NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 30 • Issue 1 • January 2026

ISSN 2046-4924

Treatments for renal cell carcinoma: NICE Pilot Treatment Pathways Appraisal

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

DOI 10.3310/GJDL0327

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Treatments for renal cell carcinoma: NICE Pilot Treatment Pathways Appraisal

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 M 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] and GJ Melendez-Torres[1]

[1]Peninsula Technology Appraisal Group (PenTAG), University of Exeter, Exeter, UK
[2]Delta Hat Limited, Nottingham, UK

[*]**Corresponding author** d.lee7@exeter.ac.uk

Published January 2026
DOI: 10.3310/GJDL0327

This report should be referenced as follows:

Lee D, Muthukumar M, Lovell A, Farmer C, Burns D, Matthews J, *et al*. Treatments for renal cell carcinoma: NICE Pilot Treatment Pathways Appraisal. *Health Technol Assess* 2026;**30**(1). https://doi.org/10.3310/GJDL0327

Health Technology Assessment

NICE TAR and DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR136008. The protocol was agreed in December 2022. The draft manuscript began editorial review in May 2024 and was accepted for publication in March 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: The National Institute for Health and Care Excellence is piloting a new approach to evaluating health technologies, which takes into consideration the full treatment pathway for a condition. This report describes the first pilot topic for the pathways process, which evaluated systemic treatments for advanced renal cell carcinoma.

Objectives: This pilot aimed to develop a decision model representing the treatment pathway that will be used to evaluate new technologies for advanced renal cell carcinoma. The pilot also evaluated a new treatment for renal cell carcinoma: cabozantinib (Cabometyx®; Ipsen, Slough, UK) plus nivolumab (Opdivo®; Bristol Myers Squibb, Princeton, NJ, USA).

Review methods: A systematic literature review was conducted to identify evidence to inform effectiveness, safety and economic model development, including systematic literature reviews, randomised controlled trials, economic evaluations, utility studies and cost and resource use data. Real-world evidence was sought following the recommendations of the National Institute for Health and Care Excellence real-world evidence framework. Structured expert elicitation informed assumptions about overall survival and progression-free survival. Network meta-analyses were conducted to evaluate the clinical effectiveness of treatments. A de novo state transition model that was constructed with a partitioned survival analysis structure was also presented. The cost perspective of the model was that of the National Health Service and Personal Social Services; the time horizon was 40 years, costs and outcomes were discounted at 3.5% per annum and a 2022 price year was used. The model allowed sequences of up to four active lines of treatment.

Information sources: The review included 118 systematic literature reviews, 30 randomised controlled trials, 122 economic evaluations, 82 studies reporting utility data and 13 studies reporting cost and/or resource use data. A total of 21 real-world evidence sources were identified. Unpublished data were provided by the manufacturer and other stakeholders (competitor companies, patient and clinical organisations). The expert elicitation recruited nine United Kingdom-based oncologists.

Results: Cabozantinib plus nivolumab was associated with better progression-free survival and overall survival than existing tyrosine kinase inhibitors as first-line treatment in the all-risk group. Using the list price of the evaluated interventions, the incremental cost-effectiveness ratio for cabozantinib plus nivolumab compared to the next non-dominated tyrosine kinase inhibitor monotherapy (pazopanib [Votrient®; Novartis, Slough, UK]) was £275,106 per quality-adjusted life-year in the all-risk population and was £379,222 in the favourable-risk population. Incremental cost-effectiveness ratios were relatively consistent across the base-case and scenario analyses. In the intermediate-/poor-risk population, the incremental cost-effectiveness ratio for pembrolizumab (Keytruda®; Merck Sharp & Dohme, London, UK) plus lenvatinib (Lenvima®; Eisai, Hatfield, UK) was £450,638 compared to cabozantinib; cabozantinib plus nivolumab and nivolumab plus ipilimumab were both dominated by cabozantinib and pazopanib monotherapy, respectively, in the base-case analysis. Quality-adjusted life-year gains were similar for cabozantinib plus nivolumab, pembrolizumab plus lenvatinib and nivolumab plus ipilimumab (Yervoy®; Bristol-Myers Squibb, Princeton, NJ, USA). Cabozantinib plus nivolumab was shown to be less effective and less expensive than pembrolizumab plus lenvatinib in most scenarios.

Limitations: Most interventions were supported by only one trial and data quality was poor. Outcomes reported in clinical trials were generally more favourable than those reported in real-world evidence, suggesting that trials may overestimate treatment benefits.

Conclusions: This pilot demonstrated the feasibility of producing a reference model, which is open source and available to relevant stakeholders without restriction. This will improve consistency in the National Institute for Health and Care Excellence's decision-making and allow for the evaluation of optimum treatment sequences for advanced renal cell carcinoma.

Future work: Future research is needed to resolve uncertainties in clinical effectiveness estimates for treatments for advanced renal cell carcinoma, including long-term effectiveness of combination treatments and their effectiveness in the favourable-risk subgroup. This would also inform further evaluation of optimal treatment sequences for advanced renal cell carcinoma.

This evaluation included the use of confidential data, including commercially sensitive data provided by the manufacturers of treatments for renal cell carcinoma. Where feasible, references to confidential data have been removed from this monograph. Any remaining instances of confidential data have been redacted.

Study registration: A final review protocol was submitted to National Institute for Health and Care Excellence in advance. Due to confidentiality issues surrounding the analysis plan, this was not deposited with PROSPERO.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136008) and is published in full in *Health Technology Assessment*; Vol. 30, No. 1. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of supplementary material

Report Supplementary Material 1 Supplementary material

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/GJDL0327).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

The supplementary materials (which include but are not limited to related publications, patient information leaflets and questionnaires) are provided to support and contextualise the publication. Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately, and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

List of abbreviations

AE	adverse event
AIC	Akaike information criterion
aRCC	advanced RCC
ASCO	American Society of Clinical Oncology
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*
BSC	best supportive care
ccRCC	clear-cell renal cell carcinoma
CDF	Cancer Drugs Fund
CNS	central nervous system
CODA	convergence diagnosis and output analysis
CPRD	Clinical Practice Research Datalink
CS	company submission
CT	computerised tomography
CTLA-4	cytotoxic T-lymphocyte-associated protein 4
DataSAT	Data Suitability Assessment Tool
DBL	database lock
DES	discrete event simulation
DIC	deviance information criterion
DoR	duration of response
DSU	decision support unit
EAG	External Assessment Group
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
eMIT	electronic market information tool
EQ-5D	EuroQol-5 Dimensions
EQ-5D-3L	EuroQol-5 Dimensions, three-level version
EQ-5D-5L	EuroQol-5 Dimensions, five-level version
ESMO	European Society for Medical Oncology
FDA	US Food and Drug Administration
FE	fixed effects
FKSI–DRS	Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index – Disease-Related Symptoms
FP	fractional polynomial
GDO	Get Data Out
HCRU	healthcare resource use
HES	Hospital Episode Statistics
HFS	hand–foot syndrome
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
IFNα	interferon alpha
IL-2	interleukin-2
IMDC	International Metastatic RCC Database Consortium
IO	immuno-oncology
IPD	individual patient data
ITT	intention to treat
IV	intravenous
IVI	Innovation and Value Initiative
KM	Kaplan–Meier
LDH	lactate dehydrogenase
LY	life-year
LYGs	life-years gained
MMRM	mixed model repeated measures
MRC	Medical Research Council
mRCC	metastatic renal cell carcinoma

MSKCC	Memorial Sloan Kettering Cancer Centre
MTA	multiple technology appraisal
mTOR	mammalian target of rapamycin
nccRCC	non-clear-cell renal cell carcinoma
NCRAS	National Cancer Registration and Analysis Service
NDRS	National Disease Registration Service
NHSCII	NHS Cost Inflation Index
NHSE	National Health Service England
NICE	National Institute for Health and Care Excellence
NMA	network meta-analysis
NOS	not otherwise specified
NSCLC	non-small cell lung cancer
NSS	nephron sparing surgery
OD	once daily
ONS	Office for National Statistics
ORR	overall response rate
OS	overall survival
PartSA	partitioned survival analysis
PAS	Patient Access Scheme
PD	progressive disease
PD-1	programmed cell death protein 1
PD-L1	programmed death-ligand 1
PF	progression-free
PFS	progression-free survival
PH	proportional hazard
PICOS	population, intervention, comparison, outcomes and study
PPS	post-progression survival
Pre-PS	pre-progression survival
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
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
RCC	renal cell carcinoma
RCT	randomised controlled trial
RDI	relative dosing intensity
RE	random effects
RECCORD	Renal Cell Carcinoma Outcomes Research Dataset
RECIST	Response Evaluation Criteria in Solid Tumours
RMST	restricted mean survival time
RN	radical nephrectomy
RWD	real-world data
RWE	real-world evidence
SACT	Systemic Anti-Cancer Therapy
SLR	systematic literature review
SmPC	summary of product characteristics
STA	single technology appraisal
STEER	structured expert elicitation resources
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
VAS	visual analogue scale
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor

Note

This manuscript is based on the Technology Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed confidential. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain language summary

There are many drug treatments currently available for advanced renal cell carcinoma, which is a type of cancer that begins in the kidney and then spreads to other parts of the body. We made a model that tells us which drug treatments work better and which ones cost more than others. We included the treatments people with advanced renal cell carcinoma might get one after the other, but we did not include surgery. We used our model to look at a new treatment, cabozantinib and nivolumab, as the first treatment for people with advanced renal cell carcinoma.

To make this model, we first looked for information about which treatments work best. We found 24 high-quality studies of treatments for advanced renal cell carcinoma. We agreed that 17 of those studies were the most useful. When we combined the studies looking at patients having their first treatment, we found that cabozantinib and nivolumab may be a good option for some people and may offer good value for money for the NHS. When we combined the studies looking at patients having their second, third or fourth treatments, we did not include cabozantinib and nivolumab, and we were unable to conclude which treatments are better than others due to limitations in the evidence available.

We then combined our results with what we know about people in the UK who are diagnosed with advanced renal cell carcinoma and how well different treatments work for them. Because we needed to use confidential information about how much different drugs cost the National Health Service, we cannot share all the results of our model. However, an important part of the project is that the model we developed can be used again by the National Health Service when new treatments come along. This will mean that decisions about which treatments the National Health Service should pay for will be fairer and more consistent.

Scientific summary

Background

The National Institute for Health and Care Excellence (NICE) is piloting a new approach to evaluating health technologies, which takes into consideration the full treatment pathway for a condition. The 'pathways' process aims to increase the efficiency of reimbursement decisions in the NHS. The approach is based on developing comprehensive and adaptable core models for specific disease areas to which new treatments can be added and compared over time. The approach can also be used to evaluate optimum treatment sequences.

This report describes the first pilot topic for the pathways process, which evaluated systemic treatments for advanced renal cell carcinoma (aRCC), which includes locally advanced and metastatic renal cell carcinoma (RCC). The project aimed to develop an evidence base to inform the development of a decision model representing the treatment pathway. The model is available open source without restriction, allowing reuse for future appraisals while maintaining confidentiality of proprietary data. As part of this phase of the pilot, the project evaluated a new treatment for aRCC, cabozantinib (Cabometyx®; Ipsen, Slough, UK) plus nivolumab (Opdivo®; Bristol Myers Squibb, Princeton, NJ, USA).

The RCC originates in the tubules of the kidney and is the most common type of kidney cancer (80% of cases). Clear-cell RCC (ccRCC) is the most common subtype of RCC (around 75%), with the remainder comprising papillary, chromophobe and others. The disease is staged from 1 to 4, according to degree of spread. Stage 3 indicates locally advanced cancer (with regional lymph node involvement), and stage 4 indicates metastatic disease beyond the regional lymph nodes. Stage 3, locally aRCC that is unresectable with surgery may instead be treated with systemic treatment in the first line. Prognostic risk scores are available to predict survival in people with RCC, including the International Metastatic RCC Database Consortium (IMDC) risk model, which is used to inform systemic treatment options alongside prior treatment exposure.

Objectives

The objectives of this analysis were to develop an open source, core decision model for systemic treatments for aRCC and use this model to estimate the costs, effects and cost-effectiveness of:

- all-risk and favourable-risk populations: cabozantinib + nivolumab versus pazopanib (Votrient®; Novartis, Slough, UK) versus tivozanib (Fotivda®; Recordati, Hemel Hempstead, UK) versus sunitinib (Sutent®; Pfizer, Sandwich, UK) as a first-line systemic therapy in people with untreated aRCC
- intermediate-/poor-risk population: cabozantinib + nivolumab versus pazopanib versus tivozanib versus sunitinib versus cabozantinib versus nivolumab + ipilimumab (Yervoy®; Bristol-Myers Squibb, Princeton, NJ, USA) versus pembrolizumab (Keytruda®; Merck Sharp & Dohme, London, UK) + lenvatinib (Lenvima®; Eisai, Hatfield, UK) as a first-line systemic therapy in people with untreated aRCC.

In addition, this analysis considered treatment options at second line and beyond, including axitinib (Inlyta®; Pfizer, Sandwich, UK), cabozantinib, lenvatinib + everolimus (Afinitor®; Novartis, Slough, UK), sunitinib, everolimus, pazopanib, nivolumab, and including tivozanib as an off-label treatment, in people with previously treated aRCC of any risk group.

Consistent with NICE methods, the analysis did not consider the cost-effectiveness of treatments for RCC that are not routinely commissioned in the NHS. This includes avelumab plus axitinib, which is currently only available to people with aRCC through the Cancer Drugs Fund.

Methods

The External Assessment Group (EAG) conducted a systematic literature review (SLR) to identify relevant published evidence and real-world data sets. The methods used were consistent with NICE-preferred methods and best practice guidance for the conduct and reporting of SLRs. As no randomised controlled trials (RCTs) directly compared every comparator head to head, the EAG conducted a network meta-analysis (NMA) to indirectly compare treatments. Development of the decision model commenced with a review of published cost-effectiveness studies, structured expert elicitation to provide estimates of parameters for which no data existed, development of the model in R (The R Foundation for Statistical Computing, Vienna, Austria), populated with effectiveness, quality of life and resource use/cost data and reporting of incremental analyses.

Systematic literature review of clinical effectiveness evidence

Searches were conducted to identify previous SLRs and meta-analyses and RCTs published since the most recent relevant SLRs. Database searches were complemented by additional hand-searching of grey literature. Ongoing RCTs were identified by review of relevant trial registries. Relevant data were extracted from study reports into a bespoke database and assessed for quality using the Cochrane risk of bias tool (v2).

The EAG received a submission of evidence from the manufacturer of cabozantinib , which was appraised and used to inform the broader project. This included SLRs and NMAs, which were reviewed and compared against the EAG's own methods. New data from the company, and other stakeholders, were extracted and included in the EAG's analyses.

Relevant real-world evidence (RWE) was sought following the recommendations of the NICE RWE framework. Additional sources were identified from stakeholder submissions.

Network meta-analysis

Evidence networks were constructed by line of treatment and risk status. Second- to fourth-line treatments were pooled, as trials generally included patients who were previously treated at multiple lines. Thus, networks were estimated for first-line treatment, second-line + and first-line treatments stratified by IMDC risk subgroup. Outcomes were estimated for progression-free survival (PFS) and overall survival (OS) as time-to-event outcomes. As these appeared to violate assumptions of proportional hazards (PH), a fractional polynomial (FP) approach was employed as the primary analytic technique, with a PH model employed as a secondary analysis. Standard NMA using a binomial likelihood was also used to estimate the overall response rate, discontinuation due to adverse events (AEs) and risk of incurring AEs of grade 3 or higher.

Decision model

A SLR was undertaken to identify previous economic evaluations of relevance to the decision problem, studies reporting quality-of-life data and UK cost and resource use studies. Methods used were consistent with those used for the clinical SLR. Learnings from these studies were used to inform the design of the EAG's decision model as well as to provide pointers to potential model input parameters.

Structured expert elicitation was conducted to inform assumptions about long-term OS and PFS in the decision model. Experts who participated in the elicitation process were experienced clinicians treating people with RCC in the NHS and had no specific personal or financial conflicts of interest for the appraisal. The elicitation process was designed using the structured expert elicitation resources developed by the University of York and delivered using R Shiny. Experts were asked to provide estimates of OS and PFS by line and risk groups in the patient population they see in practice following different treatments, including best supportive care (BSC). These estimates were used to elicit a probability distribution from each expert representing the relative likelihood of different values, which were then mathematically aggregated and used to inform survival and Indirect Treatment Comparison (ITC) model selection.

A de novo state transition model was constructed in R, but with a partitioned survival analysis (PartSA) structure also presented. The model allowed up to four active lines of treatment followed by BSC, and health states were defined by treatment line and on/off treatment at each line. Overall structural assumptions included that OS was dependent on progression status and line of treatment.

The cost perspective of the model was that of the NHS and Personal Social Services, the time horizon was 40 years with a weekly transition period and costs and outcomes were discounted at 3.5% per annum. The price year was 2022.

The RWE derived from clinical practice in NHS hospitals was used to inform baseline risk, while relative effectiveness data were drawn from the NMAs. Health-state utilities were assigned on the basis of treatment line and health state [progression-free (PF) vs. progressive disease], adjusted for age and sex. An adjustment for AEs by treatment was also applied. Resource use categories were acquisition cost of drugs, administration, routine disease management, AE and end-of-life care costs. For the evaluation of cabozantinib plus nivolumab, subsequent treatment costs and outcomes were based on a weighted average of all possible treatment sequences, using data from UK RWE.

The model was used to estimate the costs, effects and incremental cost-effectiveness of the different treatments stated in *Objectives*. A number of scenarios explored the impact of uncertainty in the base case.

Results

During the technical engagement phase of the appraisal, some analyses were updated to include additional data or information. This summary reflects those findings.

Systematic literature review of clinical effectiveness evidence

One hundred and eighteen SLRs and meta-analyses and 30 RCTs (of which 6 were ongoing) were identified in the SLR. Of the 24 complete RCTs, earliest recruitment was in 2006 and the latest datacut was from December 2019. Trials included between 3 and 200 centres, with at least 14 including a UK centre. Sample sizes varied between 22 and 1110 participants. None of the studies were considered to be at low overall risk of bias: nine were considered at high risk and eight were considered to be unclear. Despite the moderate evidence base identified, most interventions were supported by only one trial and data quality was poor, resulting in a number of uncertainties. Notably, almost all evidence identified was based on a ccRCC population, which may be associated with improved treatment response relative to other histologies. There was a lack of high-quality data for time to treatment discontinuation (TTD) and time to next treatment (TTNT) alongside poor reporting of the subsequent therapies received by participants. It was not possible to explore the effect of adjuvant pembrolizumab on the effectiveness of first-line therapies.

A total of 4 databases, 12 publications and 5 stakeholder submissions provided information on relevant RWE sources. Data extracted included treatment patterns, OS, PFS, TTNT and TTD. No data sources provided information on health-related quality of life or resource use/costs, although relative dose intensity was reported in one source which was used to calculate drug costs in scenario analysis.

Network meta-analysis

There were a number of challenges in the evidence base used to inform the NMAs. A paucity of data in the favourable-risk group prevented the use of a FP approach. The analyses also lacked meaningful data for pembrolizumab plus lenvatinib in the intermediate-/poor-risk population due to the inaccessibility of confidential data. Broad results of the NMAs suggested that cabozantinib + nivolumab was associated with better PFS and OS than tyrosine kinase inhibitors (TKIs) as first-line treatment in the all-risk group. In the intermediate-/poor-risk group, PFS for cabozantinib + nivolumab was equal to cabozantinib monotherapy up to month 15, but OS was higher for the combination through to 55 months, at which point the survival with other treatments was superior.

Decision model

Searches for literature relevant to the economic model yielded 162 papers. Modelling methods and outcomes of the cost-effectiveness analyses of various combinations varied across the available literature, including within prior NICE Health Technology Assessments. Most previous economic evaluations used a simple three-state PartSA model based on progression status. While these are relatively simple to implement, NICE committees have previously expressed concern about how such models handle subsequent lines of therapy, in particular where trial data (on which survival functions are modelled) do not match common subsequent treatment patterns in the

UK. A state transition model allows the impact of different assumptions to be tested rather than being rendered unquantifiable.

The expert elicitation recruited 9 oncologists from the 38 ones contacted. Notable results included: the clinical experts expected a higher proportion of patients to be both alive and PF at 3 years with sunitinib compared with CheckMate 9ER KM curves; cabozantinib + nivolumab outcomes were in line with projections from the trial evidence; type of prior treatment appeared to influence outcomes estimates, particularly for combination therapies, including TKI following receipt of TKI monotherapy; and clinician opinions were in general more optimistic than the data observed in RWE.

There was uncertainty in the cost-effectiveness results due to concerns about the reliability of effect data from the NMAs. Results were sensitive to assumptions about the dose and duration of treatment that would be received by patients in practice and methods used to account for non-linear pricing. The broad results, using the list price of the evaluated interventions, indicated that cabozantinib plus nivolumab was not cost-effective compared to TKI monotherapy in either an all-risk or favourable-risk population. This finding was demonstrated consistently across the base-case and all-scenario analyses. In the intermediate-/poor-risk population, sunitinib monotherapy was the most cost-effective treatment. Quality-adjusted life-year gains were not hugely different between cabozantinib plus nivolumab, pembrolizumab plus lenvatinib and nivolumab plus ipilimumab. However, cabozantinib plus nivolumab was shown to be less effective and less expensive than pembrolizumab plus lenvatinib in the majority of scenarios. When compared to nivolumab plus ipilimumab, cost-effectiveness outcomes were influenced by poor surrogacy between OS and PFS within the CheckMate 214; two methods were presented to alleviate this issue: use of TTNT as a proxy for PFS and use of the PartSA structure, and both had a notable impact on results. Results were generally consistent between the state transition and PartSA structures when using UK RWE to model baseline risk. Interestingly, cabozantinib monotherapy dominated cabozantinib plus nivolumab, pembrolizumab plus lenvatinib and nivolumab plus ipilimumab. This lacked face validity and was attributed to issues related to the CABOSUN trial; there was uncertainty about how to address this within the model.

Outcomes reported in clinical trials were generally more favourable than those reported in the UK RWE and Systemic Anti-Cancer Therapy database, suggesting that trials may overestimate treatment benefits. Modelled OS compared very well to observed OS within the UK RWE in the all-risk and intermediate-/poor-risk populations and acceptably in the favourable-risk population. Modelled OS compared less well to CheckMate 9ER data when using this as the primary data source, which was thought to be due to substantial differences between subsequent treatments given in the trial and UK practice.

Conclusions

The availability of a decision model for systemic treatments for aRCC will improve consistency in NICE's decision-making and allow for the evaluation of optimum treatment sequences for RCC. However, limitations in the evidence present challenges, and further modelling approaches may be needed to account for data scarcity associated with some treatments, such as complexities with dosing and titration, limitations of using PFS as a surrogate for OS and discrepancies between results generated from FP and PH NMAs. Optimal treatment sequences, the long-term effectiveness of combination treatments and the effectiveness of combination treatments in favourable-risk RCC also remain areas of uncertainty.

This pilot has demonstrated the feasibility of (1) the production of a reference model capable of evaluating treatment sequences as well as the introduction of a new technology within the treatment pathway and (2) the incorporation of UK RWE into economic analysis. The pilot has also highlighted the importance of thoroughly describing areas of uncertainty which are often underdeveloped within the single technology appraisal process, such as the impact of subsequent treatments on outcomes and the importance of using data reflective of the UK population to describe baseline risk.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136008) and is published in full in *Health Technology Assessment*; Vol. 30, No. 1. See the NIHR Funding and Awards website for further award information.

Chapter 1 Background

Description of the health condition

Renal cell carcinoma (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 > 80% of cases.[1] Clear-cell RCC (ccRCC) is the most common subtype, quoted as accounting for approximately 75% of cases.[1]

The 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.[2]

The scope for this appraisal is people with locally 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 the cancer is unresectable. Due to this, people with stage 4 RCC or stage 3 unresectable RCC have been included in this appraisal. For simplicity, throughout the report, we use the term aRCC to refer to people with locally advanced (stage 3) and metastatic (stage 4) RCC that is treated using systemic treatment.

Overall survival (OS) data for RCC from 2013 to 2019 were available from the Get Data Out (GDO) 'Kidney' data set, published by the National Cancer Registration and Analysis Service (NCRAS). For stage 4 clear-cell patients, 12-month survival ranged from 58.5% to 62.2%. The most severe histological subtype with the lowest 12-month OS estimates were patients with stage 4 RCC not histologically confirmed [not otherwise specified (NOS)], ranging from 13.1% to 18.4%.

The data suggest that there has been a sustained improvement in 12-month 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.

For stage 4 ccRCC, 60-month survival ranged from 19.1% to 20.1%. Patients with stage 4 RCC NOS have the poorest 12-month prognosis/lowest survival rates (ranging from 2.1% to 2.7%).

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 Centre (MSKCC) scores, are widely used tools that incorporate various factors, including performance status (PS), 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.

Histology is another key prognostic factor, with ccRCC being the most common subtype and it is generally associated with a poorer prognosis compared to other subtypes.[3] The presence of metastasis is a well-established prognostic factor in RCC, indicating the extent and aggressiveness of the disease.[3] Differentiating between visceral metastases and bone metastases (BM) is also important, as patients with BM often exhibit a less favourable outcome and suboptimal response to certain treatments, such as tyrosine kinase inhibitors (TKIs).[3]

Nephrectomy is an additional prognostic factor in aRCC. In select patients, nephrectomy has shown benefits, especially in a 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.[4] 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.[4]

Timely initiation of systemic therapy may also be a significant prognostic factor for patient outcomes, though this has been challenging to determine through published evidence. While studies have suggested that initiating systemic treatment without delay following diagnosis is associated with improved response rates and survival,[3,5] other studies have been unable to replicate this finding and either find no association or the reverse.[6–9] There can be many reasons for treatment delays,[6–9] including factors associated with both patient and healthcare service characteristics, and clinical experts advise that patients treated earlier may be those with more advanced disease that is not amenable to treatment with surgery and radiotherapy and may have poorer outcomes with systemic therapies.

Sarcomatoid features within the tumour represent another important prognostic factor in aRCC.[3,10] 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 OS rates.

Other prognostic factors in aRCC include age, tumour stage, PS[11,12] and laboratory parameters such as haemoglobin levels, LDH levels and calcium levels.[13] 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.

Current treatment pathway

The treatment pathway for RCC can be divided into interconnected decision points based on the disease staging system and line of therapy. The treatment pathway is based upon people with clear-cell histology (as are the majority of trials; see *Objectives of the pilot process and this assessment*). 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.[14]

Treatment for early-stage to locally advanced renal cell carcinoma

Surgery [partial or radical nephrectomy (RN)] is usually possible, and is the preferred treatment, for people with early-stage to locally aRCC and is usually curative.[15] After tumour resection, the cancer can be graded. Risk of recurrence is greater in higher-grade cancers.[16] 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.[16] This results in advanced, unresectable tumours.[16] The aim of adjuvant treatment is to prevent the recurrence and potential progression to advanced (unresectable or metastatic) disease.[17] Approximately, 20–40% of people who have received surgery subsequently develop mRCC.[18]

One major change is the introduction of adjuvant treatment. The National Institute for Health and Care Excellence (NICE) recommended pembrolizumab (Keytruda®; Merck Sharp & Dohme, London, UK) as an option for the adjuvant treatment of RCC at an increased risk of recurrence after nephrectomy, with or without metastatic lesion resection in October 2022.[16] 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 (IO) treatments for people who have received treatment with a programmed cell death protein 1 (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 External Assessment Group (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 (October 2022, less than a year prior to this appraisal), uptake was expected to start at 20% of the eligible population, rising to 65% in 5 years.[16] Based upon estimates of the eligible population size, the maximum uptake is expected to be 18% of the total population. Based upon data from December 2024 provided by Peter Clark at NHS England, the current (steady) uptake of adjuvant pembrolizumab in resected RCC is about 850 patients/year, which represents approximately 30% of the eligible population.

Local ablation is an alternative first-line approach of particular use in people whose renal function needs to be preserved.[19] The most commonly utilised techniques are radiofrequency ablation and cryoablation.[19]

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.[19]

Treatment for advanced and metastatic renal cell carcinoma

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)[20] and the *British Medical Journal (BMJ)* RCC best practice guidelines (July 2022).[19] 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 *Figure 1*.

Active surveillance or surgery

Treatment options for patients with aRCC include active surveillance and cytoreduction for patients with favourable-risk disease. A subset of patients with aRCC 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.[20,21] 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 with no evidence of distant metastasis.[19] 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.

Systemic treatment

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

Vascular endothelial growth factor receptor (VEGFR)–TKIs, encompassing a range of multikinase inhibitors, have emerged as the cornerstone of targeted therapies in the treatment of aRCC. These agents target VEGFRs, primarily 1–3, which play a critical role in tumour-induced angiogenesis and lymphogenesis. Standard treatments for aRCC may include various VEGFR–TKIs such as sunitinib (Sutent®; Pfizer, Sandwich, UK), pazopanib (Votrient®; Novartis, Slough, UK), tivozanib (Fotivda®; Recordati, Hemel Hempstead, UK) and cabozantinib (Cabometyx®; Ipsen, Slough, UK). These inhibitors act by impeding the activity of VEGFRs, thereby disrupting the signalling pathways involved in angiogenesis and lymphogenesis. VEGFR–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 VEGFRs. This potency can range from low (e.g. sorafenib [Nexavar®; Bayer, Reading, UK]) to intermediate (e.g. sunitinib) to high (e.g. cabozantinib and lenvatinib [Lenvima®; Eisai, Hatfield, UK]). On the other hand, selective inhibitors demonstrate an increased selectivity for VEGFRs and display intermediate (e.g. pazopanib) or high (e.g. axitinib [Inlyta®; Pfizer, Sandwich, UK] and tivozanib) in vitro inhibitory activity specifically against VEGFRs.

In 2015, nivolumab (Opdivo®; Bristol Myers Squibb, Princeton, NJ, USA) an anti-PD-1 inhibitor was approved for vascular endothelial growth factor (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 aRCC. 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 aRCC patients.[22]

First-line systemic treatment (untreated advanced renal cell carcinoma)

In the first-line treatment of aRCC, 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 first-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 first-line PD-1 inhibitor therapy, in combination with VEGFR-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. 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 (Yervoy®; Bristol-Myers Squibb, Princeton, NJ, USA). One clinical expert considered that the cabozantinib + nivolumab combination may be particularly beneficial for patients with BM due to the cabozantinib component of the treatment.

Nivolumab + ipilimumab is a recommended first-line treatment for patients with intermediate- and poor-risk diseases (TA780[23]). 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 (ICIs). NICE recommendations only allow the use of pembrolizumab + lenvatinib in patients who are 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[24]), pazopanib (TA215[25]), tivozanib (TA512[26]) are alternative treatments, in addition to cabozantinib for those with intermediate- and poor-risk disease (TA542[27]). While checkpoint inhibitors are generally preferred unless there are strong contraindications, clinical feedback to the EAG indicated that the use of first-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 real-world evidence (RWE) (UK RWE, 2022[28]) shows that 60% of patients were treated with a first-line single agent TKI in the period 2018–22 (sunitinib 25%, tivozanib 8%, pazopanib 18%, cabozantinib 9%). Yet, nivolumab + ipilimumab (23.4%) and avelumab (Bavencio®; Merck and Pfizer, UK) + axitinib (12.7%) only became available via Cancer Drugs Fund (CDF) from 2019 to 2020, respectively, and pembrolizumab + lenvatinib received its recommendation outside of the study period, which

may perhaps reflect the high usage of first-line single agent TKIs in the study period. Of note, ESMO guidelines consider sunitinib or pazopanib as 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.

Second-line and subsequent lines of systemic treatment (previously treated advanced renal cell carcinoma)

The advent of ICI combinations as the standard first-line therapy for aRCC has raised questions about the best second-line treatment strategy in this new treatment landscape. Currently, limited data are available regarding the optimal second-line treatment option for patients who have progressed on a first-line ICI-based combination therapy. International guidelines, such as those from the ESMO,[20] acknowledge the lack of robust prospective data specifically focusing on second-line treatment after first-line PD-1 inhibitor-based combination therapy.

Treatment options for second-line therapy could include a TKI, a PD-1 inhibitor or a mammalian target of rapamycin (mTOR) inhibitor. ICIs 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 first-line combination as a potential second-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 VEGFR-targeted therapy (e.g. avelumab + axitinib and pembrolizumab + lenvatinib), treatment options in the second line include axitinib,[29] cabozantinib,[30] lenvatinib + everolimus[31] and everolimus (Afinitor®; Novartis, Slough, UK) (TA432[32]), depending on the first-line treatment combination received.

While the majority of patients receive cabozantinib, in certain cases, lenvatinib + everolimus may be considered as an alternative 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 second-line treatment and is often reserved for later lines of therapy. Otherwise, first-line options of sunitinib (still on label as second-line treatment) or pazopanib (off-label as second-line treatment) or tivozanib (off-label as second-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 second-line treatment), pazopanib (off-label as second-line treatment) or tivozanib (off-label as second-line treatment). Clinical advice to the EAG indicated that cabozantinib is typically chosen as the next treatment option (although the EAG notes that it is off-label following nivolumab + ipilimumab), as administering another round of checkpoint inhibitor therapy is generally considered to be futile and is also not allowed in the UK.

In patients who were initially treated with VEGFR-directed TKI monotherapy, the recommended treatment options after disease progression include nivolumab (TA417[33]) or cabozantinib (TA463[30]), both of which demonstrated OS benefit in the second-line setting. Other options that can be considered include axitinib (TA333[29]) and lenvatinib + everolimus (TA498[31]).

While approved for second-line and third-line treatments, clinical advice to the EAG indicated that everolimus and axitinib are typically reserved for fourth-line treatment. Yet, given the toxicity of everolimus, only a small proportion of patients would be eligible to receive it. It is uncommon for patients to go beyond the fourth line, and very few would require a fifth line of treatment. This is in line with the UK RWE data set identified for this pilot.[28]

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 the disease progression, symptom control and palliative care without active treatment.[25]

The treatment pathway overview is summarised in *Figure 1*.

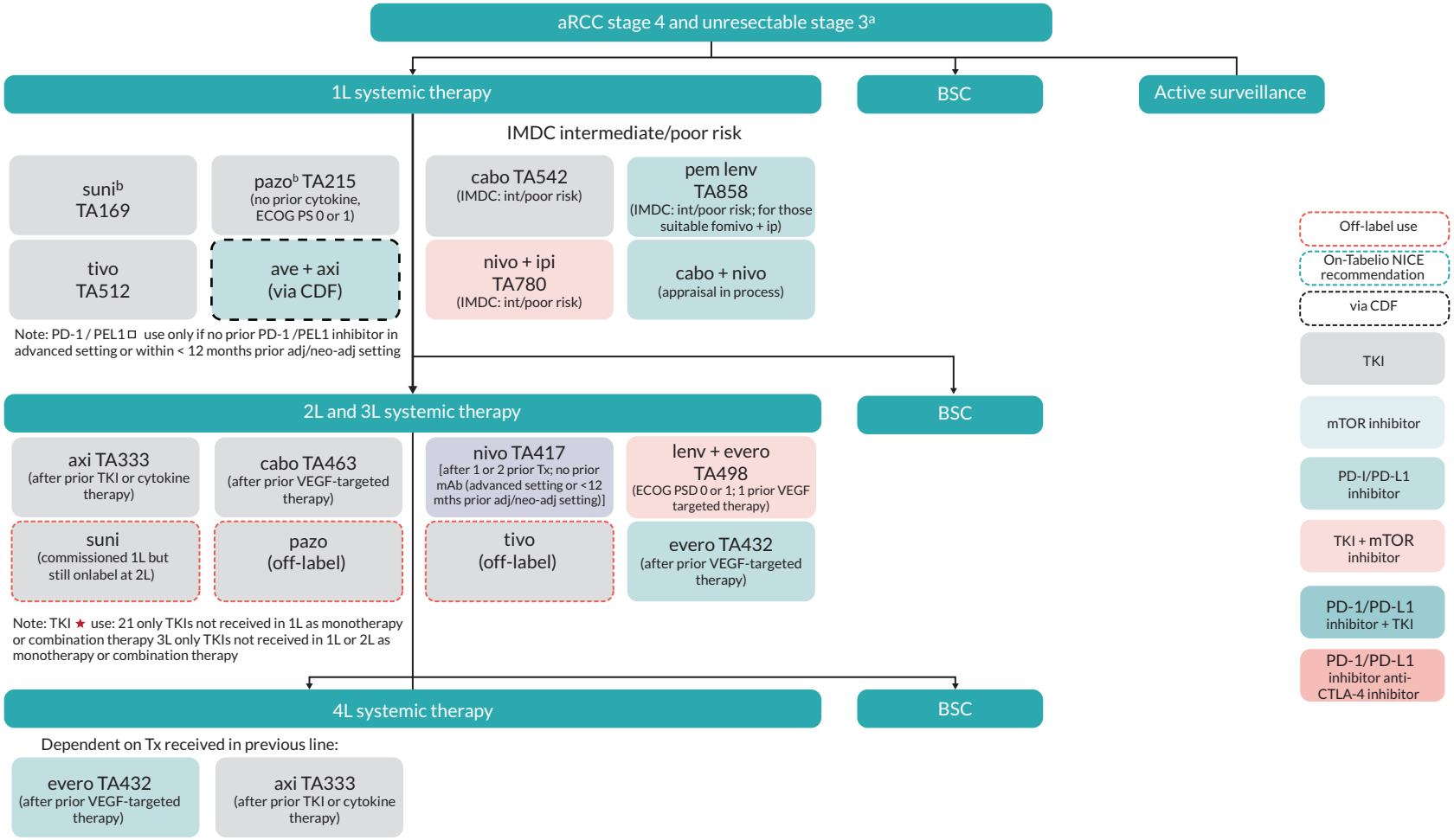


FIGURE 1 Treatment pathway for aRCC: overview. a, Cancer has spread into surrounding tissues outside Gerota's fascia or into adrenal gland. Cancer has spread to another part of the body. May or may not spread to lymph nodes. Eligible for systemic treatment; b, also considered potential alternatives to PD-1 inhibitor-based combination therapy in IMDC favourable-risk disease (ESMO guideline recommendations, 2021).1L, first-line; 2L, second-line; 3L, third-line; 4L, fourth-line; adj, adjuvant; ave, avelumab; axi, axitinib; cabo, cabozantinib; ECOG, Eastern Cooperative Oncology Group; evero, everolimus; int, intermediate; ipi, ipilimumab; lenv, lenvatinib; nivo, nivolumab; pazo, pazopanib; pem, pembrolizumab; suni, sunitinib; tivo, tivozanib; Tx, treatment.

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.

The National Institute for Health and Care Excellence 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 multicomparator 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.[34] 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 while maintaining the confidentiality of proprietary data.

An attractive model for this type of approach is the Innovation and Value Initiative's (IVI) Open-Source Value Project (Jansen *et al.*, 2019[35]). Since the project began in 2018, IVI has developed three disease models – one in rheumatoid arthritis, one in non-small-cell lung cancer (NSCLC) 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).[36] 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 *Model implementation*). 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.

Decision problem

The platform model to be developed encompasses all stages of the treatment pathway for RCC, including all treatments within the treatment pathway for first-line and subsequent line systemic treatment (see *Current treatment pathway*). Within the pilot and as summarised in this report, the EAG appraised the clinical and cost-effectiveness of one new treatment: cabozantinib + nivolumab for untreated aRCC. A summary of the decision problem for the appraisal of this treatment is provided in *Table 1*.

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 first-line treatment of aRCC on the basis of the CheckMate 9ER Phase 3 trial,[37] first by the European Medicines Agency (EMA) on 26 March 2021[38] and then by the Medicines and Healthcare products Regulatory Agency on 13 May 2021.[39] The marketing authorisation holder for cabozantinib is Ipsen Pharma. The marketing authorisation holder for nivolumab is Bristol-Myers Squibb (BMS) Pharma EEIG.

TABLE 1 Summary of decision problem

	Final scope issued by NICE	Decision problem addressed
Population	People with untreated aRCC or mRCC	Per the scope, all evidence identified was for adults
Intervention	Cabozantinib plus nivolumab (submission led by Ipsen)	Per the scope
Comparator(s)	• Pazopanib • Tivozanib • Sunitinib • Cabozantinib (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Nivolumab plus ipilimumab (only for intermediate- or poor-risk disease as defined in the IMDC criteria) • Pembrolizumab plus lenvatinib (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	• OS • PFS • Response rates • DoR • ToT/TTNT • AEs of treatment • HRQoL	Per the scope, dependent upon data availability; limited data are available for time on treatment and TTNT within published literature
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per QALY. 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 PSS 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 subgroups will be considered: • Intermediate-/poor-risk advanced mRCC 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

AE, adverse event; DoR, duration of response; PFS, progression-free survival; PSS, Personal Social Services; QALY, quality-adjusted life-year; ToT, time on treatment; TTNT, time to next treatment.

Note
Final scope available at: www.nice.org.uk/guidance/gid-ta11186/documents/final-scope

Chapter 2 Review methods

This manuscript contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report, and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Assessment group methods for reviewing clinical evidence

The EAG conducted a systematic literature review (SLR) to identify published evidence and real-world data (RWD) sets relevant to the decision problem. The methods used were consistent with NICE-preferred methods and with best practice guidance for the conduct of SLRs.[40,41] This section provides a description of the methods used to identify relevant evidence and how this was handled. The review included the following types of evidence:

- published RCTs of systemic treatments for the target population
- published economic evaluations of relevant interventions and comparators
- studies reporting quality-of-life data in the form of utilities and UK cost and resource use
- RWD, with a focus on UK settings
- input from clinical experts in the NHS
- evidence submission from the manufacturers of cabozantinib plus nivolumab.

Identification of systematic literature reviews and randomised controlled trials

Search strategies and screening process

Search strategies for all evidence types were developed by an information specialist and quality was assured by another information specialist. The search strategies used a combination of indexed keywords (e.g. Medical Subject Headings) 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. Full search strategies are supplied in *Appendix 1*.

Articles were independently assessed for inclusion by two reviewers using the prespecified inclusion/exclusion criteria. Discrepancies were resolved by discussion, with involvement of a third reviewer, where necessary. All duplicate papers were double-checked and excluded.

Some additional information about the searching and screening process for some evidence types is provided in the following subsections.

Search for randomised controlled trials

Systematic searches were conducted to identify (1) the clinical effectiveness SLRs and meta-analyses and (2) randomised controlled trials (RCTs) published since the most recent relevant systematic reviews.

The most recent, highest-quality and most comprehensive SLRs were sought to identify RCTs relevant for this appraisal. Four SLRs were identified and screened for RCTs: Heo *et al*. (2021), Liao *et al*. (2022), Riaz *et al*. (2021) and NICE TA858.[15,42–44] The search date of these SLRs was then used to inform the date from which we had to run the top-up RCT searches.

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 *et al*. (2022) and TA858.[15,43]

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 (as described in *Appendix 1*). Twenty-nine potentially relevant reports were identified by searching for RCT trial numbers in the economic studies EndNote

[Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] database, which were then sifted down to 23 studies (covering 16 of the 30 RCTs) during full-text review.

Search for real-world evidence

In line with the recommendations in the NICE RWE framework,[45] 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 aRCC 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 observational studies filter[46] and the NICE UK filter.[47] Search strategies are provided in *Appendix 1*. Results (*n* = 2683) were exported into EndNote and screened by one reviewer using the pre-specified inclusion criteria (see *Inclusion and exclusion criteria*).
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 reviewer using the inclusion criteria.
3. Web search [Google (Google Inc., Mountain View, CA, USA) and Bing (Microsoft Corporation, Redmond, WA, USA)]: 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.

Articles identified from the RWE searches were assessed in a targeted manner by one reviewer using the pre-specified inclusion/exclusion criteria (see *Inclusion and exclusion criteria*). 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 and understand the proportion of planned doses that are missed.
- Look at how HRQoL changes over time.
- Inform healthcare resource use (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).

Search for economic evaluations

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

- previous NICE TAs from 2017 onwards – 10 included
- systematic reviews of cost-effectiveness studies from 2017 onwards – 2 included
- studies evaluating cabozantinib + nivolumab – 7 included
- sequencing models – 6 included
- western (Europe, USA, Canada, Australia and New Zealand) studies by recency of data – 44 included.

Contact with study authors

Where key data were missing in the published clinical effectiveness studies, the EAG wrote to the authors. No responses were received via this route which could be included, as agreement was required from the companies funding the relevant trials. Additional data were received for CheckMate 214 from BMS.

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.

Inclusion and exclusion criteria

The inclusion criteria for the review are shown in *Table 2*. Studies that were partially relevant (e.g. a mixed population of participants with aRCC and other disease stages) would be included if results specific to the eligibility criteria were available in subgroup analyses, or if 80% of the population included in the analysis met the eligibility criteria.

For identifying RCTs, SLRs that included RCTs of pharmacological treatments for aRCC published since 2020 were included in the first round of screening. 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 aRCC of systemic treatments funded within the NHS were included, where they reported at least one outcome from OS, progression-free survival (PFS), time to next treatment (TTNT), time to treatment discontinuation (TTD), response rates, adverse events (AEs) 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 the EAG anticipated the need to use sorafenib as a linking treatment in the network meta-analysis (NMA).

Data extraction and quality assessment strategy

All relevant published evidence were extracted in one single entry in the data extraction matrix, which was developed and piloted a priori. Discrepancies were resolved by discussion, with the involvement of a third reviewer where necessary. For time-to-event outcomes, both summary hazard ratios (HRs) and figures for Kaplan–Meier (KM) curves from the last datacut were extracted. Digitisation of curves using standard methods (the Guyot algorithm[48]) was conducted, assuming censoring linearly across time intervals.

Quality assessments of individual studies were assessed by one reviewer in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) and were 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 Health Technology Assessment (HTA) programme.[49] 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 identification of 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 (TE) from the previous period.

A determination of overall domain bias was made based on the worst-rated of the subdomains – for example, overall selection bias would be determined by the worst-rated of the randomisation, allocation concealment and baseline

TABLE 2 Inclusion and exclusion criteria for the review

	Include	Exclude
RCTs and SLRs of RCTs		
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 aRCC used in the systemic setting Round 2 (RCTs and extensions of RCTs): cabozantinib + nivolumab, pazopanib, tivozanib, sunitinib, cabozantinib, nivolumab + ipilimumab, pembrolizumab + lenvatinib, axitinib, lenvatinib + everolimus, everolimus, nivolumab, avelumab + axitinib[a] 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 • ToT • response rates • DoR • AEs of treatment[b] • 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 TE modifiers Round 2: non-randomised trials, observational studies, case reports, editorials and commentaries
RWE		
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 aRCC used in the systemic setting	Any pharmacological treatment for aRCC not used in the systemic setting
Outcomes	Studies reporting at least one outcome from: • OS • prognostic variables • PFS • prognostic variables • TTP • TTNT • time to treatment discontinuation • HRQoL • current treatment pathways (sequences) being used • risk scores • health costs	Studies not reporting an included outcome

TABLE 2 Inclusion and exclusion criteria for the review (*continued*)

	Include	Exclude
Study design	RWE	
Other	Geography: UK Time: collection of data within the last 10 years with a focus on data sets including more recent data (2018 onwards)	Geography: other than UK Time: collection of data > 10 years
Cost-effectiveness studies, utility studies and cost and resource use studies		
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)	Cabozantinib + nivolumab, pazopanib, tivozanib, sunitinib, cabozantinib, nivolumab + ipilimumab, pembrolizumab + lenvatinib, axitinib, lenvatinib + everolimus, everolimus, nivolumab, avelumab + axitinib[a]	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/BSC	Any other treatments
Outcomes	Economic evaluations ICER expressed as cost per LYG 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	

DoR, duration of response; ICER, incremental cost-effectiveness ratio; LYG, life-year gained; QALY, quality-adjusted life-year; TEAE, treatment-emergent AE; ToT, time on treatment; TTP, time to progression.

a 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.

b Grade 3+ TEAEs and the total number of TEAEs leading to discontinuation will be extracted. Additional lower-grade AEs of interest may be extracted following clinical advice.

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 judgement represented the worst-rated of these domains. Performance and detection biases were omitted from key domains, as 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.

Data extraction of identified RWE was at the trial level. Included observational studies were extracted by one reviewer into tables set up in a Microsoft Word document (Microsoft Corporation, Redmond, WA, USA) and were checked by a second reviewer. Discrepancies were resolved by discussion, with the involvement of a third reviewer where necessary.

For critical appraisal, Risk Of Bias In Non-randomized Studies – of Interventions was used to appraise the quality of non-randomised comparative cohort studies. For RWE identified from external data sets, 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.[45] These criteria were considered when conducting quality appraisal.

The quality of cost-effectiveness studies evaluating cabozantinib + nivolumab was assessed using the Philips 2004 checklist for decision analytical models.[50] No quality assessment was used to evaluate other economic evaluations, as these were not directly relevant to the decision problem for the first technology under evaluation. Utility, cost and resource use studies were not quality assured, given the absence of any validated approach for this.

Handling of the company submission

The company submission (CS) was appraised and new information was used to inform the broader project. New data presented by the company that were not in published reports (e.g. new datacuts and information about trial methods contained in the trial clinical study report) were extracted and included in our appraisal and analyses. Most prominently, the CS included a new datacut from CheckMate 9ER with data up to a median of 44 months. The company provided Excel files for the relevant time-to-event end points, specifying the number of events and censors per end point for PFS, OS, TTD and time to progression (TTP) that were used in the EAG's NMAs and economic model.

Indirect treatment comparison

The 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 second-, third- and fourth-line RCCs due to trials generally including patients who were previously treated at multiple lines and similar comparator sets.

The feasibility of 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 *Prognostic factors*). We further considered trial results (including interactions in forest plots), any relevant discussion from TA858 and information in the CS. Due to clinical salience and consistency (and inconsistency) of reporting, we focused on risk, age, line, BM, 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 preclude formal testing via for example meta-regression for differences between groups and consider how analyses might have been impacted by distribution of effect modifiers across the network (see *Effect modifiers across the network*). In some proportional hazards (PH) NMAs in first line, we performed sensitivity analysis of the 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 the line of treatment (first line or second-line-plus) and for first-line treatment further stratified by the IMDC risk subgroup.

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

The set of selected trials from the search process (see *Critique of randomised controlled trials identified in the review* and *Description and critique of the design of the studies*) were processed according to steps 2 and 3 of the algorithm outlined by Dias *et al.*,[51] namely (2) identify all the trials that compare two or more comparators in the population of interest and (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.[15] As described above, these nodes were sorafenib and placebo.

Network meta-analyses 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 (ToT) and TTNT indicated insufficient studies available.

Continuous and binary outcomes were further grouped with respect to similarity of follow-up times and were combined using odds ratios (ORs), 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 HRs from the longest follow-up combined after log transformation using an inverse variance method. We also describe these as 'PH 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 Response Evaluation Criteria in Solid Tumours (RECIST)-defined progression or death assessed by blinded independent central review (BICR), with investigator assessment (IA)-assessed PFS used if BICR was not available.

Fractional polynomial network meta-analyses

The first strategy used fractional polynomial (FP) analyses, as, based on previous appraisals in RCC, it is expected that there may be issues in justifying PH for all end points. Model selection compared second-order FPs (except 'repeated powers') drawn from the set of powers defined by −2, −1, −0.5, 0, 0.5, 1, 2, 3 as the standard.[52]

Pseudo-individual patient data (IPD) 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 1 week (consistent with the model cycle length), but model fits were poor. The EAG elected to use 8 weeks in order to obtain stable results and reduce coding manipulations (2 months is the value coded by Wiksten[53]).

Initial FP model selection used frequentist fixed-effects (FE) 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).[53] Frequentist code was largely based on that provided by Wiksten.[53] The selected FP model(s) were submitted to Bayesian analysis in the next stage.

A Bayesian analysis of selected models was carried out by introducing random-effects (RE) and comparing these to FE models. RE were only be considered on the basis of 'time-invariant' heterogeneity, that is only using between-study variance on intercept terms.[52] The general framework used RE in a Bayesian framework with Markov chain Monte Carlo estimation, including informative priors from Turner *et al.*[54] where available and appropriate and vague or weakly informative priors otherwise. Turner *et al.* offer priors for a set of generic scenarios in health care and associated types of outcomes. Specifically, an informative prior for the variance of log-normal (LN) (−3.95, 1.79^2) was used, which Turner offers for pharmacological versus 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 10th value. A common strategy in Bayesian analysis is to thin Markov chains, that is to retain every *n*th value from within that chain, to reduce autocorrelation between subsequent iterations and improve the quality of inference. 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 R package (Roche, Basel, Switzerland).[55] Fitted curves were compared to the life-table estimates of the hazard, following the equation given by Collett (p. 29).[56]

To summarise, each FP analysis fits 28 models under any risk and prior treatment subgroup; see for example *Table 11* for the case line 1 PFS all risk. Any model selected from these fits is further fitted with FE or RE 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, FE approach
- select models with delta AIC≤ 5
- for each selected model, run Bayesian models (FE and RE) and calculate:
  - DIC
  - area under survival curve, up to 40-year time horizon [i.e. restricted mean survival time (RMST)]
- select models where RMST > threshold for every treatment curve over a 40-year time horizon
- select models best conforming to expert elicitation landmark distributions
- select model with minimum DIC comparing RE and FE.

Under expert elicitation, the expected survival at 5 years (conditional on surviving to 3 years) and 10 years (conditional on surviving to 5 years) were calculated for each model curve for the 1L intermediate-/poor-risk and second-line+ populations. These were compared with the elicitation distributions (see *Structured expert elicitation*). 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 (CI) 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 FPs, including singularities at zero 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 2 weeks to avoid implausibly low survival curve estimates.

Proportional hazards network meta-analyses and network meta-analyses of other outcomes

Finally, meta-analyses on PH estimates were undertaken of survival outcomes, overall response rate (ORR), discontinuation due to AEs and the risk of TEAEs 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 first-line NMA. This is because (1) the results appeared to be reported as HRs for the difference between treatment sequences rather than between treatments and (2) the crossover trials served only to connect tivozanib to the main network, and previous TAs considered that an assumption of similar effectiveness to sunitinib was appropriate.[15,57]

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 10th iteration. We used standard inconsistency and convergence checks on these models. A shorter burn-in as compared to the FP NMA was justified by a simpler model.

Chapter 3 Evidence included in the review

Studies identified and included

Randomised controlled trials

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams for the individual SLR and RCT searches can be found in *Appendix 1*, *Figures 18* and *19*. In total, 118 SLRs and meta-analyses were identified and 30 RCTs were identified – 20 from the SLRs and a further 10 from the RCT top-up search and other supplementary search techniques.

Published cost-effectiveness studies, utility studies and cost and resource use studies

In total, 162 papers were identified across the 3 searches. Some publications contained information relating to more than one review. One hundred and twenty-two papers containing relevant economic evaluations were identified; 82 papers were identified containing utility data (discussed in *Utility values*) and 13 were identified containing cost and resource use data (discussed in *Resource use and costs*). The PRISMA diagram is shown in *Appendix 1*, *Figure 20*.

Identified real-world evidence

The search and screening process for RWE is described in *Assessment group methods for reviewing clinical evidence*.

A total of four relevant databases were identified in the review of RWE (*Table 3*). Of these, data were only publicly available for the NCRAS (#1)[58] database. These data were included. Three databases [Systemic Anti-Cancer Therapy (SACT) data set (#2),[59] Clinical Practice Research Datalink (CPRD) (#3)[60] and Hospital Episode Statistics (HES) (#4)[61]] 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 time frame 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 (see *Table 3*). The authors for each of the 12 published reports containing potentially relevant data sources were contacted for access to additional data. A 3-week period was allowed for a response, with one follow-up e-mail sent. A total of four studies were excluded: four [Marchioni *et al.*, 2021 (#6)[62]; International mRCC Database Consortium (IMDC, #7)[63]; Schmidinger *et al.*, 2020 (#10)[69]; Maroun *et al.*, 2018 (#8)[64]] 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 *et al.*, 2020 (#15)][74] 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: Renal Cell Carcinoma Outcomes Research Dataset (RECCORD)[6]; UK RWE, 2022[28]; Nathan *et al.*, 2022[70]; Brown *et al.*, 2021[71]; Hack *et al.*, 2019[72]; Hawkins *et al.*, 2020[73]; NICE TA780[23].

The authors of Challapalli[65] were contacted, and access to IPD from the wider UK RWE data set used for this analysis was granted to the EAG. At the time of the NICE appraisal, these data had not been published and were therefore redacted at the request of the data holders. However, additional information was published following the NICE appraisal in a number of additional publications,[66–68] allowing additional information to be presented unredacted in this HTA monograph.

In addition to the data sets and studies identified in the EAG's review, a further four potential sources were identified in stakeholder and CSs (see *Table 3*). In addition, to these sources, the company also provided Hospital Audit Data 2022 from the same data set reported in Maroun[64] in its response to clarification question A1 (see NICE TA964[80]). Following scrutiny against the EAG's population, intervention, comparison, outcomes and study (PICOS) criteria specified in *Inclusion and exclusion criteria*, two studies were excluded on geographical location as they did not report data for UK participants: one study was conducted in Germany (Hilser *et al.*, 2023), and one study was a multicentre study in 32 worldwide institutions (Santoni *et al.*, 2019). Three studies that met the specified PICOS criteria were included (Kidney Cancer UK: Quality Performance Audit of kidney cancer services in England;[75] Nathan *et al.*, 2023;[78] IQVIA Hospital Audit Data, 2022[79]). Given that no RWE was identified evaluating the cabozantinib + nivolumab combination, the geographical criterion was relaxed to include the Hilser *et al.* (2023)[77] study.

TABLE 3 Identified potential sources of RWE

#	Name	Identified from	Included
Databases			
#1	NCRAS[58]	Web search + Health Data Research UK Innovation Gateway	Yes. Publicly accessible data for the aRCC population
#2	SACT data set[59]	Web search + Health Data Research UK Innovation Gateway	No. Data that would be required from the SACT data set for this project are not available in the public domain and cannot be accessed within the timescales of this project
#3	CPRD[60]	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	HES[61]	Web search + Health Data Research UK Innovation Gateway	No. Data that would be required from the HES data set for this project are not available in the public domain and cannot be accessed within the timescales of this project
Publications			
#5	RECCORD registry (Wagstaff *et al.*, 2016)[6]	Observational studies search	Yes (full text)
#6	REMARCC[62]	Observational studies search	No. Study reported data for North American and European centres. The authors were contacted for data from the UK centres, but no data were provided
#7	International mRCC Database Consortium[63]	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 *et al.*, 2018)[64]	Observational studies search	No. Study published in Maroun *et al.* (2018)[64] 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 response to clarification question A1. These data were included (see below)
#9	UK RWE data set 2022[28]	Observational studies search	Yes (access to data set). The authors were contacted and access to the data set was granted following contact with authors of Challapalli *et al.*[65] Additional information was published following the NICE appraisal in a number of additional publications[66–68]
#10	Real-world experience with sunitinib treatment in patients with metastatic renal cell carcinoma: clinical outcome according to risk score (Schmidinger *et al.*, 2020)[69]	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	Avelumab plus axitinib in advanced renal cell carcinoma (aRCC): 12-month interim results from a real-world observational study in the United Kingdom (Nathan *et al.*, 2022)[70]	Observational studies search	Yes (conference abstract)

TABLE 3 Identified potential sources of RWE (*continued*)

#	Name	Identified from	Included
#12	Cabozantinib and axitinib after VEGF therapy in patients with aRCC: a retrospective cohort study (Brown *et al.*, 2021)[71]	Observational studies search	Yes (conference abstract)
#13	Real-world experience of nivolumab therapy in metastatic renal cancer patients: a 3 year multi-centre review (Hack *et al.*, 2019)[72]	Observational studies search	Yes (conference abstract)
#14	Treatment patterns and health outcomes in metastatic renal cell carcinoma patients treated with targeted systemic therapies in the UK (Hawkins *et al.*, 2020)[73]	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 *et al.*, 2020)[74]	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 (February 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[23]	During grey literature screening/ data extraction	Yes (report)
Stakeholder submissions (company and other stakeholders)			
#17	Kidney Cancer UK: Quality Performance Audit of Kidney Cancer Services in England[75]	Stakeholder submission	Yes (report)
#18	Real-world data on cabozantinib in previously treated patients with metastatic renal cell carcinoma: focus on sequences and prognostic factors (Santoni *et al.*, 2019)[76]	CS	No. Study reported data for 32 worldwide centres, no data from UK centres were reported
#19	Cabozantinib + nivolumab in adult patients with advanced or metastatic renal cell carcinoma: a retrospective, non-interventional study in a real-world cohort (Hilser *et al.*, 2023)[77]	CS	Yes. Study reported data for German centres only, no UK centres were 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	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 (Nathan *et al.*, 2023)[78]	Company response form	Yes (conference abstract + poster)
#21	IQVIA Hospital Audit Data[79]	Company clarification response to question A1	Yes. The company provided Hospital Audit Data 2022 from the same data set as reported in Maroun *et al.* (2018)[64] in its response to clarification question A1. These data were included

REMARCC, Registry for Metastatic Renal Cell Carcinoma.

A summary of the information sources scrutinised is provided in *Table 3*.

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.

Critique of randomised controlled trials identified in the review

In total, 30 trials were identified for inclusion in the review. Of these, six are ongoing and are addressed below in *Ongoing studies*. The remaining 24 trials are described below and summarised in *Table 4*.

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 datacuts in published records drawing from December 2019. Trials included as few as 3 and as many as 200 centres, with at least 14 trials including UK centres; and trials had sample sizes across arms comparing relevant treatments of between 22 and 1110 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 first line of a treatment reimbursed only at second line would have been deprioritised. In one situation (NCT01136733), we deprioritised a trial arm in a three-arm trial but retained the relevant comparison.

Design of the studies

An overview of study design characteristics for the included trials is shown in *Appendix 2*, *Table 34*.

Of the 24 included trials, 18 were parallel trials and 6 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,[98] 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 6 studies,[81,96,101,103–105] 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.

Although 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 3 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).

Population

Inclusion/exclusion criteria

Included trials included participants aged ≥ 18 years, with histologically confirmed RCC, measurable via RECIST guidelines, and with participants having adequate PS [generally defined as Eastern Cooperative Oncology Group (ECOG) PS of 0 or 1, or as Karnofsky Performance Score of ≥ 70%]. All trials required participants to have locally aRCC or mRCC, 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 (CNS) metastases.

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

TABLE 4 Randomised controlled trials included

Study name	Lead reference	Population	Clear-cell type (%)	Risk score (IMDC or MSKCC)	Trt line	Comparison
ASPEN (NCT01108445)	Armstrong *et al.*, 2016, *Lancet Oncol*[81]	Locally advanced and metastatic (*N* = 108)	0	Mixed	1L[a]	Sunitinib vs. everolimus
AXIS (NCT00678392)	Rini *et al.*, 2011, *Lancet*[82]	Locally advanced and metastatic (*N* = 723)	100	Mixed	2L	Axitinib vs. sorafenib
BERAT (EUDRACT 2011-005939-78)	Grünwald *et al.* 2022, *Oncol Res Treat*[83]	Metastatic (*N* = 22)	NR	NR	2L	TKI (axitinib and sunitinib) vs. everolimus
BIONIKK (NCT02960906)	Vano *et al.*, 2022, *Lancet Oncol*[84]	Metastatic (*N* = 202)	100	Mixed	1L[b]	Nivolumab vs. nivolumab + ipilimumab, nivolumab + ipilimumab vs. VEGFR-TKI (suni + pazo)
CABOSUN (NCT01835158)	Choueiri *et al.*, 2018, *Eur J Cancer*[85]	Metastatic (*N* = 157)	100	Intermediate and poor	1L	Cabozantinib vs. sunitinib
CheckMate 025 (NCT01668784)	Motzer *et al.*, 2015, NEJM[86]	Locally advanced and metastatic (*N* = 821)	100	Mixed	2L and 3L	Nivolumab vs. everolimus
CheckMate 214 (NCT02231749)	Motzer *et al.*, 2018, *NEJM*[87]	Locally advanced and metastatic (*N* = 1096)	100	Mixed	1L	Nivolumab + ipilimumab vs. sunitinib
CheckMate 9ER (NCT03141177)	Choueiri *et al.* 2021, *NEJM*[37]	Locally advanced and metastatic (*N* = 651)	100	Mixed	1L	Cabozantinib + nivolumab vs. sunitinib
CLEAR (NCT02811861)	Motzer *et al.*, 2021, *NEJM*[88]	Locally advanced and metastatic (*N* = 1069)	100	Mixed	1L	Pembrolizumab + lenvatinib vs. lenvatinib + everolimus vs. sunitinib
COMPARZ (NCT00720941)	Motzer *et al.*, 2021, 2013, *NEJM*[89]	Metastatic (*N* = 1110)	100	Mixed	1L	Pazopanib vs. sunitinib
CROSS-J-RCC (NCT01481870)	Tomita *et al.*, 2020, *Clin Genitourin Cancer*[90]	Metastatic (*N* = 120)	100	Favourable and intermediate	1L	Sunitinib vs. sorafenib
ESPN (NCT01185366)	Tannir *et al.*, 2016, Eur Urol[91]	Metastatic (*N* = 72)	16.7	Mixed	1L[a]	Everolimus vs. sunitinib
Hutson *et al.*, 2017 (NCT00920816)	Hutson *et al.*, 2013, *Lancet Oncol*[92]	Metastatic (*N* = 288)	100	Favourable and intermediate	1L[a]	Axitinib vs. sorafenib
JAVELIN RENAL 101 (NCT02684006)	Motzer *et al.*, 2019, *NEJM*[93]	Locally advanced and metastatic (*N* = 886)	100	Mixed	1L	Avelumab + axitinib vs. sunitinib
METEOR (NCT01865747)	Choueiri *et al.*, 2015, *NEJM*[94]	Locally advanced and metastatic (*N* = 658)	100	Mixed	2L and 3L	Cabozantinib vs. everolimus
NCT01136733 (NCT01136733)	Motzer *et al.*, 2015, *Lancet Oncol*[95]	Locally advanced and metastatic [*N* = 153 (101 relevant)]	100	Mixed	2L	Lenvatinib + everolimus vs. everolimus

continued

TABLE 4 Randomised controlled trials included (*continued*)

Study name	Lead reference	Population	Clear-cell type (%)	Risk score (IMDC or MSKCC)	Trt line	Comparison
RECORD-1 (NCT00410124)	Motzer *et al.*, 2008, *Lancet*[96]	Metastatic (*N* = 410)	100	Mixed	2L and 3L	Everolimus vs. placebo
RECORD-3 (NCT00903175)	Motzer *et al.*, 2014, *J Clin Oncol*[97]	Metastatic (*N* = 471)	85	Mixed	1L[a]	Sunitinib vs. everolimus
SWITCH (NCT00732914)	Eichelberg *et al.*, 2015, *Eur Urol*[98]	Locally advanced and metastatic (*N* = 365)	87	Favourable and intermediate	1L	Sunitinib vs. sorafenib
SWITCH II (NCT01613846)	Retz *et al.*, 2019, *Eur J Cancer*[99]	Locally advanced and metastatic (*N* = 377)	87	Favourable and intermediate	1L	Pazopanib vs. sorafenib
SWOG 1500 (NCT02761057)	Pal *et al.*, 2021, *Lancet*[100]	Locally advanced and metastatic [*N* = 152 (94 relevant)]	0	Mixed	1L[c]	Cabozantinib vs. sunitinib
TIVO-1 (NCT01030783)	Motzer *et al.*, 2013, *J Clin Oncol*[101]	Metastatic (*N* = 517)	100	Favourable and intermediate	1L and 2L	Tivozanib vs. sorafenib
TIVO-3 (NCT02627963)	Rini *et al.*, 2020, *Lancet Oncol*[102]	Metastatic (*N* = 350)	100	Mixed	3L and 4L	Tivozanib vs. sorafenib
VEG105192 (NCT00334282)	Sternberg *et al.*, 2010, *J Clin Oncol*[103]	Locally advanced and metastatic (*N* = 435)	100	Favourable and intermediate	1L and 2L[d]	Pazopanib vs. placebo

NEJM, *New England Journal of Medicine*; NR, not reported; Trt, treatment.

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

b This trial is not currently included in the first-line network because it includes a non-standard design.

c This trial is not included in the first-line network as the definition of PFS is not consistent with other trials and given a different histological profile.

d This trial is not included in the first-line network as no other trials were compared to placebo and therefore inclusion did not add any value to the network.

Baseline characteristics

An overview of the sample characteristics in the prioritised and deprioritised trials is shown in *Appendix 2*, *Table 35*.

Histology Of the 24 trials, 17 included patients with ccRCC 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 ccRCC and non-clear-cell RCC (nccRCC).[83,91,97–99] The remaining three trials specifically targeted participants with predominantly nccRCC histology.[81,106]

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[83,90] did not enrol any participants assessed as having poor risk, and a further three prioritised[98,99,101] and two de-prioritised trials[92,103] enrolled a very low number of participants assessed as being at poor risk (i.e. ≤ 5% of the trial sample). One prioritised trial[85] 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 first line of systemic therapy. Of these 17 trials, 14 were only in participants receiving first-line treatment. The remaining three trials enrolled patients to receive first-line and second-line treatments; for these trials, the proportion of patients receiving their first systemic treatment ranged from 93% to 53%. Ten trials in the first-line setting were prioritised for inclusion.

Correspondingly, 10 trials enrolled participants receiving second line or later therapy. Distinguishing between participants receiving second-line and third-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 first-line and second-line patients, an additional two trials enrolled only participants for the second line of treatment. Of the remaining five trials, four enrolled participants across second line and third line, with ranges of second-line treatment between 20% and 72%; and one trial enrolled only participants at the third and fourth lines of therapy, with 60% of participants at third line. Seven trials in the second-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 11 trials that reported data on prior TKI use were prioritised for inclusion, and this included 5 trials[94–96,102,105] that enrolled only participants with prior TKI, 5 trials[85,89,104,107,108] that enrolled participants only without prior TKIs and 1 trial[82] 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 6 trials with no participants who had previously received prior immunotherapies were included.

Interventions and comparators

An overview of the intervention characteristics used in the included trials is shown in *Appendix 2*, *Table 36*.

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 BSC as a comparator, but placebo was used as a comparator in two trials,[96,103] 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).

Outcomes

The outcomes reported in the 24 trials are summarised in *Table 5*. The account of outcomes is derived from publicly available trial reports.

Overall survival

Overall survival 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 2 years in one case and up to 7 years in one case; five trials had final follow-ups of between 2 and 4 years. An additional trial reported a 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 < 2 years, one reported a median follow-up of 2 years and the final three trials reported a median follow-up of between 3 and 6 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 the 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. IO therapies followed by IO therapies) was only inconsistently presented across trials.

Progression-free survival

Progression-free survival on first treatment was also included in all 24 trials. Twenty-three 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 that included clinical progression and symptomatic deterioration (investigator-assessed). Where PFS censoring rules were mentioned in trial protocols, the trials specified US Food and Drug Administration (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. Ten trials assessed PFS via BICR, 2 used an independent review committee with no or unclear blinding and the remaining 12 were investigator-assessed. All combination therapy trials were assessed via independent central review except CheckMate 214.

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 TTNT outcomes. Six trials reported time to treatment discontinuation.

Duration of response and response rate

Duration of response (DoR) 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 AEs, though discontinuation due to AEs was also reported for nearly all trials in some form.

TABLE 5 Outcomes reported by RCTs included in the review

Trial name	OS	PFS	TTP	TTNT	TTD	DoR	Response rate	AEs	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[a]			X	X	X	X
JAVELIN RENAL 101	X	X				X	X	X	[b]
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

a Time to treatment failure.
b Utility data reported within the economics section of TA645, but not clinical outcomes reported by arm.

Health-related quality of life

Health-related quality of life outcomes were identified for 16 trials. Utility data identified are presented in the later sections relevant to the economic analysis (see *Utility values from CheckMate 9ER*).

Critical appraisal of the included studies

The quality assessments of RCTs included are presented in *Appendix 2*, *Table 37*.

None of the included trials were appraised as being at a low overall risk of bias. Of the 17 prioritised trials, 5 were appraised as being at a high risk of bias and 12 were appraised as being at an unclear risk of bias. Overall, results from the NMA were based on underlying evidence with various methodological shortcomings. Most notable of these was the challenge of handling high levels of attrition from study arms in analyses, a major component of which was driven by disease progression and treatment switching. Trials were also typically sponsored by industry, and were rarely blinded, which has implications for HRQoL data.

Clinical effectiveness results from trials identified in the review.

Overall survival

Overall survival results are summarised in *Appendix 2*, *Table 38*.

First line

Overall risk Nine prioritised trials evaluated OS in an overall risk population in the first-line setting. All trials included a comparison with sunitinib (seven 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 and 54.3 months. Median OS was between 29.3 and 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% CIs 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 datacut 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 first line for the favourable-risk group. All trials involved a comparison with sunitinib (seven trials) and/or sorafenib (two 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% CIs, 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 first line in an intermediate-/poor-risk population. All trials involved a comparison with sunitinib (eight trials). Sorafenib was only compared with sunitinib (two 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 and 37.8 months for sunitinib (not reported 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% CIs 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% CIs 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.

Second-line-plus

Seven trials reported OS in the second-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.

Progression-free survival

Progression free survival results are summarised in *Appendix 2*, *Table 39*.

First line

Overall risk Nine trials reported PFS for the overall risk population in the first-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 first-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 and 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% CIs 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 first-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% CIs 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 BM. 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.

Second-line-plus

In the second-line setting, eight trials evaluated PFS, all in an overall risk population. The treatments evaluated were everolimus (five trials), cabozantinib (one trial), everolimus + lenvatinib (one trial), sorafenib (three trials) tivozanib (two trials), nivolumab (one trial), axitinib (one trial) and placebo (one trial). All trials included a comparison with either placebo, everolimus or sorafenib. Median PFS was 1.9 months for placebo, between 3.7 and 5.5 months for everolimus and was 3.9–5.7 months 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 outperformed 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.

Response rates

Response rates are summarised in *Appendix 2*, *Table 40*.

First line

Overall risk Nine trials reported response rates in an overall risk population at first line. All trials involved either a comparison with sunitinib (nine trials) and/or sorafenib (three 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 the 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 first line. All trials involved a comparison with sunitinib. Response rates for sunitinib ranged between 45.8% and 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 was seen 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 first line. All trials involved a comparison with either sunitinib (five trials) or sorafenib (one 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% CIs 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.

Second-line-plus

Seven trials reported response rates in the second-line-plus, all in an overall risk population. Treatments evaluated were everolimus (five trials), sorafenib (two trials), axitinib (two trials), cabozantinib (one trial), everolimus + lenvatinib (one trial), tivozanib (one trial), nivolumab (one trial) and placebo (one 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.

Duration of response

Duration of response results are summarised in *Appendix 2*, *Table 41*.

In total, nine trials reported the DoR with treatment: five[104,107–109] in the first-line setting and four[82,95,105,110] in the second-line-plus setting.

First line

In the first-line population, the comparator in all trials was sunitinib. The median DoR for sunitinib ranged between 14.5 and 32.0 months for patients at overall risk (five studies[104,107–109]) and it was 20.8 months and 9.8 months in those with favourable and poor risk, respectively (one trial[109]). DoR 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 (one trial[109]) and for cabozantinib + nivolumab (one trial[108]), pembrolizumab + lenvatinib (one trial[107]) and NIVO/IPI (one trial[104]) in the overall risk group. In the overall risk population, and in descending order, median DoR was not reached for nivolumab + ipilimumab (with a follow-up of over 5 years in CheckMate 214[104]), and it was 26.7 months for pembrolizumab + lenvatinib, 22.08 months for cabozantinib + nivolumab and 19.4 months for avelumab + axitinib. In the JAVELIN trial,[109] unlike for sunitinib where there was a difference in DoR between favourable and intermediate-/poor-risk groups, median DoR was similar: 22.6 months and 19.3 months for favourable and intermediate-/poor-risk groups, respectively.

Second-line-plus

In the second-line population, all trials reported the DoR in the overall risk group. Two trials used everolimus[95,105] as the comparator and two trials[82,110] used sorafenib. Median DoR ranged from 8.5 to 14 months for everolimus and from 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 DoR; while NCT01136733 included a higher proportion of participants at poor risk, it also primarily included people treated at second line, while more than a quarter of participants in CheckMate 025 (28%) were receiving third-line treatment.

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

Time to next treatment

Time to next treatment results are summarised in *Appendix 2*, *Table 42*.

Results were only available for two trials, CheckMate 9ER and CheckMate 214, with data provided by the manufacturers in confidence as part of this appraisal.

Time on treatment

Time on treatment results are summarised in *Appendix 2*, *Table 43*.

Results were available for eight trials evaluating first-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 treatment discontinuation were reported, whereas all other studies reported only the duration of treatment.

The median duration of treatment was reported for sunitinib in six trials[89,90,98,104,107,108] 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 DoR in the overall risk population was also available for pazopanib (two trials[89,99]), cabozantinib + nivolumab (one trial[108]), nivolumab + ipilimumab (one trial[104]), pembrolizumab + lenvatinib (one trial[107]) and tivozanib (one trial[101]). In descending order, the median treatment duration was 21.8 months for cabozantinib + nivolumab, 17 months for pembrolizumab + lenvatinib, 12 months for tivozanib, 7.9 months for

nivolumab + ipilimumab and 5.7–8 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 treatment was only available from one trial in the favourable-risk population. These 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 (confidential information has been removed) 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 months, and it was comparable with the overall risk population. Median duration of treatment for cabozantinib and for nivolumab + ipilimumab was no different than their comparator, sunitinib; 8.4 months and (confidential information has been removed), respectively. Treatment duration was substantially longer for cabozantinib + nivolumab than sunitinib, however, at a median of (confidential information has been removed).

In the second-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 (two trials), nivolumab (one trial), axitinib (one trial), tivozanib (one trial), sorafenib (two trials), and placebo (one trial).

In descending order, duration of treatment was a mean of 8.2 months for axitinib, median of 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.

Adverse events of treatment

Discontinuation due to adverse events

Adverse events are summarised in *Appendix 2*, *Table 44*.

No studies reported separate AE 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 first-line setting, in an intermediate-/poor-risk population that was the entire the trial sample.

First line

In the first-line setting, nine studies reported the rate of discontinuation due to AEs in an overall risk population. All trials involved a comparison with sunitinib (seven trials) and/or sorafenib (four trials). Pazopanib was evaluated in two trials, and 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% and 28.4% for sunitinib and between 7.0% and 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 AEs 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 AEs in an intermediate-/poor-risk population. The rate of discontinuation was similar for cabozantinib and sunitinib.

Second-line-plus

Seven trials reported the rate of discontinuation due to AEs in the second-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 AEs ranged between 0% and 16.1% for everolimus, 12.4% and 29.7% for sorafenib and 0% and 7.5% for axitinib. Rates of discontinuation were generally lower than in the first-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 < 15% of the trial arm.

Grade 3+ adverse events

Grade 3+ AEs are summarised in *Appendix 2*, *Table 45*.

First line

Nine trials reported the rate of grade 3+ AEs in an overall risk population in the first-line setting. All trials involved a comparison with sunitinib (seven trials) and/or sorafenib (four trials). Pazopanib was evaluated in two trials, and 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% and 83.3% for sunitinib, 57.1% and 75.0% for sorafenib and 62.2% and 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 was lower 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 first-line setting.

Second-line-plus

Four trials reported rates of grade 3+ AEs in the second-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+ AEs across trials, with rates for everolimus ranging between 36.8% (in the trial with the longest follow-up) and 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 > 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.

Health-related quality of life

Health-related quality of life is summarised in *Appendix 2*, *Table 46*.

First line

Overall risk Six trials reported HRQoL in an overall risk population in the first line: all six trials reported a disease specific HRQoL [Functional Assessment of Cancer Therapy Kidney Cancer Symptom Index (FKSI) total (four trials) and FKSI–Disease-Related Symptoms (DRS) (two trials)] and four trials reported generic HRQoL [EuroQol-5 Dimensions (EQ-5D) index (three trials) and EQ-5D visual analogue scale (VAS) (one trial)]. This section focuses 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 first and second lines (TIVO-1).

Baseline FKSI total scores were reported to be between 58.4 and 60.1 (reported in two 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 and 0.83 (CheckMate 9ER, CLEAR and TIVO-1). None of the trials reported meaningful differences in HRQoL between treatment arms according to established minimum/minimal(ly) important difference thresholds.[111–114] 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 first 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 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 first line: three trials reported disease-specific HRQoL [FKSI total (two trials) and FKSI–DRS (one trial)] and two trials reported generic HRQoL [EQ-5D index (one trial) and EQ-5D VAS (one trial)]. All trials involved a comparison with sunitinib. Treatment with sunitinib was followed by meaningful reductions in HRQoL (two trials). Pembrolizumab + lenvatinib was associated with a smaller reduction in disease-specific and generic HRQoL (one 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.

Second-line-plus

Four trials reported HRQoL in the second-line-plus, all in an overall risk population: four trials reported disease-specific HRQoL [FKSI total (three trials) and FKSI–DRS 9 (one 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.

Critique of real-world evidence identified for this appraisal

Description and critique of real-world evidence

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 (2009–22) and, as such, capture a substantial number of patients and treatment data. The study populations include people with locally aRCC and mRCC. Sample sizes ranged from smaller cohorts, such as the Nathan *et al.* (2022)[70] study with an advanced population of 36 patients (N = 36), to larger patient populations in the UK RWE,[28] which included 1319 patients. Interventions assessed in the available evidence typically reflect the NICE recommendations during the data collection periods covered by the included evidence. Summary study characteristics are provided in *Appendix 3*, *Table 47*.

The Kidney Cancer UK report[75] provided results from a 2-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. 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 PICOS [postoperative 30-day and 12-month all-cause survival in M0 kidney cancer patients who undergo RN or nephron sparing surgery (NSS) and metastatic kidney cancers should receive SACT or active surveillance].[75]

Hospital Audit Data (IQVIA 2022[79]) 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 data set[58] provides publicly accessible data for the aRCC population. The NCRAS forms part of the NDRS in NHS Digital. On 1 October 2021, responsibility for the management of the NDRS transferred from Public Health England to NHS Digital. The EAG has extracted publicly available data from the NCRAS, specifically the 'GDO' programme. The 'Kidney' data set contains information on the incidence, treatment rates, survival and routes to diagnosis (and other key outcomes) for patients with malignant kidney cancer in England from 2013 to 2019.
- The UK RWE data set[28] (access kindly provided by the coinvestigators: Amarnath Challapalli, Amit Bahl, Gihan Ratnayake, Ricky Frazer and John McGrane) included 1319 mRCC participants from 15 UK centres, who commenced first-line systemic therapies between June 2018 and August 2022. The data set included patients from all regions of the UK (with a focus on England), a mix of secondary and tertiary centres and patients from urban and rural settings. Access to the data set was provided following contact with the authors listed on a conference abstract identified in the searches (Challapalli, 2022[65]). The EAG was able to conduct its own analyses using this data set.

Baseline characteristics and risk scores

The included evidence all focused on people with aRCC. Median age ranged from 59 to 68 years,[6,23,28,58,70–73,75,77–79] which broadly mirrored the populations included in the clinical trials (see *Appendix 2*). Ten analyses reported sex; in these analyses, the majority of participants were male.[6,23,28,70-73,75,77,78] Baseline characteristics are summarised in *Appendix 3*, *Table 48*.

Of the 12 analyses, the RECCORD data set[6] included only patients with clear-cell histology. Six analyses[23,28,70,73,77,78] included a mix of histologies, but ccRCC consistently appeared as the most prevalent histological subtype across the studies ranging from 67% in Hilser *et al.* (2023)[77] to 91% in SACT TA780[23] data. Four[6,28,72,77] of the 12 analyses reported the proportion of participants who had undergone prior nephrectomy; this ranged from 50%[6] to 67.9%[72].

The ECOG PS was reported in five analyses,[23,70,71,77,78] 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;[23,70,77,78] one analysis[71] 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.[23]

Risk score was reported in eight studies.[23,28,70,72,73,77–79] Risk distribution was measured by a combination of IMDC (or Heng criteria),[23,28,70,72,77–79] MSKCC,[73] 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 *et al.*, 2022[70] to 100% in the SACT TA780[23] data set). The proportion of participants assessed as intermediate or poor risk broadly matched that in the clinical trial populations (see *Appendix 2*).

Outcomes

The outcomes reported in the included RWE are summarised in *Appendix 3*, *Table 49*.

Critical appraisal real-world evidence studies

The DataSAT was completed for UK RWE (2022),[28] Hawkins *et al.* (2020),[73] RECCORD (Wagstaff *et al.*, 2016[6]) and SACT TA780.[23] The details of critical appraisal of RWE are reported in *Appendix 3*. Overall, the included data sets provided

relevant information from UK practice in terms of treatment patterns and efficacy outcomes (e.g. OS, PFS, TTNT, discontinuation and dosing information). However, in interpreting the information, the EAG considered the changes in the treatment landscape over time, given the differences in treatment pathways between the study periods and the present.

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 SACT, resulting in different treatment sequences at different points in the pathway. A recent audit commissioned by Kidney Cancer UK[75] highlighted this variation, which 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 1 month before to any time after diagnosis ranged from 39.7% [95% CI (33.7 to 46.1)] to 70.7% [95% CI (59.6 to 79.8)]. These variations were broadly similar from 1 month to 4 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 *et al.*, 2016); Hawkins *et al.*, 2020; UK RWE]. The studies were all UK studies and were aligned with the NICE Pathways for locally aRCC and/or mRCC, meaning that the received treatments were consistent with NICE-recommended systemic therapies. The broad time period across the three analyses (2008–22) 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 *Appendix 3*, *Table 50*. 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 CDF.

As noted, the interventions received by participants in the earlier data sets[6,73] reflected the treatments available during the study period; that is, 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 *et al.* (2020)[73] and RECCORD (Wagstaff *et al.*, 2016)[6] 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 *et al.* (2020)[73] and RECCORD (Wagstaff *et al.*, 2016[6]) 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 *et al.* (2020)[73] and RECCORD (Wagstaff *et al.*, 2016)[6] data sets, respectively], reflecting the timing of the NICE recommendations. In third line, the majority of participants received everolimus or axitinib.

A summary of treatments used from first line to fourth line from three RWE sources (data collection period 2008–22) is provided in *Appendix 3*, *Table 51*. 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 first line: avelumab + axitinib (13%), nivolumab + ipilimumab (23%), pazopanib (18%), sunitinib (25%), cabozantinib (9%) and tivozanib (8%) aligned with NICE recommendations. The data indicate that a small proportion (5%) of patients are treated with interventions not recommended by NICE (e.g. in clinical trials). At second line, the data indicate that 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 the risk group, the proportions treated were similar apart from a higher proportion of patients receiving nivolumab + ipilimumab in first-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 being the most common treatment at third line (48%) and axitinib being the most common treatment at fourth line (43%). A full breakdown of interventions received in the cohort is provided in *Appendix 3*. The EAG conducted an analysis to show the pathway of care from first-line to fourth-line treatment as shown in *Figure 2* (data are reported in *Appendix 3*, *Table 52*).

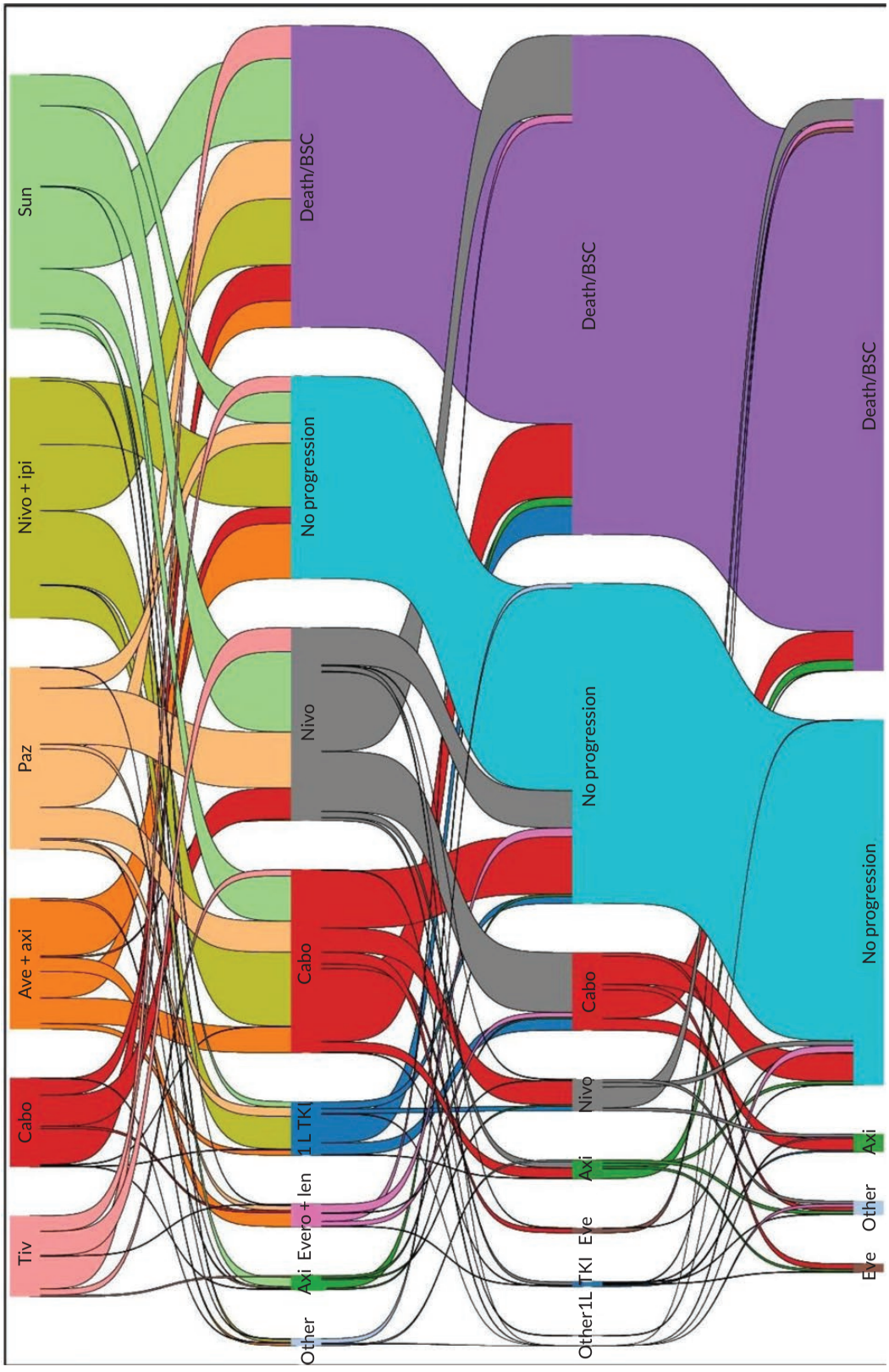


FIGURE 2 Sankey diagram for UK RWE. Note: Patients receiving treatments not currently prescribed in the NHS have been removed from first line for readability. Evero + Len, everolimus + lenvatinib; Nivo + Ipi, nivolumab + ipilimumab; Paz, pazopanib; Sun, sunitinib; TIV. Source: UK RWE (2022).[28]

Overall survival

Overall survival was reported in eight sources. The studies evaluated various interventions and lines of therapy and typically reported median OS as well as OS rates at different time points, and a summary is provided in *Appendix 3*, *Table 53*. OS data for RCC were sourced from the NCRAS-published 'Kidney' data set via the GDO platform.[58]

The Kidney Cancer UK audit report[75] reported postoperative 12-month all-cause survival in M0 kidney cancer patients who undergo RN or NSS. A total of 241 (2.8%) of M0 patients who had RN or NSS died within 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% of cases. Circulatory disease and other cancers were the underlying causes for over 30 deaths each (14.3% and 13.4% of patients, respectively).

In the UK RWE (2022[28]) data set, the median OS for patients who received first-line treatment was 25.16 (95% CI 23.39 to 27.47) months. The survival estimate was 68.9% at 12 months, falling to 27.3% at 48 months. For those patients who received a second-line treatment, median OS from second-line treatment initiation was 17.25 months, with a 1-year survival estimate of 63.1%. For those patients who received a third-line treatment, median OS from third-line treatment initiation was 10.55 months, with a 1-year survival estimate of 47.3%. For those patients who received a fourth-line treatment, median OS from fourth-line treatment initiation was 5.32 months, with a 1-year survival estimate of 18.8%. The analysis found that risk score was a significant predictor of survival time. A log-rank test stratifying OS at first line by favourable or intermediate/poor status generated $p < 0.0001$, with a Cox HR of 2.59 [95% CI (2.09 to 3.22)]. Refer to *Appendix 3* for KM curves.

Similarly in the Hawkins *et al.* (2020)[73] analysis, the median OS decreased with each subsequent treatment. The Hawkins *et al.* (2020)[73] study found that the 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 first-line and second-line treatments, 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 (first line), 14.3 months (second line)] when compared with intermediate-risk [median OS; 15.8 months (first line); 8.9 months (second line)] and poor-risk patients [median OS; 6.1 months (first line) and 3.3 months (second 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 *et al.*, 2016[6]) data set, median OS was measured from first-line treatment initiation and was 23.9 (95% CI 18.6 to 29.1) months over the 13.8-month 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 first-line treatment (20.9 months). Median OS was significantly longer in participants who switched to second-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 first-line therapy is causing prolongation of survival. A similar 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 first-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 *et al.* (2020)[73] 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)[28] data set. Several factors could explain the lower median OS observed in Hawkins *et al.* (2020) when compared to RECCORD and the UK RWE.

Firstly, the RECCORD (Wagstaff *et al.*, 2016)[6] study only included patients with ccRCC, which constituted 80% of the cohort in Hawkins *et al.* (2020)[73] and 82% of the UK RWE data set. 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 data set[28] and was 64 years

in the Hawkins *et al.* (2020)[73] data set. 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 *et al.* (2020)[73] compared to RECCORD[6] is the inclusion of patients on clinical trials in the RECCORD[6] data set as well as a small number of patients receiving interleukin-2 (IL-2) or interferon alpha (IFNα). Hawkins *et al.* (2020)[73] suggest that the inclusion of these patients in RECCORD[6] could have contributed to a higher median OS. Hawkins *et al.* (2020)[73] conducted a subgroup analysis of 89 patients excluded from the main analysis because they received IL-2 or IFNα at any point during the study. This analysis revealed a substantially longer median OS (47.5 vs. 12.9 months for first-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 *et al.* (2020)[73] 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 *et al.* (2020)[73] analysis compared to similar studies.

These differences (patient selection, age and treatment mix) could, in part, explain the differences between the median OS in the UK RWE (2022)[28] and the Hawkins *et al.* (2020)[73] data set, and 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 *et al.* (2020),[73] the majority of participants received sunitinib (60.7%) or pazopanib (37.7%), whereas the UK RWE[28] data set 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 *Treatment patterns* and *Appendix 3*).

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 first-line treatment showed survival rates at 6-, 12- and 18-month time points 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 6-, 12- and 18-month time points and 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.[23]
- Cabozantinib and axitinib as second-line treatments demonstrated similar median OS.[71] Median OS was lower in RWE than in clinical trials for both cabozantinib (vs. everolimus) and for axitinib (vs. sorafenib) (see *Appendix 2*).
- Nivolumab in second and subsequent lines of treatment showed a 12-month survival rate of 56.88%. OS data were not reported for CheckMate 025 (median OS not reached) with which to compare (see *Appendix 2*).[72]
- Avelumab + axitinib first-line treatment showed a 12-month OS rate of 86%.[70] OS data were not reported for JAVELIN Renal 101 (not estimable) with which to compare (see *Appendix 2*).

Progression-free survival

Four sources reported data on PFS. A summary is provided in *Appendix 3*, *Table 54*.

The UK RWE (2022[28]) cohort reported a median PFS for first-line treatment of 11.93 months (95% CI 10.81 to 13.86), reducing to 3.68 months (95% CI 2.23 to 4.60) in the cohort of patients receiving fourth-line treatment.

In a retrospective cohort study (February 2016–April 2019; England), evaluating nivolumab in the second and subsequent lines of treatment (Hack *et al.*, 2019),[72] 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.

In a retrospective cohort study (Hilser *et al.*, 2023)[77] evaluating patients with mRCC receiving cabozantinib + nivolumab first line, the PFS rate at 6 months was 81.9%. This was broadly aligned with the rate reported in the CheckMate 9ER trial for cabozantinib + nivolumab (79.6%) (see *Progression-free survival*).

A prospective cohort study (August 2019–January 2022; UK), evaluating patients with aRCC receiving avelumab + axitinib first line via an early access scheme (Nathan *et al.*, 2022),[70] reported median duration of follow-up and PFS of 12 months.

Three sources reported TTP:

- In the UK RWE (2022[28]) data set, median TTP at first line was 10.12 months (95% CI 9.03 to 11.27). The correlation of TTD and PFS (first line) and TTP (first line) was 0.87 (Spearman's correlation). Refer to *Appendix 3* for KM curves of TTP by line of treatment and for TTP on first-line treatment risk stratified.
- In the RECCORD study (Wagstaff *et al.*, 2016),[6] at the time of analysis, disease progression had been experienced by the majority (66.1%) of patients on first-line therapy (median duration of follow-up: 13.1 months, 95% CI 12.0 to 14.1 months). Median time to disease progression was 8.8 months (95% CI 7.7 to 9.9 months). There was a significant association between the time from RCC diagnosis to first-line treatment and disease progression ($p = 0.019$). Estimated TTP was the shortest for patients who had started first-line treatment within 100 days of diagnosis [16.8 months (95% CI 14.1 to 19.5 months)].
- Hack *et al.* (2019)[72] reported that 59.3% had disease progression in the cohort of mRCC patients who received nivolumab in second-line-plus treatment. TTP was not reported.

Additional outcomes

Three sources reported TTNT, five sources reported data on discontinuation and none of the studies reported HRQoL data or UK costs. The UK RWE did report data that enabled the calculation of relative dosing intensity (RDI) which could be used to calculated drug costs. Information about these outcomes is provided in *Appendix 3*, *Tables 55* and *56*.

Critique of published cost-effectiveness studies, utility studies and cost and resource use studies

Cost-effectiveness evaluations of cabozantinib plus nivolumab

Seven publications reported an economic evaluation of cabozantinib + nivolumab (*Appendix 4*, *Tables 57* and *58*).[121–128] All publications used data from CheckMate 9ER [with the majority using the March 2020 database lock (DBL)]. The four papers that were 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 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, the quality-adjusted life-year (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 fits. Quality assessment was conducted using the Phillips checklist and is included in the original EAG report.[129]

One study explored the difference that a state transition versus a partitioned survival analysis (PartSA) model structure made upon outcomes, and it was concluded that there was little difference. Drug costs, quality of life and effectiveness inputs were key drivers in the majority of models, with 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.

Other published economic evaluations

Data were extracted from 43 additional studies; 26/43 (60.5%) of the studies looked at first-line therapies, and 17/43 (39.5%) investigated second-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. All clinical effectiveness and utility inputs were derived from trials, or from previous NICE TAs.

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 second-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 computerised tomography (CT) scan and death from RCC. Both of those studies by Raphael (2017, 2018[130,131]) seem to discuss the same state transition model evaluating the cost-effectiveness of the perspective of the Canadian healthcare system.

Sixteen first-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)[132] evaluated two combinations: lenvatinib + pembrolizumab and lenvatinib + everolimus; both combinations resulted in a similar QALY gain. Yfantopoulos *et al.* (2022)[133] 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 QALYs than sunitinib across all studies. Pazopanib yielded a slightly higher number of QALYs than sunitinib, albeit by a negligible margin of < 0.1 in all studies except one, which used RWE (Nazha, 2018[134]), where sunitinib exhibited better performance. For the second 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.

Utility studies

A total of 82 studies were identified in the literature containing utility values for people with aRCC (first, second 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 TAs, European and western (non-European) studies containing utility values (published from 2017 onwards) were considered to be most relevant for consideration. Using the prioritisation criteria, 34 studies were identified.

- 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),[135] Amdahl *et al.* (2017),[136] Porta *et al.* (2021),[137] Henegan *et al.* (2022),[138] Motzer *et al.* (2021),[139] Mouillet *et al.* (2017),[140] Cella *et al.* (2019),[141] Cella *et al.* (2021),[142] Cella *et al.* (2022),[143] Bedke *et al.* (2022)[144] and Buckley *et al.* (2019).[145] 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 EuroQol-5 Dimensions, five-level version (EQ-5D-5L) mapping.

Ten published NICE TAs were identified, which met the prioritisation criteria. 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 first-line treatments (and among second-line treatments), and 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 approach was used. Treatment-specific utility values were not commonly used within the 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 first- and second-line treatments, specifically TA645 and TA498, respectively (see *Utilities used in the model* for more detail).

Utility values from CheckMate 9ER

The company submitted additional utility estimates based on HRQoL data collected in the CheckMate 9ER trial. The EAG identified limitations with these data, and these were not used in the economic model. A full critique of these estimates is available in the EAG report for this appraisal, though the specific utility estimates were confidential and so were redacted.[129]

Cost and resource use studies

A total of 13 studies were identified in the literature containing cost and resource use data (see *Appendix 4*) for people with aRCC across different lines of therapy (namely, first, second and subsequent lines), of which there were 10 NICE TAs and 3 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 drug and administration costs, disease management or health-state costs based on the healthcare resource utilised and terminal care costs. Some studies also reported AE 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 TAs or feedback from clinical experts. Unit costs associated with the HCRU were derived from NHS reference costs and Unit costs of Health and Social Care from Personal Social Services Research Unit (PSSRU), etc. Summary of cost and resource use information from published studies and from previous NICE TAs has been provided in *Appendix 4*, *Tables 59* and *60*).

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 TAs. However, the source of resource use frequency was quite varied across the studies. *Appendix 4* compares the different sources for resource use inputs and provides a 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.

Description and critique of the evidence presented by the company

The CS for cabozantinib with nivolumab comprised a main submission, an appendix and a subsequent submission with updated efficacy data from CheckMate 9ER. The EAG requested IPD from the company to enable the NMA and survival analysis to be run as robustly as possible, but this was not received.

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 with locally aRCC or mRCC 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 6 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 (September 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 DBL (median follow-up was 32.9 months) was presented. This was later superseded by a fourth DBL with a 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 PH models. In response to clarification question A21 on the validity of the PH 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 intention-to-treat (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 *Critical appraisal of the included studies*. 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 was pragmatically accepted as presenting low risk of bias due to the use of interactive voice or web response systems 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 AEs 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, which raise questions about the generalisability of this trial. Emerging observational evidence on the use of cabozantinib + nivolumab suggests that AE 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 the 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 on both generalisability and achievability of the observed effect sizes. The company was unable to provide correct data on the continuation of treatment post progression within the time frame of the appraisal; however, given the time to treatment discontinuation curves were similar to the PFS curves, this was not considered as a major issue.

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 6*).

By means of comparison, considering earlier follow-up points for the company's primary outcome, PFS rates were: 79.6% versus 59.9% at 6 months, 67.9% versus 48.3% at 9 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 in *Table 7*. Combined intermediate/poor data were also provided for certain outcomes. In particular, it is notable that findings for OS do not suggest a TE 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 no benefit in the favourable-risk group in HRQoL measured

TABLE 6 Key results from 44-month follow-up for CheckMate 9ER

Outcome	Cabozantinib + nivolumab (*n* = 323)	Sunitinib (*n* = 328)
BICR-observed PFS events	230	248
Median PFS months (95% CI)	16.56 (12.75 to 19.48)	8.38 (6.97 to 9.69)
HR PFS (95% CI)	0.59 (0.49 to 0.71), $p < 0.0001$	
Median OS months (95% CI)	49.48 (40.31 to N.E.)	35.52 (29.24 to 42.25)
HR OS (95% CI)	0.70 (0.56 to 0.87)	
Increase in ORR (95% CI)	56.0% (50.4 to 61.5)	28.0% (23.3 to 33.2)
Median TTR months	2.83	4.32
Median DoR months	22.08 (17.97 to 26.02)	16.07 (11.07 to 19.35)
Median PFS-2 months	44.65 (35.94 to N/A)	25.07 (20.96 to 32.36)
HR PFS-2 (95% CI)	0.63 (0.51 to 0.78), $p < 0.0001$	
Number of patients remaining on treatment[154]	57	32
Median TTD months	(confidential information has been removed)	(confidential information has been removed)
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-L1 inhibitor-based regimen (101/288; 35.1%)
Median TTNT	(confidential information has been removed)	(confidential information has been removed)

N/A, not applicable; TTD, time to discontinuation; TTR, time to response.

TABLE 7 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 to 24.71) Con: 13.86 (9.56 to 16.66)	Int: 16.59 (11.86 to 20.04) Con: 8.67 (7.00 to 10.38)	Int: 9.92 (5.91 to 17.56) Con: 4.21 (2.92 to 5.62)
HR PFS (95% CI)	0.72 (0.49 to 1.05)	0.63 (0.49 to 0.80)	0.37 (0.24 to 0.57)
Median OS (95% CI)	Int: N.A. (40.67 to N/A) Con: 47.61 (43.63 to N/A)	Int: 49.48 (37.55 to N/A) Con: 36.17 (25.66 to 45.96)	Int: 34.84 (21.36 to N/A) Con: 10.51 (6.83 to 20.63)
HR OS (95% CI)	1.07 (0.63 to 1.79)	0.75 (0.56 to 1.00)	0.46 (0.30 to 0.72)
ORR % (95% CI)	Int: 67.6 (55.7 to 78.0) Con: 45.8 (34.0 to 58.0)	Int: 56.4 (49.0 to 63.6) Con: 27.7 (21.4 to 34.6)	Int: 41.0 (28.6 to 54.3) Con: 10.3 (4.2 to 20.1)

Int, intervention; N/A, not applicable; cabozantinib with nivolumab. Con, control, sunitinib.

by the FKSI-19, with quality of life declining from baseline in both risk groups.[155] This heterogeneity in effectiveness creates uncertainties in generalisability and in decision risk. Data on the potential for effect modification by PD-L1+ status were presented as commercial in confidence.

Treatment-related AEs 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. The most common treatment-related AEs were diarrhoea, hand–foot syndrome (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 EuroQol-5 Dimensions, three-level version (EQ-5D-3L) UK utility index, 0.04 (0.01–0.07), reaching significance at most time points, with small–to-moderate effect sizes (0.2–0.5).[141] Patients were less likely to be bothered by side effects of cabozantinib + nivolumab regardless of risk (intermediate-/poor-risk OR, 0.50; 95% CI, 0.34 to 0.75; favourable-risk OR, 0.51; 95% CI, 0.28 to 0.91).[155] This analysis, however, needs to be considered in the context of the higher rates of discontinuation and dose reduction seen for cabozantinib + nivolumab.

Indirect comparisons

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

The majority of included trials were associated with either first-line or second-line-plus populations, but in one prioritised trial, TIVO-1,[101] the study population was mixed. In both cases, analyses by line of treatment were available.

Networks were formed for first- and second-line-plus treatments for the outcomes OS, PFS and ORR, taking into account the availability of information (as HR, KM curves or response rates) and at first line for two IMDC risk categories: intermediate/poor and favourable. Network diagrams for first-line PFS and OS (all risk) are shown in *Figure 3*. Other networks in draft form are supplied in *Report Supplementary Material 1*.

Many networks are not complete. Following the precedent in TA858 and other previous RCC appraisals, two treatments (sorafenib and placebo) were introduced as connecting nodes. At first line, for PFS, this connects tivozanib and results in a complete network, but for OS, tivozanib is excluded (see *Report Supplementary Material 1*). 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 to be 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 first-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 *Report Supplementary Material 1*.

For line 2+ networks under FP analyses, the BERAT trial was removed from the network; indeed, BERAT was only helpful for some NMAs in other outcomes. The BERAT trial gives uninformative estimates of TE [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 Register record for more details[156]). There are two corollaries: that (1) inference to treatment with axitinib is lost and that (2) 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 PH 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

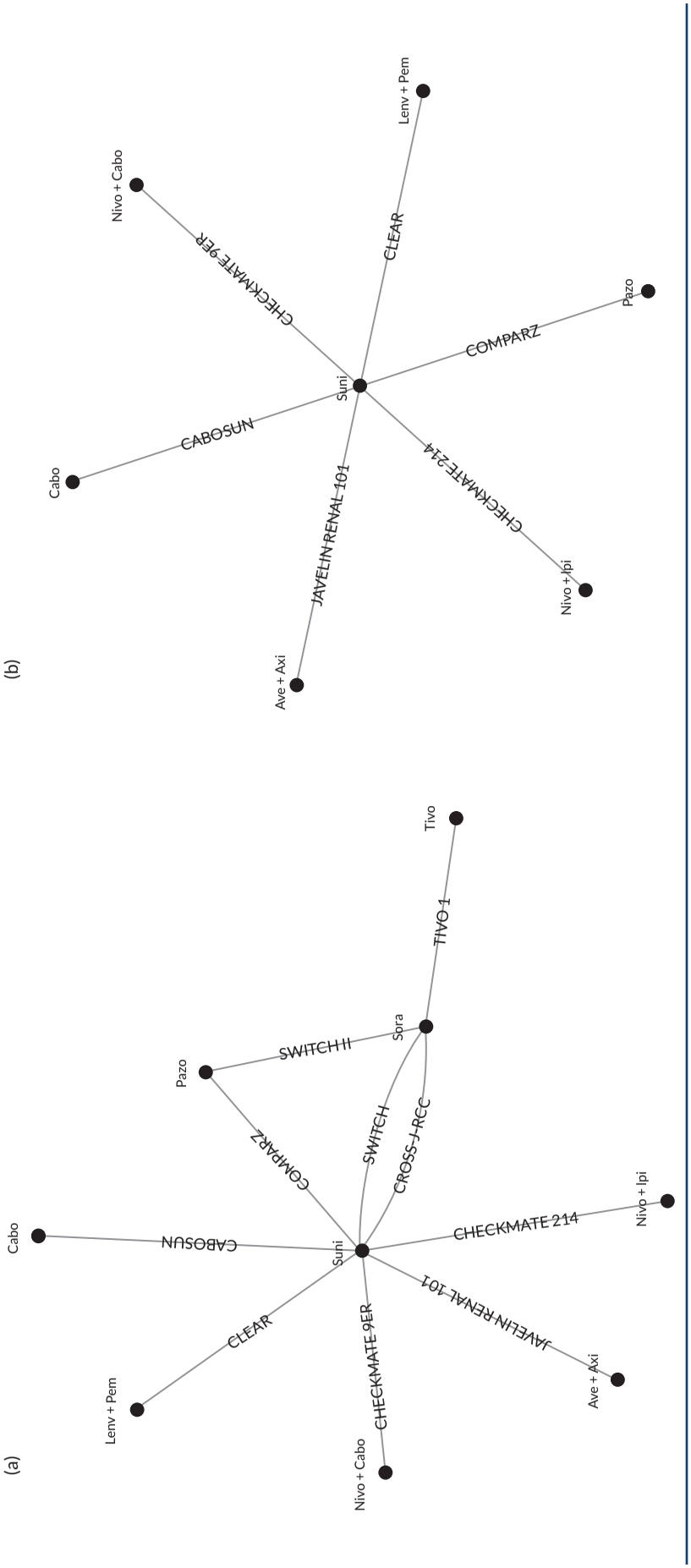


FIGURE 3 First-line network diagram for (a) PFS and (b) OS, each with summary HR and KM information.

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 TEAEs, as described below. Within subsequent cost-effectiveness analysis, given the difficulties in making comparison to axitinib within the NMA, we test the assumption of equivalence with everolimus consistent with previous TAs.

As can be seen in *Figure 3*, for first-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 (see *Report Supplementary Material 1*) with no available risk subgroup KM 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 KM 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 KM curves were available were CheckMate 9ER and JAVELIN Renal 101, and OS data were not available for JAVELIN Renal 101. Given this, only time-invariant NMA was conducted for the favourable-risk subgroup. PH NMAs at second-line-plus included all relevant comparators with the exception of pazopanib, as a reliable link could not be made to the network.

Investigation of proportional hazards

Report Supplementary Material 1, *Figures 1–4* contain log-cumulative hazard plots for included trials. Results of tests for PH using Schoenfeld residuals (i.e. Grambsch–Therneau tests) and based on EAG's digitisation of curves are provided in *Table 8*. 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-PH across the network and for both outcomes. This is including with respect to key trials in the analysis, including CheckMate 9ER (also discussed in *Analyses conducted by the company*).

The EAG scrutinised log-cumulative hazard plots alongside tests of PH. For PFS, visual assessment of PH 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 came 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 PH. 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 PH is unlikely to be valid within either the first-line or second-line-plus aRCC setting.

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 first-line network are shown in *Figure 4*. KM curves for PFS for first- and second-line-plus can be found in *Report Supplementary Material 1*.

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. In an e-mail communication, the trial sponsor, BMS, provided some rationale for the anomalous result. Based on this information, the EAG concluded that the distribution of PD-L1 status at baseline may be relevant. This characteristic was poorly

TABLE 8 Results of tests for PH 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 PH; No represents evidence of violation of PH. Lenvatinib arm dropped from analysis for three-arm NCT01136733 trial.

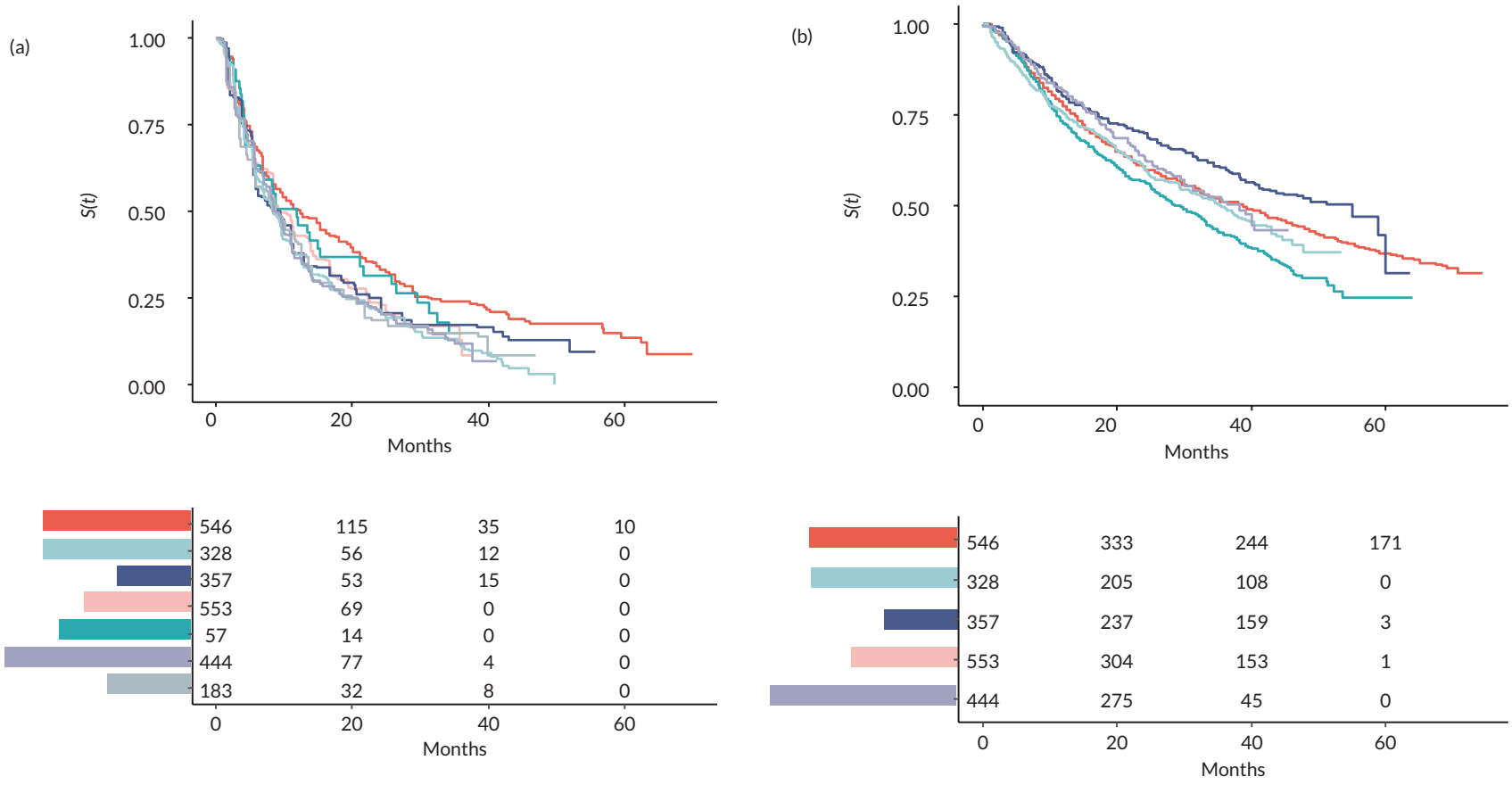


FIGURE 4 Survival data for (a) PFS and (b) OS for the central node (suni) of the first-line network; all-risk population.

reported across trials, meaning that it was not possible to confirm whether trial populations were comparable. However, the EAG did not consider it to be likely that this factor would solely explain the finding. The EAG did consider it as plausible, however, that this could be a chance finding or that it could be due to the use of IA 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. KM curves for OS for first- and second-line-plus can be found in *Report Supplementary Material 1*.

Summary information for select potential effect modifiers is shown in *Table 9*. IMDC risk category is a primary effect modifier according to clinical advice.

A network graph for PFS of first-line treatments overlaid with the proportions in risk subgroups is shown in *Figure 5* (following Cope *et al.*[157]). 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.

Report Supplementary Material 1 presents the balance of other TE modifiers across the first-line network; network graphs with TE modifiers are likewise shown in *Report Supplementary Material 1*.

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 BM enrolled; cabozantinib was considered by one of the experts consulted to be particularly effective in patients with BM, 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. Network graphs showing subsequent TKIs, IOs and other treatments are shown in *Report Supplementary Material 1*.

Results of time-dependent network meta-analysis

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 *Report Supplementary Material 1*.

As explained above, sunitinib plays a central role in the first-line networks and was selected as the reference treatment along with CheckMate 9ER as the reference study (*Report Supplementary Material 1*). For second-line-plus networks, everolimus was chosen as the reference treatment and CheckMate 025 as 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 *Indirect treatment comparison* is given in *Table 10*. 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 as noteworthy; however, the EAG generally preferred RE models where these were supported by visual inspection and by the estimability of chosen models.

First-line progression-free survival all risk

The results of the frequentist model selection for PFS (first-line trials) are summarised in *Table 11*, which shows AIC values by the two exponents of each FP fit. The model with lowest AIC has FP exponents, −2 and −0.5. In this instance, no other models attained AIC values within five points of the minimum.

TABLE 9 Summary information for select effect modifiers

Trial name	Age (median)[a]	Risk status (%)[b]			Line		BM (%)[a]	% 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 two risk factors, split between favourable and intermediate NR		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

2L+, second-line-plus.
a Where results were available by arm the figures are shown separated by a bar (|).
b In some cases, these do not add up to 100% due to rounding and risk status not having been recorded for some patients.

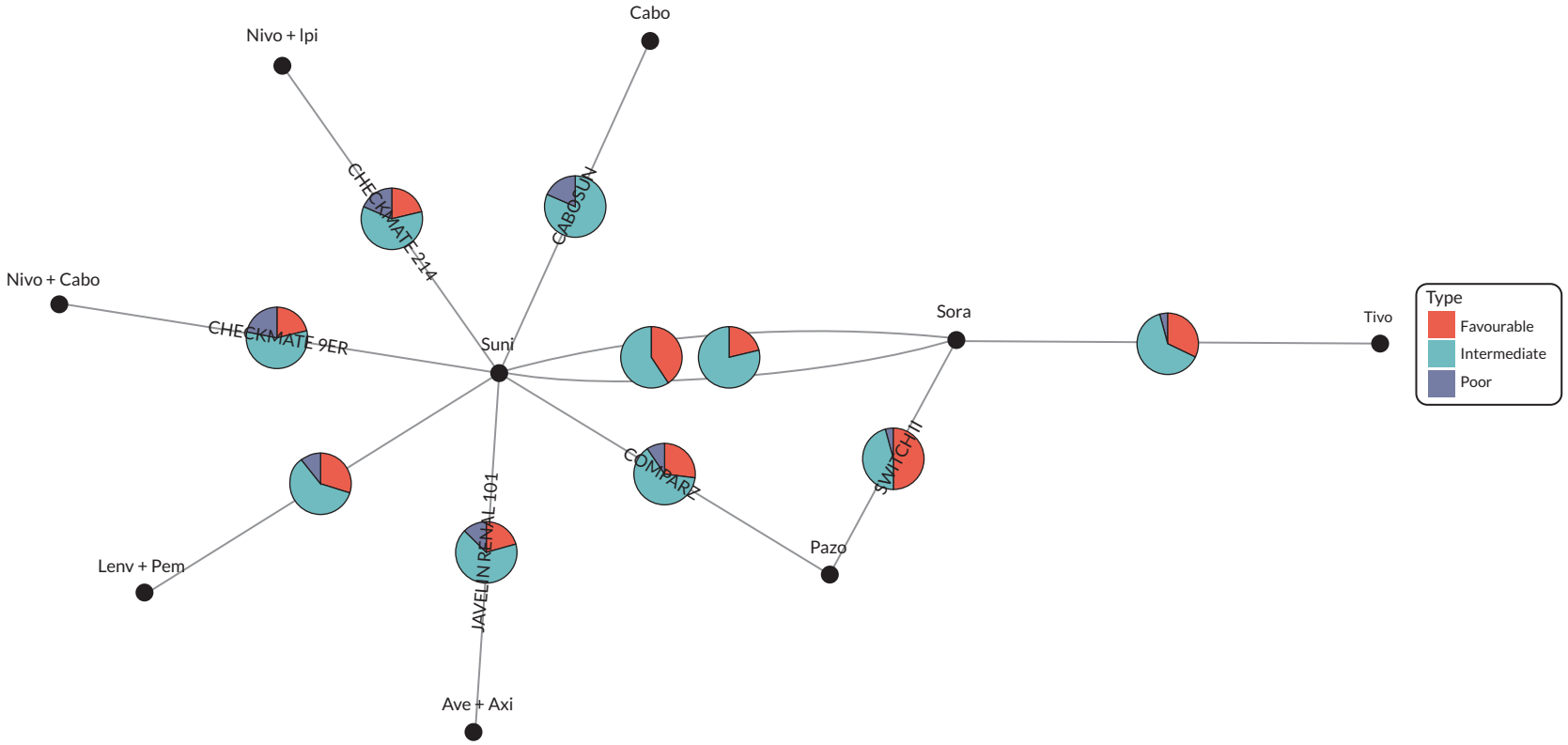


FIGURE 5 First-line network with proportions of IMDC risk subgroups overlaid. The locations of pies are jittered when there are multiple trials between treatments. Note: Three crossover trials (CROSS-J-RCC, SWITCH and SWITCH II) and one parallel group trial (TIVO-1) did not include (or included very few) poor-risk patients, and the CABOSUN trial did not include favourable-risk patients. Sora, sorafenib.

TABLE 10 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
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

TABLE 11 Akaike information criterion values for FP fit, first-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	–	–	–	–	–	–	–	–

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

The fitted log-hazards under the NMA with the best-fitting (by AIC) FP model are shown by trial in *Report Supplementary Material 1*. The trials approach a relatively constant hazard after about 20–40 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 FE and RE is shown in *Table 12* (see also *Report Supplementary Material 1*). In this case, the RE model has lower DIC. HRs from fitting by frequentist and Bayesian (RE) methods are shown in *Figure 6*. Results are qualitatively similar. Survival curves under the Bayesian approach are shown in *Report Supplementary Material 1*.

A number of observations on the presented survival curves bear noting. First, HR plots in *Figure 6* suggest that, over time, treatments with higher HRs than sunitinib are other TKIs, whereas all other treatments than pembrolizumab + lenvatinib 'settle' into HRs < 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 a 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 time frame 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 *Report Supplementary Material 1*. It should be noted that cabozantinib, nivolumab + ipilimumab and pembrolizumab + lenvatinib are only recommended for intermediate and poor-risk patients.

First-line overall survival 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 (*Report Supplementary Material 1*); 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 (*Report Supplementary Material 1*) 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 2 months (Klaassen, 2023).[158] The log-hazard (*Figure 7*) and survival curves (*Report Supplementary Material 1*) 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 *Report Supplementary Material 1*. As with PFS in first-line, cabozantinib has an unusually high level of uncertainty, likely due to the shorter time frame of follow-up. Compared to PFS findings, findings for OS in this line are considerably more equivocal, possibly due 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 in about month 50.

First-line progression-free survival intermediate/poor risk

Findings for PFS in first-line for patients with intermediate or poor risk are presented in *Figure 8*, with additional information given in *Report Supplementary Material 1*. The optimal model had polynomial terms of −2.0 and −0.5 and performed well in terms of AIC (*Report Supplementary Material 1*). 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 KM 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, which is reflected in a longer-term survival benefit emerging near the 60-month point (*Report Supplementary Material 1*). Cabozantinib monotherapy was predicted to have PFS similar to, or above, cabozantinib + nivolumab throughout the time period.

First-line overall survival intermediate/poor risk

Findings for OS in first-line for patients with intermediate or poor risk are presented in *Figures 9* and *10*, with additional information given in *Report Supplementary Material 1*. The optimal model had polynomial terms of −0.5 and 0.5 and performed well relative to other models with AIC (*Report Supplementary Material 1*). Similar patterns of uncertainties in predicted survival curves (*Report Supplementary Material 1*) were seen as in the analysis of PFS in intermediate and poor risk mentioned above. HR functions over time show a 'fanning out', with corresponding survival curves suggesting

TABLE 12 Comparison of fixed and RE Bayesian models for PFS for first-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

meanDev, mean deviation; pD, effective number of parameters.

Note
Using FP model with exponents previously selected by frequentist methods.

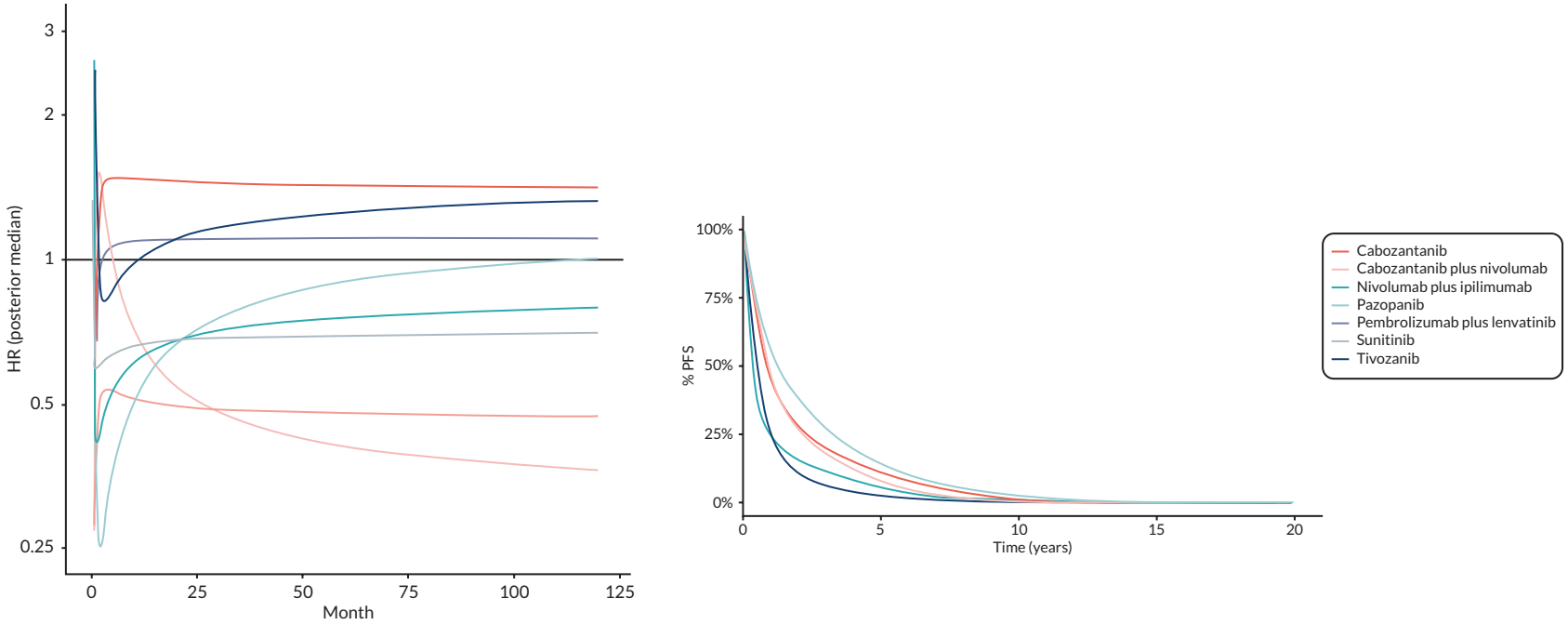


FIGURE 6 Time-dependent HRs for PFS for first-line all risk. Left: Bayesian analysis (RE). Right: Prediction survival following implementation in the economic model. The reference treatment is sunitinib (central node in the network).

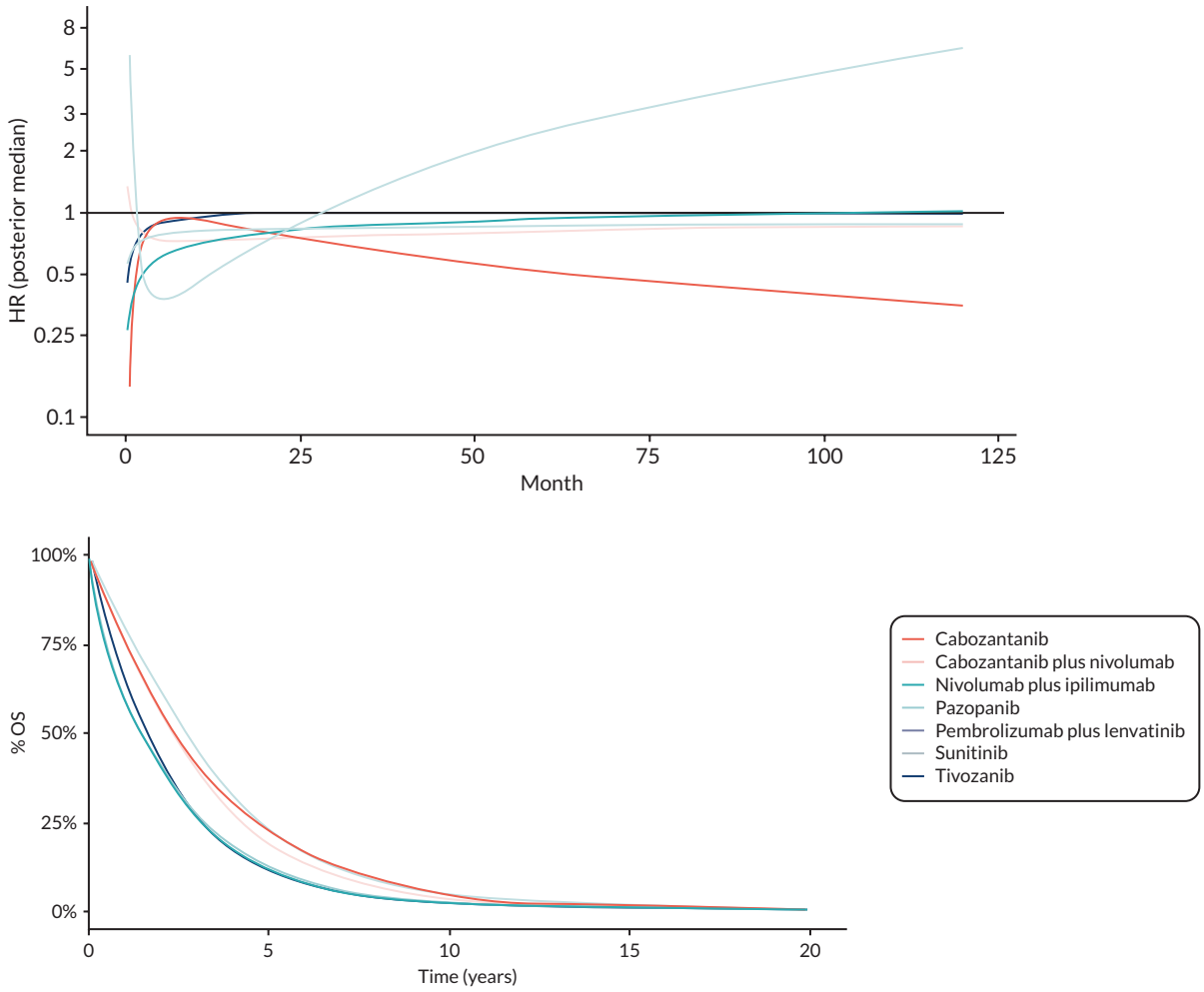


FIGURE 7 Hazard ratios and survival curves for OS for first-line all-risk (Bayesian analysis). Top: Bayesian analysis (RE). Bottom: Prediction survival following implementation in the economic model.

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.

Second-line-plus

Findings for second-line and beyond outcomes are presented in *Report Supplementary Material 1*. 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 time points. Findings for PFS (see *Report Supplementary Material 1*) 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 second-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 time points for which hazards where available, possibly due to the relatively limited follow-up time available from relevant trials to inform longer-term estimates (see *Report Supplementary Material 1*). It should be stressed that predicted survival plots (see *Report Supplementary Material 1*) reflect substantial uncertainty.

Interpretation and limitations

The EAG's FP NMAs sought to compare different treatments in each network on the basis of time-varying HRs; that is, by constructing the estimated HR for each treatment against a common comparator as a function of time. Using a

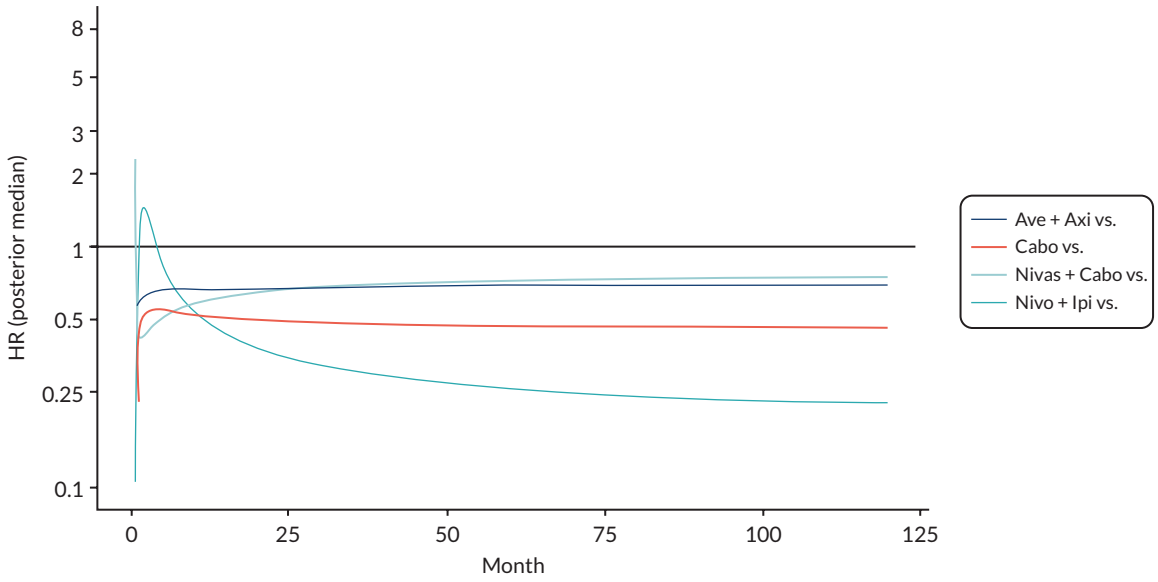


FIGURE 8 Hazard ratios and survival curves for PFS for first-line intermediate/poor risk (Bayesian analysis).

multipronged assessment process, the EAG was able to select appropriate and justifiable models for each evidence network analysed. Importantly, the evidence of non-PH in a range of included trials (see *Investigation of proportional hazards*) justified preference for a FP method over a method assuming PH (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[53] of a subset meant that it was practical for a large number of models to be efficiently assessed. At the frequentist model selection stage, all second-order FP 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 FE and RE (albeit time-invariant). When RE 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 FP NMA for the favourable-risk group. This was justified on the basis of sparse availability of relevant KM 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 the interpretation of OS analyses in ways that are difficult to quantify across the network.

Finally, FP NMAs require a choice of the 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 the optimal model choice. While expert elicitation for PFS outcomes was helpful, particularly at the 5-year time point, it did not resolve all uncertainties in situations of multiple relevant choices. Thus, in the cost-effectiveness model, scenario analyses using PH NMAs are used as well.

Results of the time-invariant network meta-analysis

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

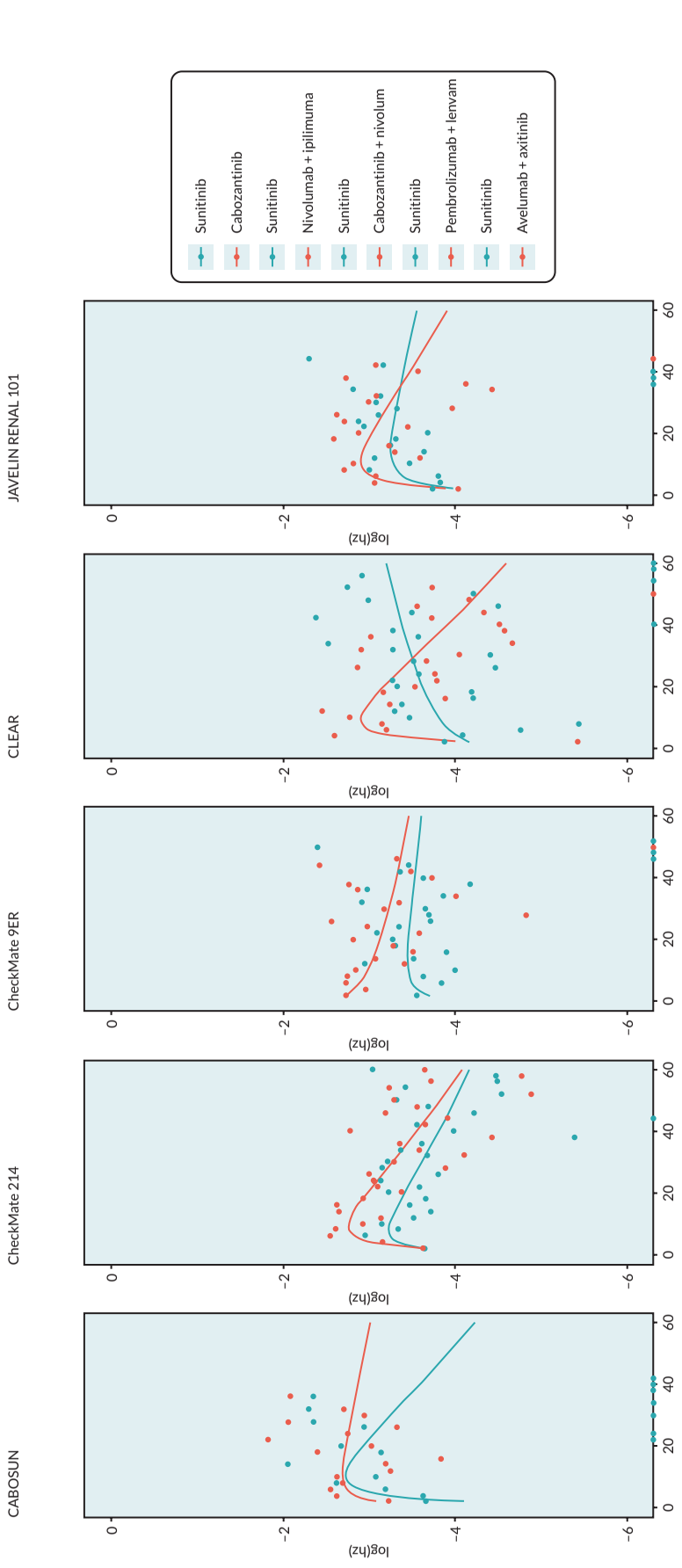


FIGURE 9 Log hazards for OS for first-line intermediate/poor risk.

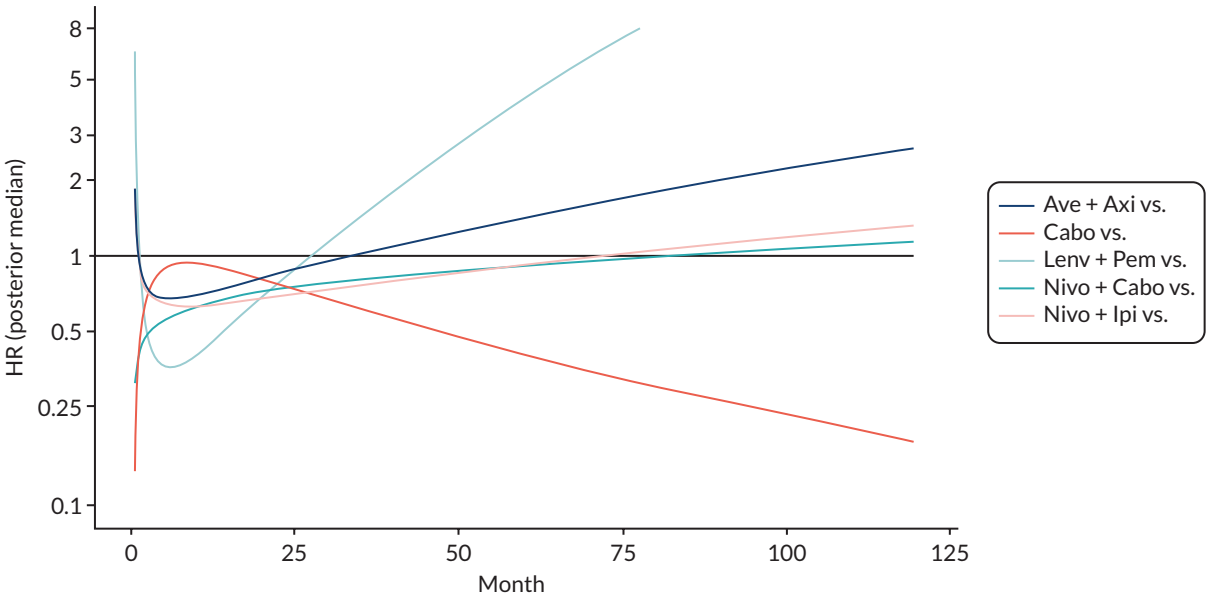


FIGURE 10 Hazard ratios and survival curves for OS for first-line intermediate/poor risk (Bayesian analysis).

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 performed sensitivity analysis of 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.

Progression-free survival in first-line intention-to-treat population

Base-case analysis

Our PH NMA of PFS in the first-line ITT population included all 10 relevant identified trials with first-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 FE analysis. Results are presented in *Table 13*, which 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 it 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 versus sunitinib enrolled only intermediate or poor-risk patients, for which the magnitude of TEs 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. Finally, even in the intermediate-/poor-risk group, the sunitinib arm of CABOSUN performed worse for both OS and PFS than in the trials of the IO combination therapies.

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 DIC for our consistency model was 18.37, with a total residual deviance of 10.40. By 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.

TABLE 13 Progression-free survival in first-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)	–

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

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Preferring investigator-assessed progression-free survival instead of blinded review progression-free survival
Where PFS was presented at the latest datacut with both investigator-assessed and BICR, we preferred blinded review-based PFS. However, two trials (CABOSUN and COMPARZ) presented PFS at last datacut assessed via both methods. We used a FE analysis and found that results were very similar to the base-case analysis (*see Report Supplementary Material 1*).

We compared our model against an unrestricted mean effects model to evaluate the presence of inconsistency. The DIC for our consistency model was 17.75, with a total residual deviance of 9.78. By 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 *Report Supplementary Material 1*. 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.

Overall survival in first-line intention-to-treat population

Our PH NMA of OS in the first-line ITT population included six relevant identified trials with first-line groups. We excluded trials testing sequences of treatments (CROSS-J-RCC, SWITCH and 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 FE 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 14*, which 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Overall response rate in first-line intention-to-treat population

Our NMA of ORR in the first-line ITT population included all 10 relevant identified trials with first-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 FE analysis. We included the whole-population estimate from TIVO-1 in order to ensure that tivozanib was represented in the network, since line-specific estimates for ORR were not available for this trial. Results are presented in *Table 15*, which 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 DIC for our consistency model was 39.53, with a total residual deviance of 21.47. By 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

TABLE 14 Overall survival in first-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)	–

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

TABLE 15 Overall response rate in first-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, 2.497)	0.81 (0.309, 2.429)	–	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 + ipi	0.434 (0.298, 0.626)	0.414 (0.276, 0.623)	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, 0.669)	0.444 (0.293, 0.668)	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, 2.063)	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, 0.399)	0.252 (0.159, 0.399)	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, 0.418)	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, 0.759)	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)	–

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

Discontinuation due to adverse events in first-line intention-to-treat population

Our NMA of discontinuation due to AEs in the first-line ITT population included all 10 relevant identified trials with first-line groups. A FE 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 RE model using a stabilising prior distribution from Turner *et al.* (2015[54]) in the form of a log-normal 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 AEs is inconsistently reported across included trials. In four trials, we extracted data from PRISMA flow charts describing discontinuations due to AEs; in another five trials, we extracted data from the text describing withdrawals or any TEAE 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 that tivozanib was represented in the network, since line-specific estimates for discontinuation were not available for this trial.

Results are presented in *Table 16*. 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Risk of treatment-emergent adverse events of grade 3 or higher in first-line intention-to-treat population

Our NMA of risk of TEAEs of grade 3 or higher in the first-line ITT population included all 10 relevant identified trials with first-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 FE analysis. We included the whole-population estimate from TIVO-1 in order to ensure that tivozanib was represented in the network, since line-specific estimates for grade 3 or higher AEs were not available for this trial. Results are presented in *Table 17*, which 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 DIC for our consistency model was 37.42, with a total residual deviance of 19.23. By 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Progression-free survival in first-line intermediate or poor-risk population

Our PH NMA of PFS in the first-line intermediate or poor-risk population included findings from nine trials (all first-line trials except for SWITCH II). We included the estimate from TIVO-1 of PFS in the intermediate- or poor-risk population spanning first- and second-line patients to ensure that tivozanib was represented in the network; otherwise, all estimates drew from first-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 FE model. Results are presented in *Table 18*, which suggested that 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

TABLE 16 Discontinuation due to AEs in first-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)	–

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

TABLE 17 Risk of AEs of grade 3 or higher in first-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)	–

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

TABLE 18 Progression-free survival in first-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)	–

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

Overall survival in first-line intermediate- or poor-risk population

Our PH NMA of OS in the first-line intermediate or poor-risk population included findings from six trials. Similar to the PH NMA of OS in the first-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 FE model. Results are presented in *Table 19*, which 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Progression-free survival in first-line favourable-risk population

Our PH NMA of PFS in the first-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 first- and second-line patients to ensure that tivozanib was represented in the network; otherwise, all estimates drew from first-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 FE model. Results are presented in *Table 20*, which 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 it was statistically superior to nivolumab + ipilimumab. Visual inspection of density plots and trace plots did not suggest non-convergence of our model.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Overall survival in first-line favourable-risk population

Our PH NMA of OS in the first-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 and SWITCH). The resultant network was star-shaped with one trial per comparison. Thus, we estimated a FE model. Results are presented in *Table 21*, which 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.

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

Overall response rate in first-line favourable-risk population

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

Density and trace plots and rank probability distributions are presented in *Report Supplementary Material 1*.

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 *Interpretation and limitations*. 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 PH was tenuous for at least one outcome in each included trial. While it is possible to interpret the HR from a model where the PH assumption has been violated as a time-average effect, it is likely preferable to use survival curves directly in indirect treatment comparisons. This

TABLE 19 Overall survival in first-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)	–

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

TABLE 20 Progression-free survival in first-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)	–

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

TABLE 21 Overall survival in first-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)	–

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

was the basis for our FP NMA. However, a competing issue that is posed by FP NMAs is the need to undertake model selection. Like all extrapolation analyses, this introduces a degree of subjectivity to the analysis, but it is likely to provide 'higher-fidelity' estimates of relative TEs.

Second, we used the most mature datacut available for each trial in all NMAs. This is a challenge for both FP 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 time points should have been used across trials for analysis, generally because more mature data (e.g. for OS) 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 the 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 the evolving shape of survival curves and trends in effect size when we jointly considered different levels of trial maturity and different treatments. In *Figures 11* and *12*, 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.

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 FP 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.

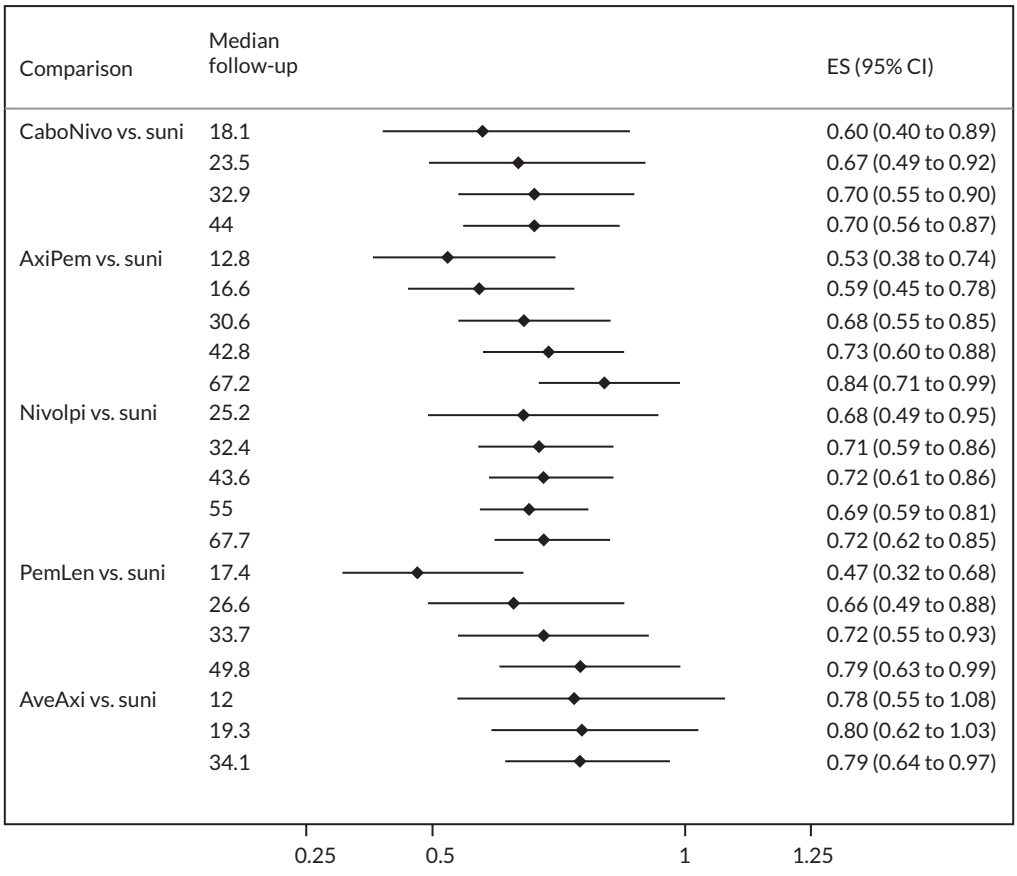


FIGURE 11 Plot of cumulative OS over sequential datacuts in key trials.

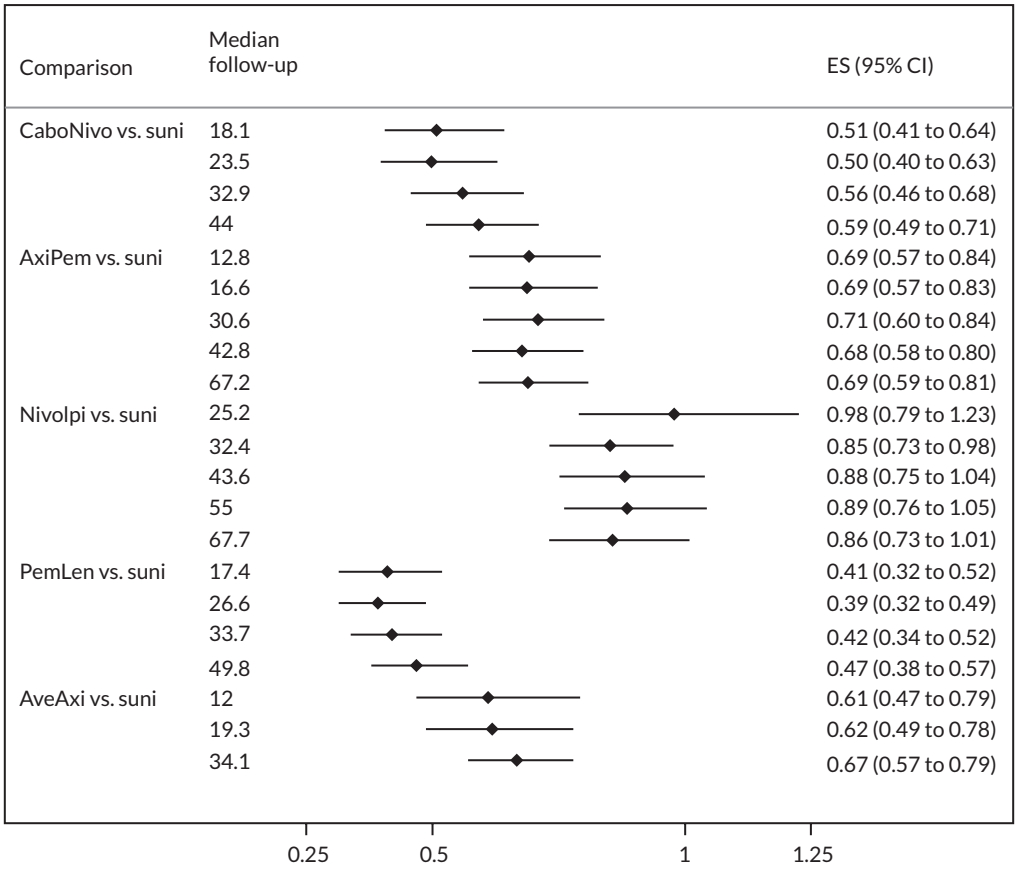


FIGURE 12 Plot of cumulative PFS over sequential datacuts in key trials.

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 first-line population. NMAs of safety outcomes should thus be regarded with some caution.

Finally, NMAs for second-line patient populations relied on a linking trial with a small sample size and documented issues with protocol administration. This means 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.

Conclusions from the External Assessment Group network meta-analyses

External Assessment Group NMAs included both FP NMAs for OS and PFS and PH NMAs for the same outcomes and NMAs for ORR and AE 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 the 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 FP NMA generated unreasonably pessimistic estimates of pembrolizumab + lenvatinib's effectiveness due to the differences in events accumulated early in the time horizon, biasing results against this treatment; by contrast, the PH NMA provide an unduly favourable estimate of effectiveness, given the convergence of hazards between treatment arms and the clear violations of the PH assumption.

Focusing on cabozantinib + nivolumab and the comparators relevant to the decision problem in each risk group, FP 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.

The 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.

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 second- or third-line context after progression on a PD-1/PD-L1 checkpoint inhibitor.[159]
- SUNNIFORECAST, which compares nivolumab + ipilimumab in combination against standard of care in a first-line context in advanced nccRCC.[160]
- A Study to Compare Treatments for a Type of Kidney Cancer Called TFE/translocation renal cell carcinoma, which compares axitinib + nivolumab against nivolumab and against axitinib in a population with multiple lines.[161]
- Cabozantinib or Sunitinib Malate in Treating Participants with Metastatic Variant Histology Renal Cell Carcinoma, comparing each treatment in a population with multiple lines.[162]
- REduced Frequency ImmuNE checkpoint inhibition in cancers (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.[163]
- A Study of Subcutaneous Nivolumab Monotherapy which is expected to complete in March 2025.[164]

Conclusions of the clinical effectiveness evidence

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 it disagreed on the full range of appropriate comparators, the relevance of TTNT 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 NMAs (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.

In relation to the External Assessment Group'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 AEs, including a high rate of AEs, leading to discontinuation in the first-line setting. In the FP 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-line-plus in the overall risk population, lenvatinib plus everolimus, nivolumab monotherapy and cabozantinib monotherapy performed best. While PH analyses suggested that IO/TKI combinations outperformed IO/IO combination (nivolumab plus ipilimumab), this was not borne out in the FP analyses.

However, despite the number of treatments available for aRCC across lines and risk groups, the EAG considered that the evidence base in aRCC 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-line-plus), most newer treatments were supported by only one trial. There was variation in some outcomes across trials, which 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 the patterns in TE 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-PH across outcomes, meaning that the results of PH NMAs are likely to be unreliable for some comparisons; at the same time, FP NMAs were highly uncertain due to similar deficiencies in the evidence base. The narrative synthesis was also conducted based on HRs that assumed PH, or on effects reported at a single follow-up time point, and therefore these findings may also be unreliable. FP 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 TE) based on the PH 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, FP NMAs produced results for pembrolizumab and lenvatinib biased in the other direction. FP NMAs were not conducted in the first-line favourable-risk population due to data limitations.

Additional outcomes were narratively synthesised, including DoR, ToT and HRQoL. 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 DoR relative to sunitinib. The findings reported for ToT 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 aRCC, 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 nccRCC, 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 TEs may vary in people with alternative histology aRCC, although the EAG does note 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.

In relation to real-world evidence

The EAG identified a number of RWE sources and completed full assessments of quality for four sources. The EAG ultimately determined that the UK RWE data set provided the most robust and relevant natural history data for use in an economic model. Median PFS data from the UK RWE were 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 TEs, 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.

Chapter 4 Cost-effectiveness model development

Structured expert elicitation

Rationale for structured expert elicitation

The maximum follow-up available within the available clinical trials identified is just over 7 years (CheckMate 025[165]). 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.[154] While this is relatively long when compared to the length of follow-up usually available within a NICE TA, this is nevertheless still short when compared to model time horizons of 40 years in the more recent published examples for first-line treatments. Given this and the fact that recent changes to the treatment pathway are expected to impact on outcomes, we conducted a structured expert elicitation exercise to inform expected long-term survival (see *Approach to elicitation*).

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 structured expert elicitation resources (STEER) repository[166,167] which was developed in line with the Medical Research Council (MRC) protocol,[166] were used to plan and conduct this exercise. The instructions for participants are detailed in *Report Supplementary Material 1*.

Expert recruitment

We initially sought to recruit a minimum of 5 and a maximum of 10 oncologists, or urologists who treat aRCC, 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 criterion 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: for example, academic versus clinical and urban versus rural populations. Experts were identified by hand-searching RCC publications and NHS websites. Recruitment was focused on substantive skills (subject area knowledge) as recommended within the MRC protocol[166] rather than normative skills (ability to accurately assess and clearly communicate their beliefs in probabilistic form). 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 being able to participate within the required time frame
- absence of specific personal and financial conflicts of interest
- published within the field of aRCC or referred by another included expert
- at least 5 years of experience in treating people with aRCC.

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 who were able to participate and the time frame 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 on 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 (three in London, two in the south-west, two in the east of England and one in the south-east of England). One expert was recruited in Scotland outside of the planned inclusion criteria following recommendation by another clinician taking part in the exercise as an expert involved in the planned NICE RCC guidelines; this was considered to be appropriate as they were already providing advice intended for use in England. The mean number of years of experience treating people with aRCC was 15 (range 5–25) and the mean number of aRCC 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 experts stated that their centre either had an academic focus, was a university hospital or a tertiary teaching hospital. Two experts stated that their population coverage included rural as well as urban geographies.

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 first-line treatment on PFS and OS and the impact on OS of different sequence lengths for subsequent treatments. The treatments included within the expert elicitation exercise per clinical are provided in *Report Supplementary Material 1*.

Based upon expert input, landmark PFS was elicited rather than OS, as this was expected to be more intuitive and avoids issues with TE 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 the key comparators for that treatment.

Approach to elicitation

Given the time frame 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 Microsoft PowerPoint® (Microsoft Corporation, Redmond, WA, USA) slides provided within the STEER tools and included background materials for each of the trials.
- Online survey that was sent to experts on 19 June 2023 for remote individual completion within 2 weeks, using the roulette method of the STEER R tool (e.g. 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.

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 time points, 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. Tabulated results are provided in *Report Supplementary Material 1*.

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 (PF) 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 aRCC 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 who 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 KM 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 were 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 the outcomes estimates. For patients receiving cabozantinib in the second line, there was a lower proportion of patients expected to be alive and PF at 3 years after receiving prior TKI monotherapy therapy (mean 14%; 95% CI 8% to 23%) than after nivolumab plus ipilimumab therapy (mean 29%; 95% CI 18% to 40%), or IO/TKI combination treatment (mean 31%; 95% CI 22% to 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 in first line, clinicians estimated that 15% more patients would be alive or PF at 3 years in the favourable-risk group (31%) compared to the intermediate-/poor-risk group (16%). By 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 PF at 3 years compared to patients receiving sunitinib in first line without a prior line of adjuvant treatment (16%). Whereas 10% fewer patients were expected to be alive and PF 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 three 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 first-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, which 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.

External Assessment Group economic analysis

Model structure

A de novo decision model was constructed for this appraisal. Adaptation of previous models, including the model used within the TA858 multiple technology appraisal (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 TAs
- methodological guidance
- the available data (type, format and coverage)
- timelines: 3 months from kick off to preliminary assessment report prior to receipt of CS, followed by 4 months to final report. This did not allow for more complex model structures to be considered.

Nature of the disease

The goal of treatment for aRCC is to extend life and delay progression, with long-term survival considered to be a reasonable goal in the context of many active agents.[168,169]

People may go through multiple 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 second-line therapy, 34% were able to receive third-line therapy, 6% were able to receive fourth-line therapy and only 1% received a fifth-line therapy.[170]

Improving HRQoL by relieving symptoms and tumour burden is also an important clinical outcome for people with aRCC.[168] Quality of life is impacted by both the stage of the disease and treatment received. Experts consulted indicated that TKI toxicities can have a considerable impact on the 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 IO treatments, immune-related AEs 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[171] 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 IO treatments, stopping rules are in place such that treatment is only given for a fixed length of time (typically 2 years).

Surrogacy between progression-free treatment, time to treatment discontinuation and time to next treatment

A targeted review was conducted to investigate the plausibility of surrogacy between different end points in aRCC. The papers identified indicated that:

- RECIST-defined ORR and PFS are not reliable surrogate end points for median OS or the TE on OS in trials of PD-L1 inhibitor monotherapy or PD-L1 inhibitor plus ipilimumab combination treatment.[172–176]
- For targeted agents, PFS is a more reliable surrogate for OS, particularly in trials which did not allow crossover after disease progression and studies published before 2005.[177,178]
- PFS may be predictive of post-progression survival (PPS) for targeted treatments at first line (a longer PFS, meaning a longer PPS[179]); PPS is then more predictive than PFS of OS.[180]
- TTNT may be a more valuable surrogate end point for previously untreated patients receiving PD-L1 inhibitor therapy.[181]
- In a real-world setting prior to the widespread 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).[182] TTD was, however, less well correlated with OS (Spearman's correlation coefficient of 0.56).

Analysis from the UK RWE data set indicated a high level of correlation between TTD and PFS end points (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. The KM (supplied in confidence and therefore not presented) demonstrated 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).

Data supplied by BMS in response to the preliminary assessment report in confidence 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 2 years for nivolumab + ipilimumab in terms of PFS while TTD continues to decrease.

The TTNT and PFS also show substantial difference for nivolumab + ipilimumab but are similar for sunitinib. Given this, adequate surrogacy may not hold for nivolumab + ipilimumab specifically. Scenario analyses are therefore presented exploring the use of TTNT as an alternative to PFS. TTNT has the benefit of not being prone to issues with 'false progression' due to tumour flare, which may potentially be experienced when considering RECIST-assessed progression with trials including ipilimumab being particularly prone to this issue. Using TTNT as a proxy for PFS is, however, also an imperfect way to estimate the effectiveness of nivolumab plus ipilimumab somewhat, as patients who are too sick to receive a new active line of treatment (i.e. patients who go on to BSC) are only coded as having an event when they die within the KM. However, given the poor surrogacy between PFS and OS for nivolumab plus ipilimumab, it provides an additional point of evidence for consideration. The truth is likely to lie between the two analyses. For reference, the HR for TTNT in CheckMate 214 is (confidential information has been removed) compared to 0.86 (0.73 to 1.01) for PFS.

The KM 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, supplied in confidence, which include TTD for both parts of the combination, do not indicate the same sort of relationship as seen for nivolumab + ipilimumab; instead, TTD and PFS appear considerably to be more similar for both arms in CheckMate 9ER.

Conceptualisation of disease model

Given the above details, if this model is conceptualised entirely using a disease-oriented approach, as recommended by technical support document (TSD; TSD13),[183] 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 that 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.

Available data

As discussed in *Objectives of the pilot process and this assessment*, all identified RCTs provided information on OS and PFS end points 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 end points except TTD by therapy type. Anonymised IPD was also provided to the EAG for 15 UK centres, including OS, PFS, ToT (first-line only), line of treatment, risk status and other population characteristics. Data from previous modelling exercises conducted within prior NICE appraisals are 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.[184,185] This can lead to PFS in the real-world appearing to be longer relative to OS than in trials.

When comparing the sunitinib arm in the UK RWE (supplied in confidence) 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.

Key issues identified within previous economic analysis

The developed model should be able to handle the following additional issues identified in prior economic analyses (see *Critique of published cost-effectiveness studies, utility studies and cost and resource use studies*):

- matching costs and effectiveness for subsequent lines of treatment
- the potential for TE 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 were not provided by Ipsen for quality of life by time to death and data from prior appraisals were redacted.

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 DES 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 Microsoft 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 trials involved in this decision problem; the ability to model non-linearity with respect to heterogeneous patient characteristics is of no additional benefit as a model linking patient characteristics to outcomes could not be produced with the data available. Note production of a simulation model may have been possible with the UK RWE data; however, as this was only received 2 weeks before production of the draft report, this was not able to be considered.

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 time frames 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.[36]

There are a limited number of examples of use of DES within prior oncology NICE TAs,[186–188] and only one of the authors is aware of where the disease area end points were OS and PFS.[186] 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 it considered that the company's model was particularly complex.[189] The evidence review group 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'.[190]

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.[183,191,192] The application of TSD13 is discussed in *Nature of the disease* and the application of TSD15 is discussed in *The need to be able to look at multiple decision points*. 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.

The TSD19 recommends that consideration is given to both theoretical and practical considerations in determining the 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 (see *Objectives of the pilot process and this assessment*), 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 datacut, for example.[193] 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 multistate modelling approach such as that defined by Williams *et al.* cannot be implemented.[194] Limited data are available to define the split between progression and death events within PFS and the data that are available do 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.

The TSD19 recommends that state transition modelling should be used alongside the PartSA approach, given the need for further methods research at the time of publication (2017).

External Assessment Group model structure

Figure 13 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 the 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 the best use of data from RWE; however, in eventuality, this was not required as the RWE information supplied to the EAG contained PFS.

The base-case model structure is a hybrid of a partitioned survival and state transition approach based upon the approach used within TA798.[195] In this approach, rather than modelling TTP and PrePS separately using a multistate modelling approach,[194] TTP and PFS data from the UK RWE (base case) and CheckMate 9ER (scenario analysis) were extrapolated, and the difference between the two was used to define pre-progression survival (Pre-PS). TEs for other treatments were applied from the NMA and it was assumed that the TE across TTP and PFS is the same. This approach was selected due to low numbers of Pre-PS events, which would be likely to make predictions from a multistate modelling approach unstable. We refer to this hybrid simply as a state transition model throughout the rest of the report.

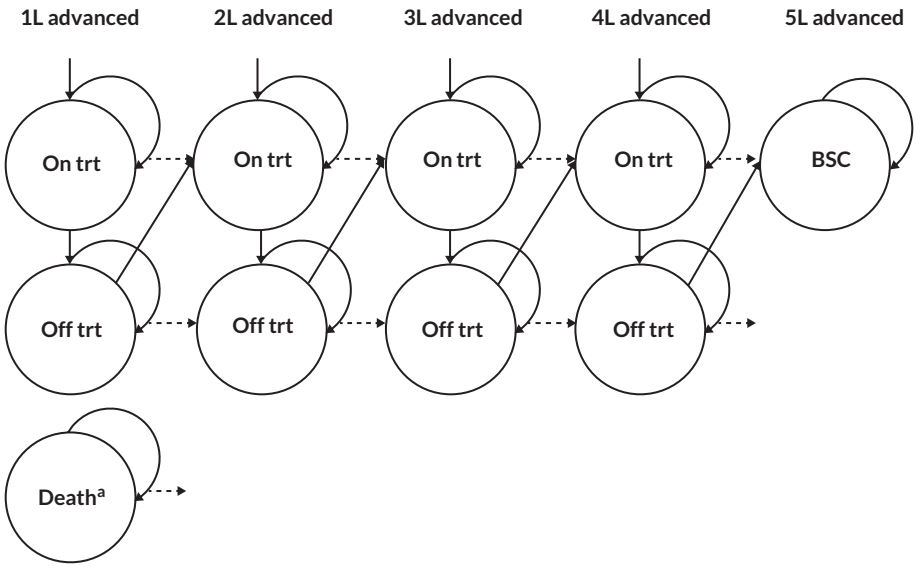


FIGURE 13 External Assessment Group model structure. Dashed line indicates the possibility to transition directly from on treatment to subsequent treatment; this does not occur in the model base case as time to treatment discontinuation is shorter than TTP, and this transition only occurs in scenario analysis where time to treatment discontinuation and PFS are set equal. a, Can be entered from any health state. 5L, fifth line; trt, treatment.

Data for ToT/TTD were also taken from the UK RWE (base case) and CheckMate 9ER (scenario analysis) and were extrapolated. PFS data were used for the relative TE 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 *Surrogacy between progression-free treatment, time to treatment discontinuation and time to next treatment*). For fixed duration treatments, the treatment duration was capped to the maximum treatment duration in the summary of product characteristics (SmPC) (base case) or included in the model using the mean number of doses received based upon the relevant trial where available (scenario analysis). RDI was taken into account in the base case.

Effectiveness data for subsequent lines following progression on first-line treatment were taken from available RWE for the reference treatment with trial data used to model relative effects based upon the NMA for other interventions. 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 second 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 for the majority of treatments of interest (see *External Assessment Group model structure*).
- 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 TE from the NMAs for PFS is applicable to TTP, Pre-PS and TTD end points; that is, there is no difference in TE for Pre-PS and progression within the PFS end point and for treatments which are given until progression; the same TE applies to TTD as to PFS.
- Patients receive subsequent treatment on progression – this is in line with how PartSA models are implemented; and it was considered as an acceptable simplification as UK RWE showed only a relatively small difference in the timing between PFS and TTNT (mean 47 days at first line).
- Transitions for first-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 first 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 third or fourth line, and the fourth-line data available from the UK RWE have 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 and 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 to be necessary as the PartSA analyses were available to cross-check against. This may be further explored in phase 2.

Model implementation

The model was implemented in R (The R Foundation for Statistical Computing, Vienna, Austria); given the complexity of the future need to evaluate large numbers of potential treatment sequences (there were 744 potential sequences of treatments that a patient could receive across the three risk-group populations based upon the 12 active treatment combinations/monotherapies that can be used in aRCC), 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.[196,197] *Report Supplementary Material 1* provides a comparison of the analytical capabilities of R and Microsoft Excel from a published example, using a side-by-side PartSA and state transition structure. The advantages to run time and analytical options clearly demonstrate for the simpler decision problem addressed by that model (only one line of treatment).

The EAG, however, note that R is less familiar than Microsoft 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 Microsoft Excel and intermediate outputs (patient flow) are provided in Microsoft Excel. In addition, NICE commissioned the decision support unit (DSU) to provide an independent external validation of the model code.

The model is intended to be made open access using GitHub® (GitHub Inc, San Francisco, USA) to improve replicability and collaboration. The model was built broadly aligning with good practice guidelines, for example, the Zorginstituut Nederland National Health Care Institute guidelines for building models in R.[198] 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.[41]The publicly available version of the decision model which was published following conclusion of the nivolumab + cabozantinib appraisal[80] uses dummy data in the correct format as inputs, where data are marked as either academic or commercial in confidence within the original data source (https://github.com/nice-digital/NICE-model-repo). The dummy data were created using the methods used to redact a Microsoft Excel model as part of a NICE submission.

Types of data which were marked as confidential and redacted to reduce the potential for back-calculation of confidential prices include:

- Patient Access Scheme (PAS) price discounts
- IPD provided by the company
- ToT input data
- relative dose intensity input data
- market share data for subsequent therapies
- reported incremental cost-effectiveness ratios (ICERs) (PAS price and list price).

A later stage of this pilot following the evaluation of cabozantinib + nivolumab was planned to involve the incorporation of a Shiny front end to the R model. Shiny is an open source R package that enables the user to build web applications using R.[199] This would allow model users to interact via an easy-to-understand user interface operating via their web browser. Unfortunately, this phase of the project was not funded by NICE.

Figure 14 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 Microsoft Excel inputs' workbook that contains data and settings for the disease model, utilities and resource use and costs
- the R output file from the FP NMA
- a Microsoft Excel output file containing the Convergence Diagnosis and Output Analysis (CODA) samples from the PH NMA
- a Microsoft Excel file containing pseudo patient-level data for the reference curves for each population, treatment, trial, line and end point for the base-case and scenario analyses
- the output from the survival analysis conducted in R (Research Design Services file; available to stakeholders for whom patient-level data access is restricted due to confidentiality).

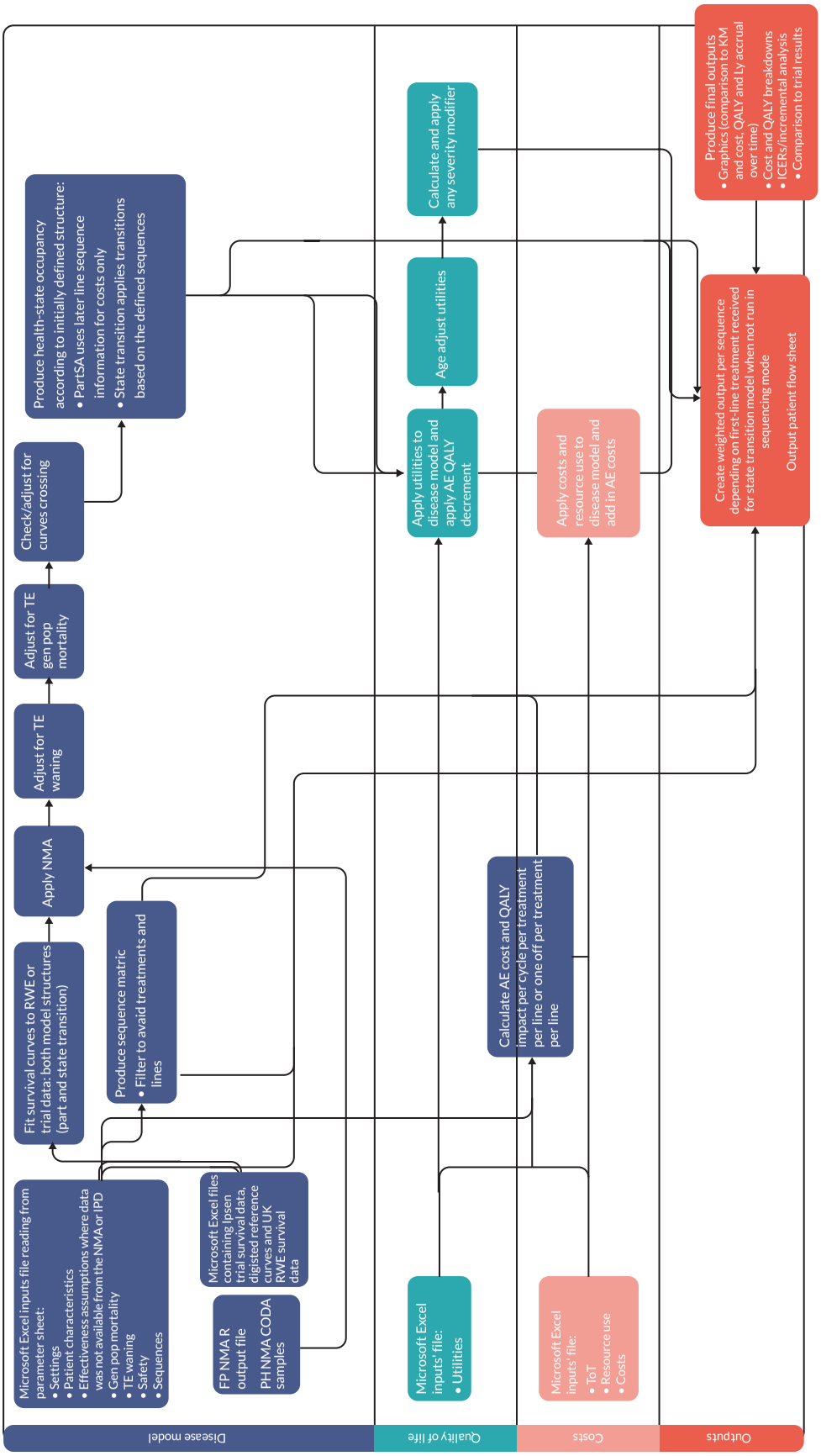


FIGURE 14 External Assessment Group model flow diagram.

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-years gained (LYG) and net monetary benefit. Base-case analyses were presented both deterministically and probabilistically, in line with the NICE manual.[41] Scenario analyses were presented deterministically.

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.

Population

The model population aligns with the decision problem population with results for the appraisal of cabozantinib + nivolumab, presented for relevant treatments for untreated aRCC or mRCC and 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 22* for how this was calculated). *Exploratory analysis looking at the impact of prior adjuvant therapy* provides details of exploratory scenario analysis that was conducted to explore the impact of this mismatch between the available clinical trial data and expected practice.

TABLE 22 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)
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%***	

SE, standard error.

Note
Scenarios for % receiving prior adjuvant treatment were calculated as the upper and lower bounds 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.

Population characteristics were taken from the UK RWE data in the base case and CheckMate 9ER in scenario analysis (see *Table 22*). Patients in the UK RWE were, on average, older and heavier than those in the CheckMate 9ER trial; other patient characteristics were broadly similar.

Treatments included

The treatments included within the decision model for the first-line setting (*Table 23*) align with those specified in the decision problem (*Table 1* and *Figure 1*).

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.

Perspective, time horizon, cycle length, discounting and price year

The model uses an NHS and PSS perspective in line with the NICE reference case.[41]

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 aRCC 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.[41] All costs were expressed in UK pounds sterling for the 2022 price year [as the latest NHS Cost Inflation Index (NHSCII) inflation index was available only until 2022 during the time this report was prepared].

Treatment effectiveness and extrapolation

Modelling of treatment effectiveness requires extrapolation of four 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, (confidential information has been removed) of patients in the nivolumab + cabozantanib arm and (confidential information has been removed) in the sunitinib arm were censored due to receipt of subsequent treatment (FDA censoring rules). EMA rules instead assume that receipt of subsequent treatment is a PFS event. TA858 demonstrated that use of EMA versus FDA censoring rules made little difference in another trial (CLEAR). Given the low proportion of patients censored due to receipt of subsequent treatment and lack of impact in prior appraisals while the use of PFS data with FDA censoring rules applied 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.
- PPS (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.

TABLE 23 Treatments included within the decision model

Treatments	1L population			Administration type and frequency	Treatment duration
	All risk	Fav risk	Poor/int risk		
Cabo + nivo[200]	x	x	x	Cabo: 40 mg orally once daily Nivo: 240 mg every 2 weeks or 480 mg every 4 weeks IV	Until disease progression or unacceptable toxicity Max 24 months for nivo
Pazo[201]	x	x	x	800 mg orally once daily	Until disease progression or unacceptable toxicity[25]
Tivo[202]	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[14]
Suni[203]	x	x	x	50 mg orally once daily, for 4 consecutive weeks, followed by a 2-week rest period	Until disease progression or unacceptable toxicity[24]
Cabo[200]			x	60 mg orally once daily	Until disease progression or unacceptable toxicity
Nivo + ipi[204]			x	Nivo: 3 mg/kg IV every 3 weeks for the first four doses Ipi: 1 mg/kg IV every 3 weeks for the first four doses Nivo maintenance: 240 mg every 2 weeks or 480 mg every 4 weeks IV, starting 3 or 6 weeks after the last dose of combination treatment, respectively	Maximum four cycles of combination treatment Monotherapy until loss of clinical benefit or unacceptable toxicity[14]
Pem + lenv[205,206]			x	Pem: 200 mg every 3 weeks of 400 mg every 6 weeks IV Lenv: 20 mg orally once daily	Until disease progression or unacceptable toxicity Max 35 3-weekly cycles for pem[14] or equivalent number of 6-weekly cycles

IV, intravenous.

Within the scenario analysis using PartSA OS, PFS and TTD required extrapolation for the reference curve at the first line of treatment only.

The reference treatment extrapolated for the first 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 first line in the UK RWE (*n* = 326). The reference treatment for second and third lines when using the UK RWE was 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 second-line-plus was everolimus, as this represented the treatment for which the most mature trial data were available (from CheckMate 025).

In line with the NICE manual[41] and discussion from other recent appraisals,[207] data for the reference treatment were 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*

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 in the original EAG report.[129] There was no indication that more flexible models were required.

Standard parametric models were therefore fitted in line with TSD 14: exponential, Weibull, log-normal, log-logistic, Gompertz, gamma and generalised gamma using the flexsurvreg package in R.

The base-case survival curve for each end point 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 end points:
  - Extrapolations where curves cross will be ruled out where possible.
  - When using the PartSA approach, the implications of selected OS and PFS curves on PPS 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 Bayesian information criterion (BIC)] – as a rule of thumb, curves with an AIC within 5 points of the best fitting curve are considered to have a similar goodness of fit.[208,209] BIC curves with a BIC difference of ≥ 6 are considered to have strong evidence of difference.
- Visual inspection.
- Statistical validity versus the NMA type to be applied (the log-normal and log-logistic curves are not consistent with the application for a FP NMA) – this issue is acknowledged but was considered to be 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), which would be 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 expect 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 data sets were identified which contained longer-term data for sunitinib than CheckMate 9ER: CheckMate 214 and KeyNote 426. These data sets were used to assess consistency with longer-term data.

Between one and three curves were selected for each end point 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 end points used within the state transition and PartSA scenarios for the reference curve for the first-line all-risk population in the model base case (sunitinib in the UK RWE). All other curve selections are presented in *Report Supplementary Material 1*.

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 24 provides a summary of where relative effectiveness has been taken from for each of treatments for each end point. 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-PH (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 (see *Treatment effectiveness waning* for further information on how treatment-effectiveness waning is addressed). For second-line and third-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 and TA645).
- Tivozanib is at best similar to pazopanib and sunitinib (TA858 and 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 and TA417).

TABLE 24 Base-case application of relative effectiveness in the economic model

	TTD	PFS	TTP	OSS
1L				
Cabo + nivo	Rel. effect = PFS	FP NMA	Rel. effect = PFS	FP NMA
Nivo + ipi	Rel. effect = PFS Rel. effect = TTNT in scenario analysis	FP NMA Rel. effect = TTNT in scenario analysis	Rel. effect = PFS Rel. effect = TTNT in scenario analysis	FP NMA
Pem + lenv	Rel. effect = PFS	PH NMA[a]	Rel. effect = PFS	PH NMA[a]
Ave + axi	Rel. effect = PFS	FP NMA	Rel. effect = PFS	PH NMA
Suni	Reference	Reference	Reference	Reference
Pazo	Equal to suni[b]	Equal to suni[c]	Equal to suni[b]	Equal to suni[+]
Tivo	Equal to suni[b]	Equal to suni[c]	Equal to suni[b]	Equal to suni[b]
Cabo	Rel. effect = PFS	FP NMA	Rel. effect = PFS	FP NMA
2L and 3L				
Nivo	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Pazo	HR to PFS	Equal to tivo[b]	Rel. effect = PFS	Equal to tivo[b]
Tivo	HR to PFS	PH NMA	Rel. effect = PFS	PH NMA
Suni	HR to PFS	Equal to tivo[b]	Rel. effect = PFS	Equal to tivo[b]
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

rel. effect; relative effectiveness.

Notes

a 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.

b Data not available in either NMA.

c PH NMA available but not used in base case.

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 HR for first line applies to TTD and TTP as discussed previously. We use the same method for TTP at second and third lines. For later lines for TTD, as data were not available in the UK RWE, we use the HR between TTD and PFS calculated at first line for all treatments:

- TTD HR to PFS: 1.19 (1.15, 1.24).

For fourth-line outcomes, we apply the HR between pooled third- and fourth-line outcomes calculated from the UK RWE to all treatments and then calculate TTP based upon its relationship to PFS at earlier lines.

- fourth-line OS HR 2.01 (1.45, 2.78)
- fourth-line PFS HR 1.74 (1.21, 2.51)
- TTP HR to PFS: 0.82 (0.80, 0.84).

Treatment effectiveness waning

Following application of NMA results, we considered the plausibility of the long-term TE predicted for each of the treatments relative to the reference treatment. The application of TE 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 (see *Report Supplementary Material 1*) and the fitted FP NMA models (see *Results of time-dependent network meta-analysis*)
- consistency between treatments with similar mechanisms of action
- precedent in prior appraisals.

Precedent was used to guide considerations. *Report Supplementary Material 1* demonstrates that within RCC, as in many other oncology indications, Committee concerns regarding uncertainty in long-term TEs in earlier submissions led to the 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 (confidential information has been removed). A similar trend is not seen for PFS.

When looking at the information available across IO/TKI combinations (see *Report Supplementary Material 1*), the longest-term data available are for pembrolizumab + axitinib (median 67.2 months), which is not recommended in England. Here, a clear trend can be seen for OS of increasing HRs (HRs getting closer to 1) with later datacuts and the OS KM 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 HRs and convergence of KMs can be seen over time for pembrolizumab + lenvatinib for which the latest datacut has a median follow-up of 49.8 months. The HRs for PFS did not demonstrate the same pattern of slippage observed for OS,[210] with changes in HR from first to last datacut being generally small: for cabozantinib + nivolumab (0.51 to 0.59), pembrolizumab + axitinib (0.69 to 0.68) 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 KM 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[211] 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 of 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, the 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 HRs 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 TE waning at 5 years to all IO/TKI combinations based on hazards, all end points. Five years was selected as the longest time point at which data are available for first-line combinations, with a reasonable number at risk remaining (at least 10% of the starting number).[212] 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 end points
- Waning applied at 10 years to all IO combinations based on hazards, all end points
- Waning applied between 5 and 20 years to all IO/TKI combinations based on hazards, all end points
- Waning applied between 5 and 20 years to all IO combinations based on hazards, all end points
- no TE waning.

These scenarios are all more optimistic than the base case due to the maturity of the available data and difficulties in 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 5 years to all IO/TKI combinations based on hazards.
- Pessimistic scenario: waning applied between 4 and 6 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 second-line and later treatments as mature data exist for CheckMate 025 (median 87.7 months) where there is no indication of convergence of the KM 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 TE waning, we retain the original hazards rather than lowering them to match the reference curve.

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 aRCC 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.

The Office for National Statistics (ONS) life tables[213] 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 to 2019 as the 2018–20 values were affected by COVID. We model mortality separately by sex, accounting for the differences in life expectancy by gender.

Report Supplementary Material 1 shows the expected general population mortality for people with age and sex profiles matching the first-line all-risk population in the UK RWE. This demonstrates that a maximum time horizon of 40 years is appropriate and also shows 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.

Adjustment for curves crossing

While 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 data set continued to receive treatment beyond progression, this was not considered appropriate.

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 that were tested with clinical experts (see appendix M of the original EAG report for more details[129]):

- Ave + axi 1L 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 in second and third lines
- Axi and evero can be used in fourth 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 *Report Supplementary Material 1*). In the base case, this was informed by the UK RWE; in scenario analysis, the use of trial data is tested.

Validation

We present the final modelled curves versus OS KM data based upon the aggregation of outcomes for each line of treatment to determine whether the model fit is appropriate. The model curve was also compared to the projections from other models previously used for NICE single technology appraisals (STAs) in the same decision point (confidential information has been removed).

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 the outcomes for cabozantinib + nivolumab even after a wait of at least a year in line with NHS criteria and is 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 time points for each treatment using information from the experts who answered the questions related to adjuvant treatment only. The average HR across the three time points for sunitinib was 0.51 and was 1.36 for cabozantinib plus nivolumab, accounting for the conditional survival format of the 5- and 10-year time points.

Exploratory analysis on the impact of prior tyrosine kinase inhibitor

Trials for second line and later often required treatment with a prior TKI (METEOR, NCT01136733, RECORD-1 and TIVO-3). Where they did not, they often had a high proportion of patients who had received prior TKI treatment (e.g. CheckMate 025). None of the trials, including second- and further line patients, were run in an era where IO combinations were available.

Based upon the responses to expert elicitation (see *Results* in *Chapter 4*), prior TKI monotherapy was expected to impact on the effectiveness of subsequent TKI monotherapy (cabozantinib was most often asked about as the most frequently used) due to similarities in the mechanism of action.

Based on fitting a basic exponential curve to the three data points available from expert elicitation, and on comparing the impact of the three types of prior treatment, there is little difference between prior nivo + ipi and IO/TKI combinations (HR 1.001). There is, however, a greater difference between prior nivo + ipi and prior TKI monotherapy (HR 1.588). An exploratory scenario analysis has been presented, including this impact. In this analysis, it was assumed that:

- The effectiveness of cabozantinib or axitinib immediately after TKI monotherapy would be impacted (these are the only TKI monotherapies allowed). Based on the UK RWE, this made up 27.2% of subsequent therapy after pazopanib, 24.5% after sunitinib, 16.7% after tivozanib and 2.9% after cabozantinib.
- The effectiveness of these treatments would be reduced. This was assumed for simplicity. In reality, it would be expected that the effectiveness of these treatments would be increased after IO combinations as the trials for these treatments included previous TKI monotherapy.

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 end point and not separately of other types of discontinuations given the data available.

Adverse event rates were taken from the 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 experiencing

a specific type of AE. This was not supplied and AEs 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 grade 3 or higher AEs which occur in > 5% of patients in any trial arm in the model. This aligns with TA858.[15] In addition, the following three AEs were included at any grade on the advice of clinical experts that these were the AEs with the most impact on the patient's quality of life and NHS resources at lower grades:

- HFS
- diarrhoea
- fatigue.

All three of these were noted as common chronic VEGF toxicities with a large impact on patients.

Reporting of specific AEs was inconsistent across the literature and producing NMAs per specific AE, given the number of interest, was not considered to be feasible; therefore, the following options are presented to capture the impact of toxicity within the model:

- Base case: NMA relative effects applied to reference treatment [sunitinib (first line) and everolimus (second-line-plus)] and trial (CheckMate 9ER[37] and CheckMate025[86]) using EAG NMA for grade 3+ AEs and all-grade NMAs from the Cochrane review[214] for the three specified grade 1 and 2 AEs, namely diarrhoea, fatigue and palmar-plantar erythrodysaesthesia syndrome.
- Scenario analysis: treatment-related naive AE rates for grade 3+ (in ≥ 5% of patients) AEs (absolute estimates) from CheckMate 9ER or comparator pivotal trials – this is the standard practice in the majority of oncology TAs.

No data were available for AEs from UK RWE for RCC specifically. One publication was identified focusing on safety outcomes for IOs, which showed that, from 2125 patient records, one-third of the patients experienced a clinically significant (grade 3+) immune-related AE.[74] RWD from Germany indicated that 32/67 (48%) of patients receiving nivolumab + cabozantinib experienced grade 3+ AEs.

The 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 were not available.

The 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 that they are equally likely to occur for the entire duration of treatment as data were 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 6 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 AEs 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.

Report Supplementary Material 1 presents the rate per patient per week for the reference treatment (sunitinib) and 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 AE. The rates were split up into IO- or TKI-induced based on the

CheckMate 9ER data (see Table 11 of the company evidence submission v2.0, dated 13 April 2023[108]) which indicated that 8 grade 3 or higher diarrhoea events were considered to be immune-mediated out of the 28 events in total, and 10 grade 1 or grade 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, they were 100% TKI-related, as mentioned in *Report Supplementary Material 1*.

Utility values

Utilities used in the model

The utility values used in the model are presented in *Table 25*.

As noted previously, the most appropriate sources identified for the base-case analyses were TA645 for patients treated at first line and TA498 for patients treated at second line. We opted to derive utilities from these NICE TAs on the basis that the utilities for first and second lines 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,[216] resulting in a PFS utility of 0.753 and a progressive disease (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 PF utility in TA498 for second-line treatment (0.69) was slightly higher than the PD utility reported in TA645 for first-line treatment (0.683), thus presenting a logical inconsistency. To mitigate this, our analysis therefore assumes that PF patients at second line will have a utility of 0.683, reflective of progressed first-line patients.

To estimate the PD utility in second-line and subsequent lines, we used the approach outlined in NICE DSU12 guidance,[215] 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; that is, second-line utility was estimated as following first-line utility in TA498/first-line utility in model * second-line utility in TA498 (0.69/0.683*0.61 = 0.616). Due to a lack of robust, published utility values for people receiving third-line treatment (or later), the same approach was used to estimate the PD utility in later lines. Overall, the decision to apply the percentage reduction in utility (in moving from PF to PD) from TA498 to estimate the utility values for PD at second, third and fourth lines was to ensure logical consistency based upon clinical feedback, that is, to ensure that patient utility decreases with disease progression.

TABLE 25 Utility values used in the model

Line of treatment	Utility	Source
1L	PF: 0.753 PD: 0.683	JAVELIN Renal 101 (TA645[57])
2L	PF: 0.683 PD: 0.616	PFS utility assumed to reflect PD in 1L. PD value estimated based on % reduction from the AXIS trial (TA498[31])
3L	PF: 0.616 PD: 0.545	Estimated based on % reduction from the AXIS trial (TA498). Approach follows NICE DSU12 guidance[215]
4L	PF: 0.545 PD:0.482	Estimated based on % reduction from the AXIS trial (TA498). Approach follows NICE DSU12 guidance[215]

The PF utility value at third line was assumed to be the same as the PD value for second-line patients, that is 0.616. To estimate the PD value at third line, we applied the percentage reduction in moving from PF to progressed in TA498, to the PF utility value in third line, which resulted in a third-line PD utility value of 0.545. The PF utility value at fourth line was assumed to be the same as the PD value for third-line patients, that is 0.545. To estimate the PD value, we again applied the percentage reduction in moving from PFS to PD in TA498, to the PFS utility value, which resulted in a fourth-line PD 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 PF and PD in second 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 the uncertainty surrounding utility values in later lines (third and fourth lines), the EAG has conducted scenario analysis, assuming a higher proportional decrease in quality of life (see below).

Due to a lack of published HRQoL data for carers and to be consistent with previous NICE appraisals for aRCC, 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)[217] and the Health Survey England (HSE) 2014 data set, as per Hernandez Alava *et al.* (2022).[218]

Disutility associated with AEs has been included in the EAG's model. These were derived from the HRQoL data collected in the CheckMate 9ER study (received by the EAG on the 9 May 2023). AEs were included as a variable in the company's mixed model repeated measures (MMRM) model, which was used to estimate the disutility associated with any grade 3 and 4 AEs. The mean disutilities associated with grade 3 and 4 AEs were provided by the company in confidence. The EAG noted that several AEs had a positive impact on patient utility, which lacked face validity, that is neutropenia and hypophosphatemia. Data were not available for specific AEs 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 AEs in the model.

The EAG noted that several specific AEs 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 AE disutilities (see *Scenario analyses conducted*).

The impact for of the three key AEs 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.

Scenario analyses conducted

Due to uncertainty surrounding health-state utilities (particularly for later treatment lines), the EAG conducted the following scenario analyses:

- First-line: use utility values from CheckMate 9ER, which reflect direct trial data.
- All lines: use CheckMate 9ER utility values for all lines; that is, CheckMate 9ER data are used for first- and second-line utility values (and no decrement is applied for third and fourth lines).
- Second-line onwards: assume the same PFS and PD utility for second, third and fourth lines; that is, 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 second line.
- Third and fourth lines: assume a higher proportional decrease in HRQoL on progression from second to third line and from third line to fourth line. This is consistent with clinical advice to the EAG. In this scenario, for third line, it will be

assumed that the decrease in HRQoL associated with moving from PFS to PD will be 10% more than that observed in second line. For the fourth line, it will be assumed that the decrease in HRQoL associated with moving from PFS to PD will be 20% more than that observed in the third line.
- Removing the impact of AEs: applied to test the impact of AEs on the ICERs, given that there is the potential for some double counting as utility data come from trials where a proportion of patients will have experienced AEs.
- Increase AE disutilities by 10%: applied to test the impact of increasing AE disutilities on the ICER. Based on clinical input to the EAG, patients are likely to experience disutility due to AEs. This analysis assumes that the impact of these disutilities increases by 10%.

A full list of scenario analyses conducted and their justification are available in *Report Supplementary Material 1*.

Resource use and costs

Disease management or health-state costs

The quantum of health-state resource use (i.e. medical oncologist outpatient consultations, CT scans and blood tests) 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,[15,27,30] which had detailed description of the HCRU with the individual components broken down and the *BMJ*- and ESMO-published RCC guidelines,[19,20] has been presented in *Report Supplementary Material 1*. 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 3 months.

The health-state costs and resource use estimates used in the model (*Table 26*) were based on NICE TA542,[27] TA858[15] and Edwards *et al.* (2018),[220] 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, there would be 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[14]), 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[15]). In addition, patients were assumed to have daily pain medication and regular specialist nurse visits in line with Edwards *et al.* (2018),[220] however, only during the last line of treatment prior to death. These assumptions were also checked with the clinical experts.

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.[152] All the previous published studies and the NICE TAs (except TA645) derived terminal care cost from this report.

The cost components of terminal care per the Nuffield Trust report have been given in *Report Supplementary Material 1*. All costs are presented from an NHS/PSS perspective and were inflated to 2022 costs using the NHSCII from PSSRU.[219] The total estimated cost of terminal care (inflated to 2022) was found to be £8714.

TABLE 26 Health-state 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 2021–2; 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),[219] 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 2021–2; 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 2021–2; 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 2021–2; 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),[219] 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 2021–2 DAPS 03 – 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 2021–2; HRG code WF01A, clinical oncology – non-admitted face-to-face attendance, follow-up
	Specialist nurse visit	0.25	£53	Frequency: based on Edwards *et al.* (2018) but assumed to be twice as frequent as consultant follow-up Unit cost: PSSRU (2022),[219] 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 *et al.* (2018) Unit cost: BNF; 50 mg/50 ml vial morphine sulphate solution for infusion

BNF, *British National Formulary*; HRG, Health Resource Group.

Drug and administration costs

A summary of acquisition costs of the treatments considered in the first-line setting and their respective dosing schedules (as provided in detail in *Treatments included*) along with the treatments in subsequent lines has been presented in *Table 27*. Please note that the unit costs for each drug were extracted from either the 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 intravenous (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, so steps are taken to minimise wastage; and 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 the packs used.

TABLE 27 Acquisition costs of treatments considered in the economic model

Treatment	Formulations	Size of pack	Dose per unit	Pack price (list price)[221,222]
Ave	Bavencio® 200-mg/10-ml infusion vials	1 vial	20 mg per ml	£768
Axi	Inlyta® 5-mg tablets	56 tablets	5 mg	£3517
Cabo	Cabometyx® 40 mg	30 tablets	20, 40 and 60 mg	£5143
Evero	Evero 10-mg tablets (generic)	30 tablets	10 mg	£373.48
Ipi	Yervoy® 50-mg/10-ml infusion vials	1 vial	5 mg per ml	£3750
Lenv	Lenvima® 10-mg capsules	30 capsules	2 mg, 4 mg, 10 mg	£1437
Nivo	Opdivo® 100-mg/10-ml infusion vials	1 vial	10 mg per ml	£1097
	Opdivo® 40-mg/4-ml infusion vials	1 vial	10 mg per ml	£439
Pazo	Votrient® 400-mg tablets	30 tablets	400 mg	£1121
Pem	Keytruda® 100-mg/4-ml infusion vials	1 vial	25 mg per ml	£2630
Suni	Suni 50-mg capsules (generic)	28 capsules	50 mg	£1388.77
Tivo	Fotivda® 1340-µg capsules	21 capsules	1.34 mg	£2052

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.

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 TAs, as provided in *Report Supplementary Material 1*. RWE data were 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.

The EAG notes that the RDI data available are inconsistent in how it was calculated and that there may be an underestimate of the RDI for some of the treatments. It is also not always possible to recoup the full cost of a drug when patients receive a lower dose or miss doses. The EAG therefore presents a scenario analysis where all RDIs are set to 100%, given the inconsistency in the methods used within the available RDIs.

There are two treatments where the price does not vary linearly with the number of mg prescribed: lenvatinib and cabozantinib. The dosing of lenvatinib is further complicated by the difference in titration practice between the product's summary of product characteristics (SPC) and what is frequently done within the NHS and the potential impact of this on dosing of pembrolizumab when it is given in combination. This is a particular issue for lenvatinib as it is given at its maximum possible dose when used in combination. Cabozantinib, on the other hand, is given at a lower dose than the maximum possible.

As part of technical engagement, the EAG consulted two clinicians (Dr Larkin and Dr Challapalli) and National Health Service, England (NHSE) regarding the issue of lenvatinib dosing and how this dosing interacts with administration of pembrolizumab. Both clinicians acknowledged the toxicity issues associated with lenvatinib when given in combination treatment, using the starting dose of 20 mg from the SPC (this is the maximum possible dose and often not tolerated). Both noted that due to this many clinicians instead titrate patients up to as close to 20 mg as possible, often starting at 10 mg and titrating up in 4-mg steps every 2 weeks (pills come in 4-mg and 10-mg sizes). NHSE added that clinical practice is varied in that some clinicians titrate up to 20 mg and others work downwards. Regardless of whether off-label titration is done or the SPC dose is used, dose adjustments are performed as a part of an oncologist's face-to-face appointment or, more frequently, via a short phone call, rather than at an additional scheduled appointment for pembrolizumab administration. The optimal dose of lenvatinib is usually achieved within the first 2–3 months. Both clinicians consulted considered that doses of either 10 mg, 14 mg or 20 mg are given in the long term, which aligns with the CLEAR trial protocol. NHSE considered that some clinicians also use the 18-mg dose. The resource use in the model already accounts for an oncologist consultation every 4 weeks. For some patients, an additional consultation at 2 and 6 weeks may be required (maximum additional cost of £328).

Because lenvatinib is priced the same for a 4-mg tablet as a 10-mg tablet, UK titration practices may result in increased costs that are not captured in the model. In order to more accurately capture the dosing of lenvatinib, the following approach has been used in the updated EAG base case:

- All patients are assumed to receive 10 mg for the first 2 weeks.
- 75% of patients are assumed to receive 14 mg for the next 2 weeks (based upon TA858 assumption that 25% of patients cannot tolerate > 10 mg, which was confirmed as reasonable by Dr Larkin).
- 18% of patients are assumed to receive 18 mg for 2 weeks and then 20 mg for 2 weeks based upon the mean RDI of 70.5% reported in the trial, and 10, 14 and 20 mg being the relevant long-term doses. This was confirmed as reasonable by Dr Larkin.
- Patients are assumed to receive 0.429 × 4 mg and 1.196 × 10 mg pills after the first 8 weeks based upon the company response to clarification questions; table 3 in TA858.

Scenario analysis is also presented using NHSE input on the long-term doses used in practice: 25% at 10 mg, 40% at 14 mg, 20% at 18 mg and 15% at 20 mg. These are broadly consistent with the above and result in a slightly higher RDI of 73.5% (not accounting for any missed doses).

Both consulted clinicians considered that patients would be unlikely to receive every 3-week dosing of pembrolizumab as a part of the protocol to address required dosing adjustments for lenvatinib.

In addition to the impact on cost, there are patient-related issues to be considered. Dr Challapalli noted that the issues with toxicity of lenvatinib are a significant concern for patients, who may worry that a lower dose might result in reduced effectiveness or try to be 'brave' and therefore not report toxicity as early as would be ideal to manage dosing. These issues are more pronounced than for other IO/TKI combinations. As noted earlier, this is because lenvatinib, unlike the other TKIs, is used at the maximum possible starting dose.

Finally, the EAG considered how to handle the dosing of lenvatinib within lenvatinib + everolimus. During this consideration, it was noted that the maximum modelled dose had been 20 mg rather than the 18 mg in the SPC. This model was amended to use 18 mg (one 10-mg tablet and 2 × 4-mg tablets). Again, it was assumed that 25% of patients would receive 10 mg in the long term in line with the dosing within lenvatinib + pembrolizumab. Given the reported RDI of 70.4%, this resulted in an estimate of 48% of patients receiving the 14-mg dose and 27% receiving the 18-mg dose long term.

Different administration costs were used for different drugs, depending on the route of administration and whether or not the drug is administered jointly based on NICE TA858/TA645 (*Table 28*). Unit costs were extracted from NHS reference costs 2021–2.[219]

Adverse event costs

Adverse event 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 *Adverse events*).

Report Supplementary Material 1 presents the costs per event of all the AEs considered as per the two options/data sources mentioned in *Adverse events*, incorporating the clinical opinion to EAG, in line with NICE TA858[15] and the unit costs derived from NHS reference costs 2021–2.[223]

Report Supplementary Material 1 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 of the original EAG report.[129] The disutilities associated with the AEs considered have been provided and described in *Utilities used in the model*. These data were 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 AEs to the same as tivozanib. Dr Larkin also

TABLE 28 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 2021–2; HRG code: SB12Z
Ipi (for first four cycles when nivo is delivered jointly with ipi)	Complex chemotherapy, including prolonged infusional treatment, at first attendance – outpatient	£440.71	NHS reference costs 2021–2; 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 2021–2; HRG code: SB11Z – deliver exclusively oral chemotherapy. Subsequent cycles: PSSRU 2022. Pharmacist time based on NICE TA645

DAPS, direct access pathology services; HRG, Healthcare Resource Group.

Note
2020–1 costs were inflated to 2022 using NHSCII annual % increase on previous year index (2.72%) from PSSRU 2022.[219]

noted that he would have expected similar treatments to be more closely grouped together. Similar AE profiles would be expected for TKI monotherapy and more AEs than monotherapy would be expected for lenvatinib + everolimus. Similar AE profiles would be expected for the IO + TKIs, with nivolumab monotherapy and nivolumab + ipilimumab expected to be different to IO + TKIs. This does appear to be the case when looking at the total cost of managing AEs and QALY impact, but this sensible grouping is not seen when looking at per-cycle impacts due to differences in the predicted TTD. The majority of AEs would be expected to occur relatively soon after initiating treatment, 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.

Subsequent treatment costs

Given different pathways are possible, following and conditional upon first-line treatments received in the 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 (as presented in *Table 29*) 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 the progression based on the mean duration of subsequent treatment.

Two relevant data sources were considered for calculating the subsequent treatment costs:

- costs based on subsequent treatments as observed in RWE (see *Critique of real-world evidence identified for this appraisal*)
- costs based on subsequent treatments from CheckMate 9ER or other relevant comparator pivotal trials (appendix N of the original EAG report[129]).

TABLE 29 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[a]	39,268.59	795.54
	Ave + axi	39,608.96	703.34
	Pazo	54,145.22	4320.70
	Tivo	56,145.46	5129.54
	Suni	53,124.52	4412.78
Int/poor risk	Cabo + nivo[a]	39,268.59	795.54
	Nivo + ipi	34,822.10	684.62
	Pem + lenv	35,686.51	663.48
	Ave + axi	39,608.96	703.34
	Pazo	54,145.22	4320.70
	Tivo	56,145.46	5129.54
	Suni	53,124.52	4412.78
	Cabo	50,797.53	5888.86

fav, favourable; PartSA, parttioned survival analysis; prop, proportion; txt, treatment.
a 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.

The UK RWE is used for subsequent systemic therapies in the model base case (see *Report Supplementary Material 1*) to better reflect clinical practice, and 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 (see *Figure 1*), were not included. It is to be noted that the subsequent radiotherapy and surgery costs were also considered (as given in *Report Supplementary Material 1*) following progression and were added as a one-off cost, with frequencies based on data from CheckMate 9ER as data were 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 *Drug and administration costs*.

Assumptions common to both RWE and trial:

- The type of subsequent treatment was assumed to be independent of the first-line risk group and was only dependent on the prior treatments received. Analysis of RWE stratifying the contingency table of treatment types at first and second lines (excluding the 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 the 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 the UK clinical practice for all subsequent lines.
- Subsequent treatments after pazopanib and sunitinib were assumed to be the same as tivozanib for third line as data were 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 second-line systemic treatment (i.e. rescaled to 100%).
- Where data were not available for the duration of subsequent treatments from one source, then data from the alternative source was used (i.e. 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.

Severity

The NICE manual is unclear as to how current practice should be defined in a multicomparator decision space such as is present here for calculation of the severity modifier. There are three clear options to define the 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.

None of the options are fully consistent with the principle of fully incremental analysis. Therefore, for a pragmatic solution, in the EAG base case, absolute and proportional shortfalls are calculated using a common reference treatment for the overall population, and each risk subgroup with QALY weightings are assigned based upon NICE's severity modifiers (*Table 30*). 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 considers this to represent the current best practice in the absence of formal NICE guidance. Pairwise analyses are presented in addition. The EAG notes, however, that pairwise analyses are generally best avoided, as excluding relevant comparators from an incremental analysis can lead to serious errors in interpretation (e.g. by leading to comparisons of interventions that are not on the efficient frontier).

TABLE 30 Quality-adjusted life-year weightings for severity

QALY weight	Proportional QALY shortfall	Absolute QALY shortfall
1	< 0.85	< 12
×1.2	0.85–0.95	12–18
×1.7	At least 0.95	At least 18

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 (2018–20)[213] were used to calculate the future life expectancy for the general population and the HSE 2014 data set was used to calculate the future quality of life for the general population.[218] 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 the absolute and proportional QALY shortfall amounts and the relevant QALY modifier to apply.

The EAG has applied the severity modifier, in line with prior precedent, on a deterministic basis.

Uncertainty

Base-case analyses were presented both deterministically and probabilistically, in line with the NICE manual.[41] Additional scenario analyses were conducted where they added value and clarity. Scenario analyses were produced deterministically due to the large number of scenarios required and the run time associated with these.

State transition model predicted outcomes for PFS and OS as shown in *Report Supplementary Material 1*.

To reduce the run time of the state transition model to make PSA feasible, all transitions from second line onwards were approximated to an exponential curve to remove the need for use of tunnel states. The calculation for these works as follows:

- Calculate the area under the curve (AUC) using the fitted curve.
- Calculate the lambda for the exponential curve as 1/AUC.

This reduces the run time from upwards of 90 minutes to 3 minutes.

Report Supplementary Material 1 shows the difference in results between the full and reduced model without tunnel states, using the deterministic list price results and the version at the time of factual accuracy checking. The total predicted costs are within 4% for all treatments, with the costs being within 1–2% for most treatments across the two models. Predicted QALYs show more deviation for TKI monotherapies than for novel therapies (6–10%) when compared to 2–4% for cabozantinib + nivolumab and pembrolizumab + lenvatinib. This is due to loss of precision in the LYs predicted in second line, which impact more upon TKI monotherapies where patients reach second line sooner. The predicted ICERs are, however, of a similar magnitude. Therefore, the reduced model results were considered sufficiently similar for the model to be run without tunnel states to examine the level of uncertainty in the results within the PSA.

The distributions used within PSA are summarised in *Report Supplementary Material 1*.

Chapter 5 Cost-effectiveness results

Cost-effectiveness results

Base case

Table 31 provides the EAG base-case list price and base-case results, both as a fully incremental analysis and as a pairwise analysis. The results presented align with those discussed at the first Appraisal Committee meeting for this topic and are deterministic, as previous probabilistic analysis – using the lambda approximation method to reduce the run speed – showed consistent results with the deterministic analysis using the lambda approximation method (see appendix R of the original EAG report[129]).

As would be expected, the life-years (LYs) and QALYs for the three TKI monotherapies are similar (these are set to have the same first-line effectiveness in the model base case). The results differ slightly as the types of second-line therapies used differ across the treatments, in line with the UK RWE, and the AE impacts also differ across treatments. In all-risk groups, tivozanib was the least effective of the three TKI monotherapies. Sunitinib was the most effective.

Most of the time spent in state for all treatments is still in first and second lines. For example, in the all-risk population, 83% of time in state is spent in first and second lines for cabozantinib + nivolumab and 69% is spent in first and second lines for pazopanib, with 17% spent in third line and 12% spent in BSC.

Cabozantinib + nivolumab is not cost-effective at list price in the all-risk and favourable-risk populations.

In the intermediate-/poor-risk population, at list price, cabozantinib + nivolumab is dominated by cabozantinib monotherapy. This is driven by the unexpectedly good performance of cabozantinib observed relative to sunitinib in the CABOSUN trial. Neither pembrolizumab + lenvatinib nor nivolumab + ipilimumab are cost-effective in comparison to cabozantinib monotherapy and other TKIs, which aligns with the conclusion of TA858. Sunitinib monotherapy is the most cost-effective treatment at list price when considering a £30,000 per QALY threshold.

When comparing to the two other novel combinations, cabozantinib + nivolumab is less effective and less expensive than pembrolizumab + lenvatinib (south-west quadrant ICER of £110,498). This is driven by two things. First, the higher effectiveness of pembrolizumab + lenvatinib predicted from the PH NMA [HR = 0.767 (0.562 to 1.049) vs. cabozantinib + nivolumab]. Second, the increased cost associated with reduced doses of pembrolizumab + lenvatinib relative to cabozantinib + nivolumab due to lenvatinib pills being priced at the same cost rather than reduced linearly with the reduced dosing. The EAG acknowledges that, due to redaction of the PFS KM for pembrolizumab + lenvatinib, the EAG analysis had to use the PH NMA for this treatment, which likely biases towards pembrolizumab + lenvatinib.

In the intermediate-/poor-risk population, qualification for the severity modifier remains dependent on which treatment is considered representative of current practice. A modifier of 1.2 applies versus sunitinib, pazopanib and tivozanib, but not the other more recent treatment options.

In the all- and favourable-risk populations, the severity modifier does not apply regardless of the comparator. As within the previous report, the QALY shortfall-related modifier has not been directly incorporated, given the uncertainty around which, if any, modifier to apply. A modifier of 1.2 equates to a willingness to pay threshold of £24,000–36,000, using the standard NICE thresholds.

Report Supplementary Material 1 presents the detailed breakdown for the PartSA results using the EAG base-case settings at list price. The three novel therapies have relatively similar predicted QALY gains in the base case (1.86 for nivo + ipi, 1.91 for cabo + nivo and 1.96 for pem + lenv). The results are similar to the previous EAG base case (the only minor amendment being in the QALYs associated with AEs for subsequent treatments).

TABLE 31 Updated EAG base case (list price)

Technologies	Costs (£)	LYG	QALYs	Inc. costs (£)	Inc. LYG	Inc. QALYs	ICER cabo + nivo vs. comparator	ICER incremental	Severity modifier
Risk population: all-risk									
Suni	77,675	2.78	1.67	–	–	–	£263,297	–	1
Pazo	78,649	2.84	1.69	974	0.06	0.03	£275,106	£35,580	1
Tivo	98,517	2.76	1.66				£223,701	(dominated)	1
Cab + nivo	223,847	3.71	2.22	145,198	0.88	0.53		£275,106	–
Risk population: favourable risk									
Suni	83,420	3.68	2.20	–	–	–	£358,676	–	1
Pazo	84,321	3.73	2.23	900	0.06	0.03	£379,222	£32,471	1
Tivo	115,279	3.66	2.19				£287,383	(dominated)	1
Cabo + nivo	251,276	4.52	2.67	166,955	0.78	0.44		£379,222	–
Risk population: Intermediate/poor risk									
Suni	75,069	2.45	1.46	–	–	–	£237,872	–	1.2
Pazo	76,064	2.50	1.49	995	0.05	0.03	£248,380	£36,780	1.2
Tivo	91,528	2.43	1.45				£205,798	(dominated)	1.2
Nivo + ipi	137,774	2.44	1.46				£123,562	(dominated)	1
Cabo	158,308	3.46	2.07	82,243	0.96	0.59	Cabo + nivo dominated	£140,523	1
Cabo + nivo	204,721	3.36	2.00					(dominated)	–
Pem + lenv	229,649	3.62	2.23	71,341	0.15	0.16	SW quadrant £110,498	£450,638	1

SW, south-west.

Note
Cost-effectiveness results are presented by first-line treatment weighting each possible follow-on sequence by the percentage of patients expected to receive that sequence as presented in *Report Supplementary Material 1*.

Scenario analysis

Table 32 presents the scenario analysis for each of the risk populations. Results in the all-risk and favourable-risk populations are broadly consistent with the previous EAG analysis; that is, cabozantinib + nivolumab is not cost-effective at list price compared to TKI monotherapies, and when the PartSA model is used, the combination is less effective than TKI monotherapies in the favourable-risk population (due to the OS HR in CheckMate 9ER being > 1).

Notable results include:

- Nivo + ipi dominates nivo + cabo in the intermediate-/poor-risk population when trial data are used in the PartSA model.
- When the PH NMA is used within the state transition structure, the most effective treatment in the intermediate-/poor-risk population is pem + lenv (2.23 QALYs), followed by cabo + nivo (2.16 QALYs) and then followed by nivo + ipi (1.82 QALYs).
- When the PH NMA is used within the PartSA structure, the most effective treatment in the intermediate-/poor-risk population is cabo + nivo (2.17 QALYs), followed by nivo + ipi (2.09 QALYs) and then pem + lenv (1.96 QALYs).
- When TTNT is used instead of PFS from CheckMate 214 within the FP NMA, nivo + ipi remains predicted to be of lower effectiveness than cabo + nivo. This is due to the HR predicted being higher in the first year (see *Table 32*) during which time a large number of events have already occurred within the sunitinib RWE reference curve.
- If all RDIs are set to 100%, the costs associated with cabo + nivo substantially increase and, at list price, the combination is dominated by pem + lenv.

The difference in ordering of the treatments when the PH NMA is used across the two different structures should be interpreted with the following caveats:

- The base-case state transition structure likely underestimates the effectiveness of nivo + ipi due to poor surrogacy between PFS and OS.
- The PH NMA likely overestimates the effectiveness of both IO + TKI combinations as it does not account for slippage in the HRs seen in the data. This is not fully mitigated by assumptions applied for TE waning as hazards are expected to cross in the long term between IO + TKI combinations and TKI monotherapy.
- The FP NMA results for pem + lenv are not considered to be reliable due to a combination of two reasons. First, the redaction of KM data in TA858, meaning that ITT data had to be used. Second, the lack of events in the placebo arm in the initial part of the CLEAR trial (both PFS and OS) makes it difficult for the FP method to produce a plausible output. For the reasons noted in the bullet point mentioned above, the base case (using the FP NMA for all other treatments and the PH NMA for pem + lenv) is likely to bias in favour of pem + lenv.

The EAG base case includes RDIs provided by the company for cabozantinib, which are likely to underestimate the cost of this combination as, although the EAG model costs treatment per pack rather than per pill, the information presented assumes that all patients come off treatment in the CheckMate 9ER trial due to either progression or unacceptable toxicity. This is not the case, as some patients were observed to discontinue for other reasons (e.g. participant request or participant withdrawing consent).

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis was conducted at the time of factual accuracy checking (*Table 33*). Conclusions from the probabilistic and deterministic analyses were identical. The CIs around the incremental QALYs were wide. In the case of cabozantinib + nivolumab in the all-risk and favourable-risk populations relative to pazopanib, the CIs crossed 0, demonstrating a high level of uncertainty. There was more certainty in the intermediate-/poor-risk population. Here, the 95% CI for incremental QALYs for the IO combination treatments did not overlap with sunitinib, pazopanib or tivozanib, other than for nivolumab + ipilimumab, where results should be viewed with caution due to the issues with poor surrogacy of PFS for OS.

Cost-effectiveness frontiers at list price are shown in *Report Supplementary Material 1* for all-risk, favourable-risk and intermediate-/poor-risk populations. In all three risk groups, pazopanib and sunitinib lie close together on the fronter. In the all-risk and favourable-risk groups, the only novel therapy included (and therefore the only treatment lying along the

TABLE 32 Scenario analyses

Parameter	Base case		Scenario	Next best comparator[a]	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
All risk							
Base case				Pazo	145,198	0.53	£275,106
Model structure							
Overall structure	State transition, four lines	1	PartSA, four lines	Suni	142,265	0.32	£445,511
		3	State transition, two lines	Pazo	159,026	0.69	£228,912
Primary data source							
Data source for baseline risk and patient characteristics	UK RWE, state transition model	6	Trial-based analyses, state transition model	Suni	153,199	0.43	£355,214
	UK RWE, state transition model	7	Trial-based analyses, PartSA	Pazo	148,612	0.29	£519,752
Effectiveness							
Preferred first-line NMA	FP NMA	11	PH NMA	Pazo	150,768	0.66	£229,908
Preferred NMA	FP NMA first line, PH NMA second line	21	PH NMA throughout, PartSA	Suni	148,284	0.54	£277,106
Preferred NMA for pem + lenv	PH NMA	13	FP NMA	Pazo	145,198	0.53	£275,106
Surrogate outcome for nivo + ipi	PFS	73	Using TTNT data as a proxy for PFS for nivo + ipi	Pazo	145,198	0.53	£275,106
Surrogate outcome for nivo + ipi	PFS	74	Using TTNT data as a proxy for PFS for nivo + ipi, PH NMA	Pazo	150,768	0.66	£229,908
TTD data source	TTD	18	TTD equal to PFS	Pazo	149,924	0.52	£290,923
	Relative effectiveness for nivo + ipi from PFS consistent with other treatments	20	Relative effectiveness for nivo + ipi from simple HR between PFS and TTD from CheckMate 214	Pazo	145,198	0.53	£275,106
Treatment effectiveness waning	5 years for IO/TKIs, all end points, based on hazards	24	Between 5 and 20 years all IO/TKIs, all end points, based on hazards	Pazo	144,690	0.52	£278,645
		26	No TE waning	Pazo	144,630	0.52	£279,065
Suni RWE 1L PFS	Log-logistic	29	Weibull	Pazo	139,299	0.40	£345,056
Impact of prior TKI treatment	Not considered	76	Exploratory analysis HR1.59 applied to TKI after TKI monotherapy	Pazo	129,002	0.61	£211,852
RDI	Applied	41	All RDI set to 100%	Pazo	178,604	0.53	£338,401

TABLE 32 Scenario analyses (*continued*)

Parameter	Base case		Scenario	Next best comparator[a]	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
Lenv dosing within pem + lenv	TA858 and RDI data	75	NHSE input	Pazo	145,198	0.53	£275,106
Data source used for utilities	JAVELIN Renal 101 for 1L, AXIS trial for 2L and assumed same proportional decrease for 3L and 4L	50	CheckMate 9ER for all lines	Pazo	145,198	0.55	£264,436
Data source used for AEs	NMA	58	Individual trials	Pazo	144,383	0.55	£263,634
Favourable risk							
Base case				Pazo	166,955	0.44	£379,222
Overall structure	State transition, four lines	1	PartSA, four lines	Tivo	130,044	−0.21	Cabo + nivo dominated
		3	State transition, two lines	Pazo	181,255	0.61	£296,395
Data source for baseline risk and patient characteristics	UK RWE, state transition model	6	Trial-based analyses, state transition model	Suni	177,707	0.32	£564,209
	UK RWE, state transition model	7	Trial-based analyses, PartSA	Tivo	138,615	−0.24	Cabo + nivo dominated
Preferred first-line NMA	FP NMA	11	PH NMA	Pazo	166,955	0.44	£379,222
Preferred NMA	FP NMA first-line, PH NMA second-line	21	PH NMA throughout, PartSA	Tivo	130,044	−0.21	Cabo + nivo dominated
Preferred NMA for pem + lenv	PH NMA	13	FP NMA	Pazo	166,955	0.44	£379,222
Surrogate outcome for nivo + ipi	PFS	73	Using TTNT data as a proxy for PFS for nivo + ipi	Pazo	166,955	0.44	£379,222
Surrogate outcome for nivo + ipi	PFS	74	Using TTNT data as a proxy for PFS for nivo + ipi, PH NMA	Pazo	166,955	0.44	£379,222
TTD data source	TTD	18	TTD equal to PFS	Pazo	175,480	0.43	£408,325
	Relative effectiveness for nivo + ipi from PFS consistent with other treatments	20	Relative effectiveness for nivo + ipi from simple HR between PFS and TTD from CheckMate 214	Pazo	166,955	0.44	£379,222

continued

TABLE 32 Scenario analyses (*continued*)

Parameter	Base case		Scenario	Next best comparator[a]	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
Treatment effectiveness waning	5 years for IO/TKIs, all end points, based on hazards	24	Between 5 and 20 years all IO/TKIs, all end points, based on hazards	Pazo	166,955	0.44	£379,222
		26	No TE waning	Pazo	166,955	0.44	£379,222
Suni RWE 1L PFS	Log-logistic	29	Weibull	Pazo	166,961	0.44	£378,766
Impact of prior TKI treatment	Not considered	76	Exploratory analysis, HR 1.59 applied to TKI after TKI monotherapy	Pazo	139,731	0.52	£267,397
RDI	Applied	41	All RDI set to 100%	Pazo	209,776	0.44	£476,487
Lenv dosing within pem + lenv	TA858 and RDI data	75	NHSE input	Pazo	166,955	0.44	£379,222
Data source used for utilities	JAVELIN Renal 101 for 1L, AXIS trial for 2L and assumed same proportional decrease for 3L and 4L	50	CheckMate 9ER for all lines	Pazo	166,955	0.46	£366,224
Data source used for AEs	NMA	58	Individual trials	Pazo	166,148	0.46	£361,257
Intermediate/poor risk							
Base case				Cabo	46,413	−0.07	Cabo + nivo dominated
Model structure							
Overall structure	State transition, four lines	1	PartSA, four lines	Nivo + ipi	63,872	0.05	Cabo + nivo extendedly dominated
		3	State transition, two lines	Cabo	63,610	0.18	Cabo + nivo extendedly dominated
Primary data source							
Data source for baseline risk and patient characteristics	UK RWE, state transition model	6	Trial-based analyses, state transition model	Cabo	71,506	−0.03	Cabo + nivo dominated
	UK RWE, state transition model	7	Trial-based analyses, PartSA	Pem + lenv	−43,677	−0.76	Cabo + nivo extendedly dominated

TABLE 32 Scenario analyses (*continued*)

Parameter	Base case		Scenario	Next best comparator[a]	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
Effectiveness							
Preferred first-line NMA	FP NMA	11	PH NMA	Cabo	50,112	0.02	Cabo + nivo extendedly dominated
Preferred NMA	FP NMA first-line, PH NMA second-line	21	PH NMA throughout, PartSA	Nivo + ipi	46,097	0.08	£549,457
Preferred NMA for pem + lenv	PH NMA	13	FP NMA	Cabo	46,413	−0.07	Cabo + nivo dominated
Surrogate outcome for nivo + ipi	PFS	73	Using TTNT data as a proxy for PFS for nivo + ipi	Cabo	46,413	−0.07	Cabo + nivo dominated
Surrogate outcome for nivo + ipi	PFS	74	Using TTNT data as a proxy for PFS for nivo + ipi, PH NMA	Cabo	50,112	0.02	Cabo + nivo extendedly dominated
TTD data source	TTD	18	TTD equal to PFS	Cabo	45,547	−0.08	Cabo + nivo dominated
	Relative effectiveness for nivo + ipi from PFS consistent with other treatments	20	Relative effectiveness for nivo + ipi from simple HR between PFS and TTD from CheckMate 214	Cabo	46,413	−0.07	Cabo + nivo dominated
Treatment effectiveness waning	5 years for IO/TKIs, all end points, based on hazards	24	Between 5 and 20 years all IO/TKIs, all end points, based on hazards	Cabo	46,413	−0.07	Cabo + nivo dominated
		26	No TE waning	Cabo	46,413	−0.07	Cabo + nivo dominated
Suni RWE 1L PFS	Log-logistic	29	Weibull	Cabo	46,393	−0.07	Cabo + nivo dominated
Impact of prior TKI treatment	Not considered	76	Exploratory analysis HR1.59 applied to TKI after TKI monotherapy	Cabo	47,047	−0.06	Cabo + nivo dominated
Costs/RDI							
RDI	Applied	41	All RDI set to 100%	Pem + lenv	57,771	−0.07	Cabo + nivo dominated
Lenv dosing within pem + lenv	TA858 and RDI data	75	NHSE input	Cabo	46,413	−0.07	Cabo + nivo dominated

continued

TABLE 32 Scenario analyses (*continued*)

Parameter	Base case		Scenario	Next best comparator[a]	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
Utilities							
Data source used for utilities	JAVELIN Renal 101 for 1L, AXIS trial for 2L and assumed same proportional decrease for 3L and 4L	50	CheckMate 9ER for all lines	Cabo	46,413	−0.11	Cabo + nivo dominated
AEs							
Data source used for AEs	NMA	58	Individual trials	Cabo	45,672	−0.05	Cabo + nivo dominated

a Next best comparator defined as next most efficient non-dominated comparator.

TABLE 33 Base-case results at list price (probabilistic, ordered in increasing cost, 1000 runs)

Tech	Costs (£) (95% CI)	QALYs (95% CI)	LYG (95% CI)	Inc. costs (95% CI)	Inc. LYG (95% CI)	Inc. QALYs (95% CI)	ICER pairwise	ICER f.inc.	Severity modifier
Risk population: all risk									
Suni	85,907 (61,027 to 119,767)	3.032 (2.441 to 3.829)	1.845 (1.458 to 2.29)						1
Pazo	86,557 (60,299 to 119,305)	3.073 (2.446 to 3.857)	1.871 (1.488 to 2.332)	650 (−9098 to 10,351)	0.041 (−0.175 to 0.242)	0.026 (−0.255 to 0.332)		25,472	1
Tivo	106,075 (75,426 to 140,435)	3.021 (2.403 to 3.794)	1.843 (1.44 to 2.31)						1
Cabo + nivo	234,537 (184,298 to 275,935)	3.793 (3.041 to 4.777)	2.34 (1.839 to 2.933)	147,980 (106,917 to 184,569)	0.72 (0.109 to 1.424)	0.47 (−0.01 to 0.983)		315,109	1
Risk population: favourable risk									
Suni	90,575 (59,928 to 125,997)	3.936 (3.117 to 4.88)	2.377 (1.825 to 2.97)						1
Pazo	91,140 (59,620 to 128,301)	3.978 (3.113 to 4.948)	2.395 (1.841 to 3)	565 (−9360 to 10,588)	0.042 (−0.183 to 0.255)	0.018 (−0.402 to 0.458)		31,936	1
Tivo	121,973 (87,182 to 161,284)	3.925 (3.056 to 4.965)	2.37 (1.811 to 2.978)						1
Cabo + nivo	270,118 (187,712 to 359,262)	4.659 (3.149 to 6.548)	2.798 (1.914 to 3.844)	178,978 (104,613 to 2665,41)	0.681 (−0.569 to 2.106)	0.403 (−0.388 to 1.384)		443,970	1
Risk population: intermediate/poor risk									
Suni	83,165 (54,213 to 115,521)	2.701 (2.111 to 3.498)	1.641 (1.272 to 2.114)						1.2
Pazo	83,516 (55,144 to 116,868)	2.735 (2.12 to 3.524)	1.661 (1.298 to 2.1)	352 (−8834 to 10,307)	0.034 (−0.168 to 0.245)	0.02 (−0.207 to 0.241)		17,740	1.2
Tivo	98,665 (68,322 to 134,045)	2.682 (2.084 to 3.513)	1.634 (1.283 to 2.092)						1.2
Nivo + ipi	118,314 (89,848 to 146,498)	2.321 (1.639 to 3.372)	1.442 (1.041 to 1.987)						1
Cabo	166,044 (122,409 to 200,914)	3.66 (2.945 to 4.512)	2.233 (1.778 to 2.747)						1
Pem + lenv	184,683 (143,856 to 225,715)	3.817 (3.072 to 4.813)	2.366 (1.845 to 2.969)	101,167 (69,623 to 134,260)	1.082 (0.426 to 1.778)	0.705 (0.21 to 1.261)		143,469	1
Cabo + nivo	212,254 (165,233 to 250,672)	3.432 (2.698 to 4.431)	2.127 (1.65 to 2.697)						1

ext, extended; f.inc: fully incremental; inc., incremental.

frontier) is cabozantinib plus nivolumab. In the intermediate-/poor-risk group, the only novel treatment to lie along the frontier is lenvatinib plus pembrolizumab.

Model validation and face validity check

Comparison of state transition and partitioned survival-analysis results

The state transition and PartSA results were broadly consistent. Where there were differences, the EAG considered the difference in results between model structures to have been adequately explained. Both model structures provide value to the Committee in decision-making, which is dependent upon how the fundamental issue driving the differences (effectiveness in the favourable-risk population) is interpreted. See *Report Supplementary Material 1*, 'Comparison of state transition and PartSA results' for more details.

Model fits to United Kingdom real-world evidence data

When comparing model predictions to the sunitinib curve for OS (used as the model reference curve) in the state transition model, these showed a good fit to OS for the all-risk and intermediate-/poor-risk populations but showed an underprediction compared to the KM curves for the favourable-risk population (*see Report Supplementary Material 1*). The EAG considered that this was due to the impact of prior risk status as a prognostic factor on outcomes and the second line. For example, a simple analysis shows a significant difference in OS and second line between patients who were intermediate/poor risk at first line and patients who were favourable risk at first line: HR= 1.974, 95% CI (1.471 to 2.649).

In the model base case, there is some underprediction compared to the UK RWE for nivolumab + ipilimumab. It also shows that there is no visible plateau for nivolumab + ipilimumab for OS within the available UK RWE time frame (the maximum time point for which is 3.6 years). When TTNT is used instead of PFS to measure the first-line effectiveness of nivolumab + ipilimumab, the STM provides a better fit to the observed RWE, potentially with a slight amount of overestimation of the OS (*Figure 15*).

(a) Base case using PFS

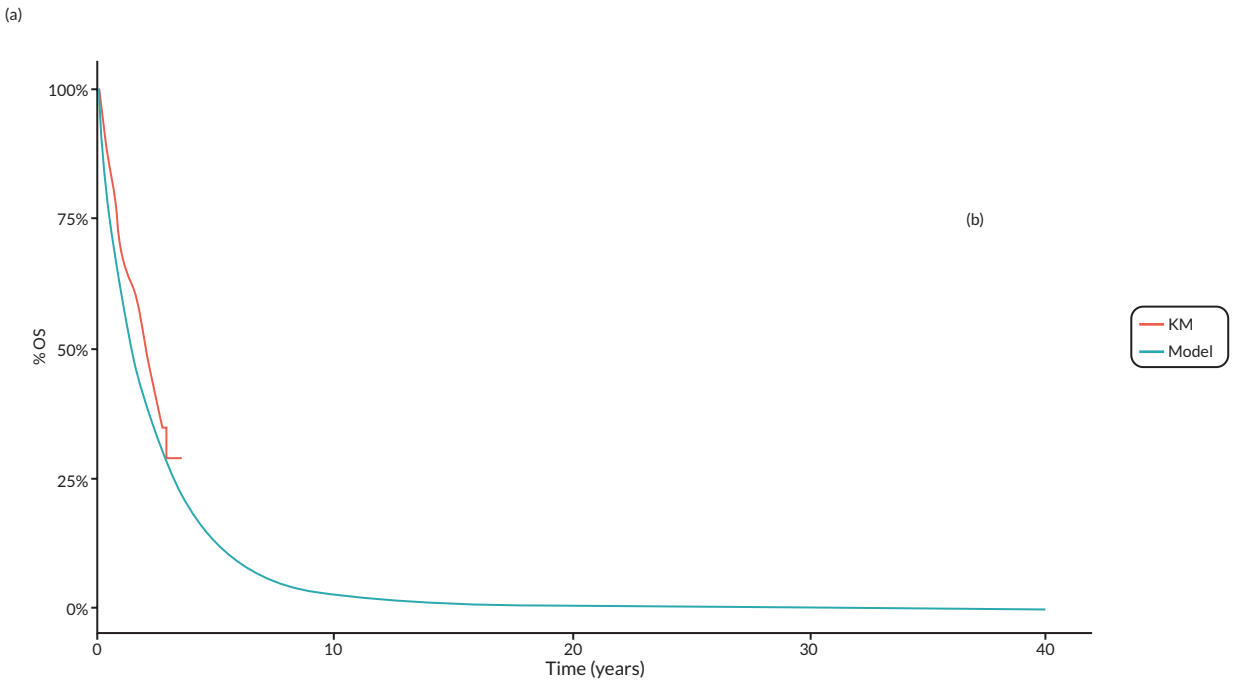


FIGURE 15 Model fit to nivolumab + ipilimumab OS in the intermediate-/poor-risk population when using sunitinib reference curve from UK RWE. (b) Using TTNT as a proxy for effectiveness (scenario analysis 73) is redacted as TTNT data were supplied commercial in confidence.

Model fits to CheckMate 9ER data in trial-based scenario analysis

Figure 16 demonstrates that the state transition model provides a good fit to the trial data for PFS for cabozantinib + nivolumab.

Figure 17 demonstrates that the PartSA analysis using CheckMate 9ER data fits well to OS for both arms (cabozantinib + nivolumab and sunitinib). The state transition model, however, underpredicts for both arms. The EAG considers that this underprediction is likely caused by the following issues:

1. CheckMate 9ER includes substantial numbers of patients receiving subsequent therapies that are not used in UK practice. For example:
   - (confidential information has been removed) of patients who received a subsequent therapy received a subsequent PD-1 therapy after cabo + nivo.
   - (confidential information has been removed) of patients who received a subsequent therapy in the cabo + nivo arm and (confidential information has been removed) of patients in the sunitinib arm received an anti-CTLA4.
   - (confidential information has been removed) of patients who received a subsequent therapy in the cabo + nivo arm and (confidential information has been removed) in the sunitinib arm received other drugs not used in UK practice, including unnamed investigational drugs.

2. CheckMate 9ER did not report third- and fourth-line subsequent therapy use, so UK RWE was used in place.
3. CheckMate 9ER potentially under-reported second-line subsequent therapy. When comparing the number of patients progressing with the number receiving subsequent treatment, (confidential information has been removed) of patients who progressed had no recorded subsequent treatment, a similar lever to that observed in the UK RWE. However, this would not be expected, as generally patients enrolling in trials have greater access to treatment. It is not clear within the CheckMate 9ER trial protocol how subsequent treatment data were collected. The EAG considers it most likely, based upon tables 2–5 of the protocol, that these data were only collected at safety visit follow-ups 1 and 2 (30 and 100 days from discontinuation). Any use after this time point would be missed.
4. Using CheckMate 025 as a reference and second-line-plus NMA based on historical trial data underpredicts the effectiveness of subsequent therapies. As noted in the EAG report,[129] all of the included second-line-plus trials were conducted before IO combinations became available at first line. Most of the trials included treatment standards and prior treatments that are now out of date. Many were conducted before even cabozantinib was in regular use at first line. Within CheckMate 025, for example, the most used previous treatments were sunitinib, pazopanib and axitinib.

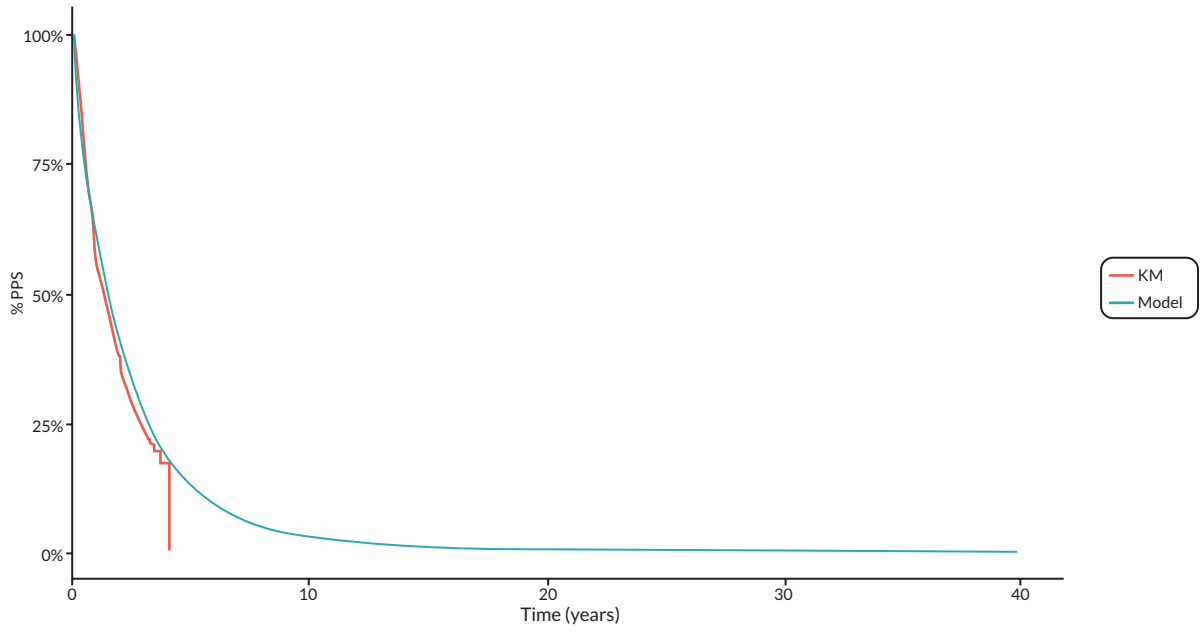


FIGURE 16 State transition model fit to cabo + nivo PFS when using sunitinib reference curve from CheckMate 9ER.

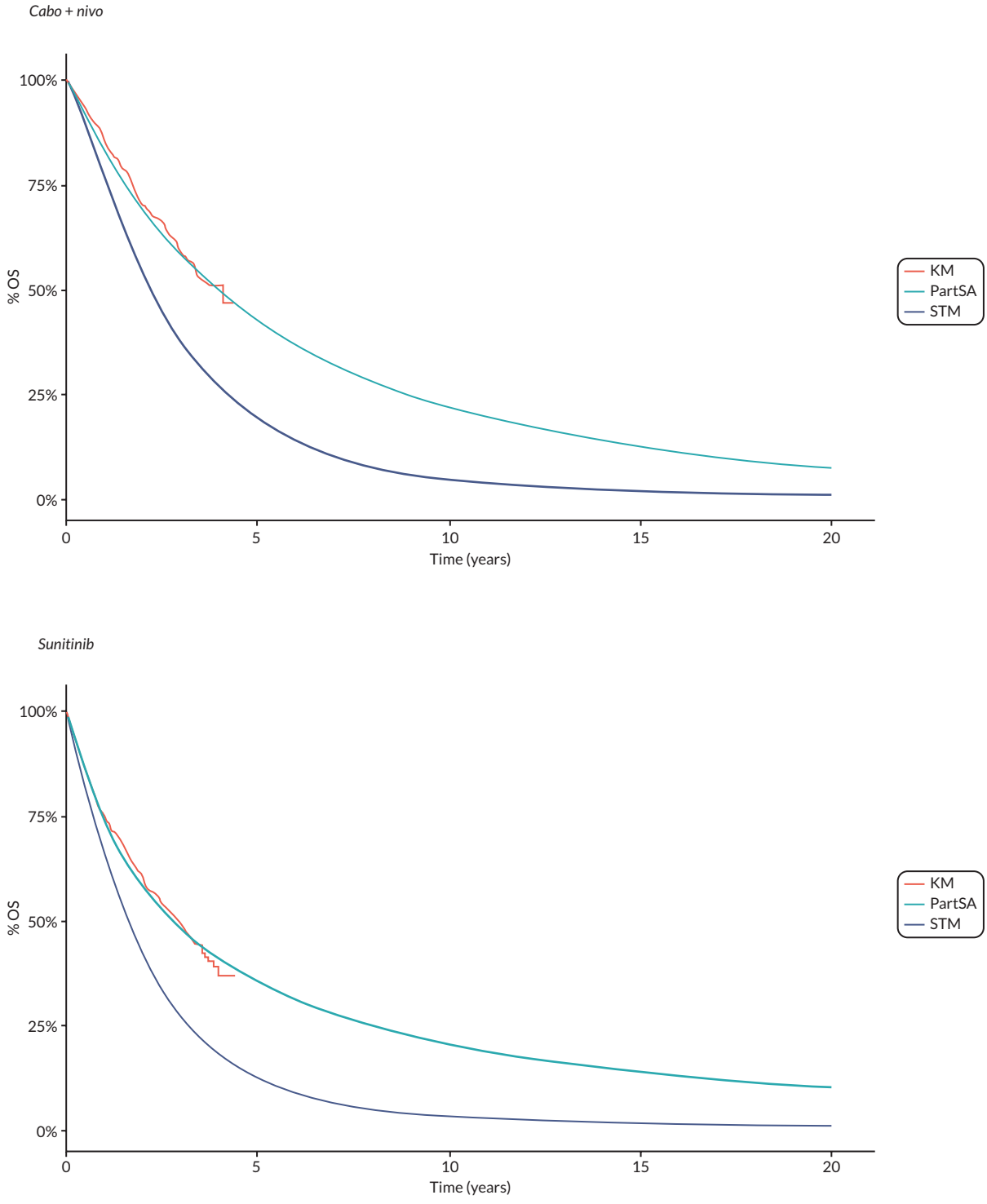


FIGURE 17 Model fit to cabo + nivo OS when using sunitinib reference curve from CheckMate 9ER. STM, state transition model.

With respect to issue (1), the decision problem for cabozantinib + nivolumab should not include either the costs or effectiveness of non-UK subsequent treatments. Therefore, the STM is likely to present a more realistic projection of expected OS. In the case of issues (2) and (3), neither the costs nor benefits of any missing treatments are included within the state transition analysis. If the standard PartSA approach considering only two lines were used the benefits, but not the costs, of third- and fourth-line treatments would be included within the appraisal. In the case of (4), this problem is somewhat mitigated when the UK RWE is used to model second-line treatments, as these data are more up to date and reflective of current practice.

Comparison to prior appraisals

The EAG noted considerable inconsistency in previous estimates across prior appraisals (see *Report Supplementary Material 1*).

The difference in LYs predicted for the same treatment across appraisals is > 100% in a number of cases. These appraisals all used the same model structure. The reason for inconsistency is largely that the data available to inform the models have changed over time. This has also been the case for this appraisal. In fact, it is a strength of this appraisal, compared to prior appraisals, that more mature data are available for several comparators. This is especially relevant, given the role of slippage in estimates of OS and PFS outcomes for certain treatments.

Rather than casting doubt on the EAG's findings, this highlights the importance of NICE's pathways pilot. The use of a common model reference framework creates the conditions for future appraisals to rationalise updated projections, account for what drives updated projections and support Committees to make empirically supported decisions as to whether the inconsistencies are justified.

Benefits not captured in the quality-adjusted life-year 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 BM. This was raised by one of the experts whom we consulted. Literature, however, is conflicting as to whether there may be additional benefit in this subgroup.[224–226]

Chapter 6 Discussion and conclusions

Discussion

This pilot was introduced to help NICE manage the high concentration of topics within a limited number of disease areas; nearly half of its HTAs are focused on just 10 areas. The central idea was to develop reusable reference models for each disease area, minimising repetition and enhancing consistency in decision-making. Additionally, NICE used the pilot to (1) explore the integration of RWE into decision-making; (2) provide innovative approaches for evaluating treatments in disease areas with multiple comparators affecting multiple lines of treatment and (3) conduct a live appraisal for cabozantinib plus nivolumab. This was an ambitious project with multiple aims and objectives.

Considerations for future reference models

- Modelling methods, and outcomes of the cost-effectiveness analyses of various combinations vary across the available literatures, including within prior NICE TAs. This, along with decisions around which comparators to include within appraisals and appraisal sequencing led to treatments being recommended, which, in hindsight and according to both this analysis and the previous MTA, were not cost-effective compared to other treatment options already available in routine practice.
- The history of previous TAs in this disease area (inconsistent modelling approaches and decisions) underlines the benefit of a common modelling framework, including all relevant comparisons, as far as practicable, to enable the consistency of decision-making using the best available data at the time.
- Reference models are of most use in disease areas where multiple appraisals are expected within a reasonably short (~5 year) period of time. This is because the initial time and resource required to develop the reference model is considerably larger – approximately, two to three times of the resources and time are required – based on the experience in this pilot – than that required to develop a model for one product. Benefits for efficiency will only be seen once multiple appraisals have been conducted. If alternative research groups are to use the model for future appraisals, time for handover of the model should be factored in.
- While interaction with stakeholders was encouraged during this pilot and all stakeholders were given the opportunity to provide feedback on the model development plans, the draft report and the model, in an ideal world, reference models should be truly co-created from the project outset (Project HERCULES[227] and the development of the CORE Diabetes Model[228] are good examples of this process). This ensures that all stakeholders are fully engaged. In the case of this assessment, the level of comments received was considerably higher from the submitting company involved in the case study appraisal in comparison to the majority of involved competitor companies. Engagement with clinical and patient stakeholders only took place in the form of evidence submissions to NICE, which may not be the ideal way to co-create future reference models.
- If NICE moves to increased use of reference models, it will be important for academics to take the lead in developing these in partnership with industry. This way, models can be developed that make the best use of all available data and consider all company value propositions without bias towards particular companies. There may also be benefits to taking a more global approach to the development of reference models, as the model structure and a large number of assumptions relating to effectiveness and safety are likely to be geography agnostic. European initiatives to increase sharing of cost-effectiveness analyses as part of HTA,[229,230] which are running in parallel to the new Joint Clinical Assessment process across European Union member states, would provide a logical springboard for these models.
- Our general modelling approach represents a shift from the use of partitioned survival models in the majority of current oncology appraisals to state transition models, though we preserve functionality for partitioned survival models. This 'return' to state transition models was 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 adds additional challenges in obtaining appropriate data and ensuring the plausibility of predictions of OS. This shift may not be necessary for future projects seeking only to provide a reference model for a disease area if mature, generalisable OS data are available.

- The timelines associated with this pilot were extremely ambitious (~5 months from scoping workshop to draft EAG assessment report and ~7 months to final report). Construction of a previous model with a similar scope took a team at the Innovation Value Initiative 2 years' time.[35] This left very little scope to deal with the difficulties arising from the nature of this project as a pilot incorporating not only multiple new processes and technical areas but also requiring application to a live appraisal. There are advantages and disadvantages to development of reference models as part of a live appraisal process, which should be carefully considered for future applications:
  - Advantages: model will be immediately used and therefore there is no risk of a model being developed and never used in an appraisal; in addition, feedback from stakeholders is more likely to be received on models developed academically due to this being in their immediate interest.
  - Disadvantages: live appraisal timelines do not allow for the time necessary to codevelop a reference model with all involved stakeholders having their full say. There is no room to deal with unexpected issues.

- Reference models require regular updates and maintenance. Funding was not put in place at the start of this pilot for this, which means that there are no resources in place to maintain the model and update for use in future appraisals.
- While our model incorporates a considerable amount of flexibility, there are areas where other functionalities might be needed for future appraisals (e.g. the ability to incorporate more complex survival analysis). This will always be the case for a reference model, as building to the maximum possible specification is unlikely to be feasible. This highlights the importance of making a priori arrangements for maintenance and updates.
- Reference models should be provided open source to allow use by all future relevant stakeholders and the ability for stakeholders to flag errors and suggest model changes to be able to evaluate upcoming products more robustly. It took 6 months to agree to a license format for the model. A standard license format for open-source HTA models would be of use for future similar models.
- Reference models need to be transparent and accessible for use by all stakeholders. This requires considerable investment. In our case, the user was able to access considerably more detail than would be possible in a usual STA model (including full intermediate calculations and the generation of report tables and figures automatically within the code); however, we were not able to produce documentation for all functions within the available time frame; nor were we able to provide a user-friendly front end for non-technical users, which should form part of the scope for future reference models.
- Quality control (QC) is also more demanding for a reference model. Within the pilot, QC was conducted internally and by the DSU who identified no major errors; however, future projects of this nature would benefit from additional allocated external and internal QC time and from the use of unit testing within functions to ensure they operate as expected.[231] More complex models, such as this one, would benefit from the incorporation of testing practices used within software engineering; however, the implementation of these practices would require additional time.

Use of sequencing/pathways models

- In our example, there were 744 possible sequences across the three risk groups, ~15,000 rows/columns in our transition matrix to allow for weekly cycles for a lifetime horizon and four lines of active treatment to apply those matrices over including the need for tunnel states to look at time dependency. This level of complexity may not be uncommon for a sequencing model. This decision problem was impossible to address without use of coding software, such as R, and even with the use of such software, long run times can be observed (in our case ~90 minutes to run the state transition model, including time dependency for later line treatments).
- Use of R led to some issues with stakeholders who were unfamiliar with the software interacting with the model. Training, the use of instructional videos and the provision of a user interface are ways to alleviate these types of issues. In our case, timelines only allowed for written instructions to be provided along with calls to walk the company and NICE team through use of the model.
- We encountered a paucity of previous health economic cost-effectiveness models suitable for addressing our decision problem. This meant that the majority of code had to be developed from scratch as available tutorial papers did not provide a suitable example code.
- The sheer scale of the decision problem in terms of the systematic review, NMAs and clinical consultation work required should not be underestimated. In this case, the volume of work roughly equated to two multiple TAs, which was extremely challenging within the allotted timelines. The size of the evidence base, availability of data and complexity of the condition and treatment pathway should be taken into account when thinking about timelines.

- Obtaining the necessary clinical data to accurately assess the impact of drug ordering can be challenging[232–234] and, in this example, heroic assumptions, such as independence of effects, needed to be made. This was particularly the case for later line treatments where trials were older and less suited to the decision problem and patient-level data were not available. It should be noted, however, that our model validated well compared to observed real-world survival data.
- The availability of patient-level data from the UK receiver operating characteristic data set was critical for producing a valid sequencing model. Patient-level data were not provided by any of the manufacturers involved in this pilot. Greater efforts should be made to invest in solutions that allow the use of patient-level data from manufacturers in future modelling of this sort, such as the methods used by Open Safely and presented by Smith *et al.*[235,236]
- Making decisions using NICE's STA process in a multiline, multitreatment decision space is extremely difficult and can lead to perverse outcomes as observed in the decision-making prior to this appraisal in relation to nivolumab plus ipilimumab.
- Additional research is needed to determine how to handle multiple technology recommendations. There is no agreed basis on which a decision-making committee can recommend more than a single option and be confident that its guidance represents an effective use of NHS resources. Statements such as 'options A and B are both cost-effective' or 'options A and B are similarly cost-effective' simply have no meaning at NICE or in broader health economic literature. Thus far, even defining 'similar' has proven elusive even in terms of clinical effectiveness. NICE's cost comparison route lacks a clear definition of similarity.
- NICE recommendations represent an accurate assessment of whether or not a treatment is cost-effective versus scoped comparators at the time the recommendation is made. NICE's current piecewise modus operandi does not consider the potential for treatments to become not cost-effective over time. If we seek to model the impact of technologies on NHS efficiency, rather than thinking about implications for innovation within industry, there are problems with this. Price changes (e.g. when coming off patent, which occurred during our work), displacements due to license changes or new entrants potentially affect the cost-effectiveness of all drugs in a pathway. The previously most cost-effective strategy at any line may change as a consequence. To more accurately represent the cost-effectiveness of treatments in clinical practice, the decision problem would need revisiting every time the state of the world changes. More research is needed to determine whether or not dynamic HTA recommendations are possible and desirable, given the associated challenges.
- Examining the incremental cost-effectiveness (or equivalently, net monetary or health benefit) of possible sequences of treatments may be one approach to take to model the difference in cost-effectiveness of different treatment pathways.[237,238] In our example, the provision of net monetary benefit per sequence was a trivial addition once the model had been set up.

Use of real-world evidence

- The identification, assessment and incorporation of RWE into our economic model was a key challenge. At the outset, the intention was to work with a vendor willing to provide such evidence to NICE. This arrangement was not realised and we were consequently required to use evidence identified during our own evidence review, performed in accordance with the NICE RWE framework.[45]
- The data source used in this appraisal was considerably richer and more complete than might be expected to be available in the majority of disease areas, covering 17 UK centres, providing information on OS, PFS and ToT for up to five lines of therapy and providing information on key disease and demographic variables. Without these data, we would have struggled to fulfil the objective of use of RWE to provide a more accurate representation of the baseline risk. In particular, the SACT database does not include data on PFS, and models built using this would be reliant on either TTD or TTNT as proxies. Our appraisal showed that assuming PFS and TTD/TTNT are similar would have been valid for older treatments, such as TKI monotherapies, but not for some of the newer IO combinations.
- Some RWE data were kept confidential from the companies involved while the data owners completed their publications. This led to protest from several of the company stakeholders, who argued that if they refused to provide data to NICE, they would face negative consequences. While of course this is not an ideal situation, we would note that as the EAG, neither we nor NICE received complete patient-level data in Analysis Data Model format from any of the involved companies and that, as with many oncology submissions, a large volume of critical data were redacted by companies (utility values from the trial, data on ToT, relative dose intensities, etc.).
- We found that when compared with clinical trials, patients in the real world had less favourable outcomes due to treatments being given to people who did not meet restrictive trial inclusion criteria, reflecting the well-known

challenges with the external validity of clinical trials. We also found that subsequent therapies used in the trials differed considerably from those used in the real world. This led to lower estimated OS when using RWE, less absolute OS gain for a given relative efficacy, and therefore less favourable (but more realistic) cost-effectiveness estimates. If NICE moves to regular use of RWE to assess baseline risk, one could expect the need for larger price discounts to ensure cost-effectiveness.

Considerations specific to advanced renal cell carcinoma

- 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 are less well reported, but what is available demonstrates that risk group is a potential TE 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 appear 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 a 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 the 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 PH may not hold within aRCC, but FP NMAs pose additional challenges relating to estimability.
- Relatedly, the duration of TE 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 in linking networks to include all treatments.
- Sparseness of networks precluded the 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.

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 TTNT 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, overoptimistic estimates of OS and PFS compared to RWE, low numbers of UK patients and low use of nivolumab after sunitinib, but it 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.
- The LYs and QALYs predicted in the base case of this appraisal are generally lower than those in previous appraisals. This is consistent with the UK RWE KM data that show reduced PFS and OS compared to trial data. This is true regardless of whether a state transition or PartSA model structure is used and is applicable for all therapies.
- The cost-effectiveness results presented are more generalisable to clinical practice in England than previous renal oncology submissions to NICE. Baseline risk, patient characteristics and treatment pathways were based upon a rich source of UK specific evidence from the UK RWE data set. As expected, use of UK RWE for baseline risk resulted in lower absolute LYs and QALYs for all treatments.
- Cost-effectiveness conclusions differ by the risk subgroup. This is because the comparators available differ and the evidence for the effectiveness of cabozantinib + nivolumab and other IO/TKIs is considerably stronger in the intermediate-/poor-risk population.
- All the results presented in this addendum are at list prices. They should therefore be interpreted with caution as PASs are available for most of the treatments involved.

- There are major uncertainties in the economic and clinical cases for cabozantinib + nivolumab in the favourable-risk subgroup. It is likely that cost-effectiveness estimates for novel treatments drawing on comparatively less mature trials may be unduly optimistic.

Extension of this work

The following additional work was described in the analysis plan for the final phases of the pilot after the appraisal of cabozantinib + nivolumab:

- Review of clinical effectiveness information focusing 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.

The open-source version of the economic model has been released and is available at: https://github.com/nice-digital/NICE-model-repo. Funding was not provided for the other pieces of additional work described in the analysis plan.

Research recommendations

Recommendations in provided in a rough priority order based upon the experience of this pathways pilot within each of the areas.

Disease area recommendations

The NICE is working on a new guideline in kidney cancer,[239] building in part upon this pilot. It was not within the scope of this pilot to provide guidance to clinicians on deciding what the most appropriate first line treatment for a particular patient should bez nor how to order treatments after first line. This will hopefully be addressed within the work of the guidelines team. The below recommendations indicate areas of high priority for future research in aRCC, which may feed into this guidance.

- Future trials for aRCC should be sufficiently powered to analyse the differences in TE by risk group. In particular, more research is needed on the effectiveness of IO/TKI combination treatment in the favourable-risk group.
- More research is needed to understand the impact of prior adjuvant treatment on the effectiveness of first-line treatment options.
- More research is needed to understand how non-clear-cell aRCC responds to different treatments.
- More research is needed to understand what the long-term survival benefits of IO treatments are in the real world.
- More research is needed to understand the optimal sequencing of treatments; particularly after IO combinations have been used.
- More research is needed to understand the long-term HRQoL impacts of different treatments in a real-world setting.

Methodological recommendations

Creation of future reference or pathways models

- Gold standard HTA-ready model templates are needed to enable the creation of reference models in a consistent, transparent and user-friendly framework. These should include code which can be adapted for use for common model types, the ability to automate reporting into HTA templates and a user interface to allow non-technical audiences to interact with the model.

- Research into how living models can be created on the back of living NMAs is warranted. This increases the feasibility of reference model creation for specific disease areas (including pathways models) and would have the capacity to dramatically reduce the time taken to appraise new treatments.
- Research into living models should be coupled with research and guidance on how and when AI should be used to reduce time requirements, particularly within literature review stages.
- Additional research is warranted to explore how models can be run on company IPD using Application Programming Interfaces (APIs) to avoid data-sharing (methodology proposed by Rob Smith and used in Open Safely).[235,236] This would enable complex models to be powered by IPD from all parties to ensure a consistent basis for recommendations.
- Training resources are required, which focus on health economic evaluation applications in R and the version control software Git. These would help make complex models, such as this, more accessible and will allow NICE and other stakeholders to interact with flexible reusable models.
- Research into the ideal model structure for a pathways model/model with multiple lines of treatment is needed. The point at which alternative model structures, such as DES, are required should also be considered.

Health Technology Assessment process

- This research highlighted considerable inconsistencies in previous NICE appraisals, including how cohesion of modelling inputs can be encouraged and research on when this is appropriate is warranted.
- This appraisal demonstrated the large difference the use of RWE can make. It improves the generalisability and estimates of the magnitude of predicted baseline LYGs, and therefore there is room for improvement with new therapies and associated ICERs. Research into the impact of RWE use in TAs, including to model baseline risk or inform the severity modifier, would be of benefit.
- Further research is warranted to explore the impact of making pairwise comparisons/'would otherwise be offered' decisions on society. Such decisions may deviate from the efficiency frontier, which depicts the set of optimal healthcare interventions that provide the best possible health outcomes for a given level of resources. These types of decisions may therefore come with an opportunity cost to society as they may not result in the most cost-effective treatments being used.

Oncology modelling

- This research highlighted the weaknesses of PartSA for the modelling of cancer treatments, particularly when the expected subsequent therapies differ substantially from those received in trials. Exploration of scenarios in which PartSA is and is not reliable is required.
- Additional research into surrogacy between TTD, PFS, TTNT and OS for oncology treatments with different mechanisms of actions is warranted, along with consideration of how this would interact with model conceptualisation for evaluation of pathways.

General statistical and modelling methods

- More research is needed on the optimal methods for time-varying NMA and how to handle unusual circumstances, such as zero events, within initial time periods.
- Additional research is needed to explore the statistical and modelling methods to account for differences in subsequent treatment between trials and practice and the impact of type of prior treatment on outcomes (extension of TSD16).[240]
- Additional research is warranted to explore the circumstances in which simplifying assumptions (such as use of deterministic lambda approximation) are reasonable to reduce run times for complex models.
- Guidance is required on code-based modelling, including the standardisation of steps to go from the conceptual model to the logic model required for coding.

Chapter 7 Equality, diversion and inclusion

The use of cabozantinib plus nivolumab was not expected to raise or address any equalities issues. This appraisal incorporated RWD with a broader sample of people with aRCC than were included in clinical trials of treatments for aRCC, meaning that the evidence base is more representative of the full target population for these treatments. Peninsula Technology Assessment Group (PenTAG) is committed to equality, diversity and inclusion in our work and the research team comprised people from across a variety of backgrounds. As part of our commitment to supporting education and development, three external individuals interested in learning about HTA were invited to contribute to the research while receiving training, and they were either acknowledged or became authors on the final report.

Additional information

CRediT contribution statement

Dawn Lee (https://orcid.org/0000-0003-4027-8456): Conceptualization (equal), Data curation (equal), Formal analysis (equal), Methodology (equal), Project administration (lead), Software (equal), Supervision (lead), Writing - original draft (equal), Writing - review & editing (equal).

Madhusubramanian Muthukumar (https://orcid.org/0000-0003-0885-4097): Data curation (equal), Formal analysis (equal), Methodology (equal), Software (equal), Writing - original draft (equal).

Alan Lovell (https://orcid.org/0009-0003-2151-0720): Data curation (supporting), Methodology (equal), Project administration (supporting), Writing - original draft (equal), Writing - review & editing (equal).

Caroline Farmer (https://orcid.org/0000-0003-3553-8970): Data curation (supporting), Formal analysis (equal), Methodology (equal), Project administration (supporting), Supervision (supporting), Writing - original draft (equal), Writing - review & editing (equal).

Darren Burns (https://orcid.org/0000-0002-5209-8041): Formal analysis (equal), Methodology (equal), Software (equal).

Justin Matthews (https://orcid.org/0000-0002-3743-1627): Formal analysis (equal), Methodology (equal).

Helen Coelho (https://orcid.org/0000-0002-4799-4300): Formal analysis (equal), Writing - original draft (equal).

Brian O'Toole (https://orcid.org/0009-0005-1418-5928): Formal analysis (equal), Writing - original draft (equal).

Laura A Trigg (https://orcid.org/0000-0002-8447-2616): Data curation (equal), Formal analysis (equal), Writing - original draft (equal).

Tristan M Snowsill (https://orcid.org/0000-0001-7406-2819): Methodology (equal), Software (equal).

Maxwell S Barnish (https://orcid.org/0000-0003-0139-6548): Formal analysis (equal), Writing - original draft (equal).

Thalia Nikoglou (https://orcid.org/0009-0007-7406-9528): Formal analysis (equal), Writing - review & editing (equal).

Amanda Brand (https://orcid.org/0000-0002-2421-1414): Formal analysis (equal).

Zain Ahmad (https://orcid.org/0009-0003-3618-7102): Methodology (equal), Formal analysis (equal).

Ahmed Abdelsabour (https://orcid.org/0009-0007-2532-4676): Formal analysis (equal), Writing - original draft (equal).

Louise Crathorne (https://orcid.org/0000-0001-8018-517X): Formal analysis (equal), Writing - original draft (equal), Writing - review & editing (equal).

Sophie Robinson (https://orcid.org/0000-0003-0463-875X): Methodology (equal).

Edward CF Wilson (https://orcid.org/0000-0002-8369-1577): Methodology (equal), Software (equal).

GJ Melendez-Torres (https://orcid.org/0000-0002-9823-4790): Conceptualization (equal), Data curation (equal), Formal analysis (equal), Methodology (equal), Project administration (supporting), Supervision (supporting), Writing - original draft (equal).

Patient data statement

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that they are stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: https://understandingpatientdata.org.uk/data-citation

Data-sharing statement

Requests for access to data from secondary research included in this project should be addressed to the corresponding author or to the data custodian (if known). Data received for the project from the UK RWD source are confidential pending their publication in due course, and requests for access should be made to the data holders at this point. Data received from stakeholders not accessible in this report are confidential and cannot be shared but may be accessible from the data owners. The open source disease model that will be developed from this research will ultimately be freely accessible from the NICE.org.uk website.

Ethics statement

The systematic reviews for this research do not require ethical approval as they are based on the inclusion of secondary research. RWD used in the research were collected and anonymised by NHS clinicians external to this research project, who received ethical approval through their respective institutions' ethical procedures. The expert elicitation exercise does not require ethical approval as this was conducted as part of routine consultation of clinical experts to inform the EAG appraisal.

Information governance statement

PenTAG, University of Exeter, is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (EU GDPR) 2016/679. Under the Data Protection legislation, PenTAG, University of Exeter, is the Data Controller and we process personal data in accordance with their instructions. You can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for the University of Exeter's Data Protection Officer here https://universityofexeteruk.sharepoint.com/sites/InformationGovernance/SitePages/Meet-the-team.aspx

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/GJDL0327.

Primary conflicts of interest: Dawn Lee: Provided private consultancy services to Neuraxpharm, Ascenian Consulting and Market Research, unrelated to possible intervention, comparators, intervention or comparator companies in this RCC appraisal. Lumanity: Was previously employed as the Chief Scientific Officer at Lumanity until September 2022. Work from December 2021 included projects and companies unrelated to possible intervention, comparators, intervention or comparator companies in this RCC appraisal. A very small shareholder in Value Demonstration Group (who are the holding company for Lumanity), Fiecon Ltd: A family member owns Fiecon, however, is not a shareholder of Fiecon. A member of Fiecon's strategic council from November 2023 (receiving financial income). Not specifically related to RCC. Provides mentoring support (receiving financial income). Not specifically related to RCC. Provide consultancy services to

Fiecon on several projects, including one topic for Eisai, unrelated to possible intervention, comparators, intervention or comparator companies in this RCC appraisal.

Darren Burns: Employed as an analyst at Delta Hat Limited from September 2022. Principal Health Economist at Lumanity/Bresmed until September 2022. From December 2021, worked on projects and for companies unrelated to possible intervention, comparators, intervention or comparator companies in this RCC appraisal. Consultancy services (financial) for Pfizer on a technology unrelated to RCC and for Servier on Multilocular cystic renal cell carcinoma (MCRCC), which is not considered as a possible comparator in this appraisal. Scottish Medicines Consortium on multiple projects

GJ Melendez-Torres: Primary investigator or coinvestigator on several NIHR-funded projects to deliver evidence syntheses of health interventions and services. None of these are anticipated to be a conflict with the present work. At the time of submission, he was a member of an NIHR Programme Grants for Applied Research subcommittee, the chair of the NIHR Academy Pre-doctoral Local Authority Fellowship Selection Committee and a member of NICE Technology Appraisal Committee A.

Disclaimer

Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

Publications

Lee D, Burns D, Wilson E. NICE's pathways pilot: pursuing good decision making in difficult circumstances. *Pharmacoecon Open* 2024;**8**:645–9. https://doi.org/10.1007/s41669-024-00490-x. Epub 2024 Apr 13. [Erratum published in *Pharmacoecon Open* 2024;**8**:783.]. https://doi.org/10.1007/s41669-024-00499-2

Lee D, Melendez-Torres GJ, Challapalli A, Frazer R, McGrane J, Bahl A. Efficacy of cabozantinib and sunitinib for the treatment of intermediate/poor risk renal cell carcinoma based upon UK real-world data. *ESMO Real World Data Digit Oncol* 2024;**6**:100087.

Lee D, Ahmad Z, Larkin JMG, Bahl A, Melendez-Torres GJ. Structured expert elicitation to inform long-term survival extrapolations in advanced renal cell carcinoma [published online ahead of print 27 August 2025]. *Appl Health Econ Health Policy* 2025. https://doi.org/10.1007/s40258-025-01000-8. PMID: 40864385.

McGrane J, Frazer R, Challapalli A, Ratnayake G, Boh Z, Clayton A, *et al*. Real world, multicentre patterns of treatment and survival in metastatic renal cell carcinoma with the UK Renal Oncology Collaborative (UK ROC): is it time to look favourably on first-line immunotherapy containing combinations in all IMDC groups? *Cancer Med* 2024;**13**:e7327. https://doi.org/10.1002/cam4.7327

Smith RA, Samyshkin Y, Mohammed W, Lamrock F, Ward T, Smith J, *et al*. assertHE: an R package to improve quality assurance of HTA models [version 1; peer review: 1 approved with reservations]. *Wellcome Open Res* 2024;**9**:701.

Lee D, Ahmad Z, Farmer C, Barnish MS, Lovell A, Melendez-Torres GJ. Slipping away: slippage in hazard ratios over datacuts and its impact on immuno-oncology combination economic evaluations. *Value Health* 2025;**28**:260–8. https://doi.org/10.1016/j.jval.2024.09.008

References

1. Cancer Research UK. *Kidney Cancer. Types and Grades*. 2020. URL: www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/types-grades (accessed 9 March 2023).
2. Cancer Research UK. *Number Stages for Kidney Cancer*. 2020. URL: www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/number-stages (accessed 3 January 2023).
3. Bukowski RM. Prognostic factors for survival in metastatic renal cell carcinoma. *Cancer* 2009;**115**:2273–81.
4. Bhat S. Role of surgery in advanced/metastatic renal cell carcinoma. *Indian J Urol* 2010;**26**:167–76.
5. Heng DYC, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, *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**:141–8.
6. 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**:159–65.
7. Cao C, Shou J, Shi H, Jiang W, Kang X, Xie R, *et al*. Novel cut-off values of time from diagnosis to systematic therapy predict the overall survival and the efficacy of targeted therapy in renal cell carcinoma: a long-term, follow-up, retrospective study. *Int J Urol* 2022;**29**:212–20.
8. Iacovelli R, Galli L, De Giorgi U, Porta C, Nolè F, Zucali P, *et al*. The effect of a treatment delay on outcome in metastatic renal cell carcinoma. *Urol Oncol* 2019;**37**:529.e1–7.
9. Woldu SL, Matulay JT, Clinton TN, Singla N, Freifeld YN, Sanli ON, *et al*. Utilization and survival implications of a delayed approach to targeted therapy for metastatic renal cell carcinoma: a nationwide cancer registry study. *J Clin Oncol* 2018;**36**:586–586.
10. 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**:35–40.
11. 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**:289–96.
12. 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**:475–80.
13. Landonio G, Baiocchi C, Cattaneo D, Ferrari M, Gottardi O, Majno M, Ghislandi E. Retrospective analysis of 156 cases of metastatic renal cell carcinoma: evaluation of prognostic factors and response to different treatments. *Tumori* 1994;**80**:468–72.
14. NHS England. *National Cancer Drugs Fund List: Version 1.256*. London. 2023. URL: www.england.nhs.uk/publication/national-cancer-drugs-fund-list/ (accessed 31 March 2023).
15. National Institute for Health and Care Excellence (NICE). *Lenvatinib with Pembrolizumab for Untreated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA858]*. London: NICE; 2023.
16. National Institute for Health and Care Excellence (NICE). *Pembrolizumab for Adjuvant Treatment of Renal Cell Carcinoma. Technology Appraisal Guidance [TA830]*. London: NICE; 2022.
17. Larroquette M, Peyraud F, Domblides C, Lefort F, Bernhard JC, Ravaud A, Gross-Goupil M. Adjuvant therapy in renal cell carcinoma: current knowledges and future perspectives. *Cancer Treat Rev* 2021;**97**:102207.
18. 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**:15.
19. Pessoa RR, Kim S. *BMJ Best Practice: Renal Cell Carcinoma. 2022 [Updated 26 July 2022. Last Reviewed 3 December 2022]*. URL: https://bestpractice.bmj.com/topics/en-gb/261 (accessed 3 January 2023).

20. Powles T, Albiges L, Bex A, Grünwald V, Porta C, Procopio G, *et al.*; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. ESMO Clinical Practice Guideline update on the use of immunotherapy in early stage and advanced renal cell carcinoma. *Ann Oncol* 2021;**32**:1511–9.
21. 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**:2957–95.
22. Rini BI, Signoretti S, Choueiri TK, McDermott DF, Motzer RJ, George S, *et al*. Long-term outcomes with nivolumab plus ipilimumab versus sunitinib in first-line treatment of patients with advanced sarcomatoid renal cell carcinoma. *J Immunother Cancer* 2022;**10**:e005445.
23. National Institute for Health and Care Excellence (NICE). *Nivolumab with Ipilimumab for Untreated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA780]*. London: NICE; 2022.
24. 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.
25. 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.
26. National Institute for Health and Care Excellence (NICE). *Tivozanib for Treating Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA512]*. London: NICE; 2018.
27. National Institute for Health and Care Excellence (NICE). *Cabozantinib for Untreated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA542]*. London: NICE; 2018.
28. Challapalli A, Ratnayake G, Baht A, McGrane J, Frazer R. *UK RCC RWE Data Set*. Bristol: University Hospitals Bristol NHS Foundation Trust; 2022.
29. 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.
30. National Institute for Health and Care Excellence (NICE). *Cabozantinib for Previously Treated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA463]*. London: NICE; 2017.
31. 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.
32. National Institute for Health and Care Excellence (NICE). *Everolimus for Advanced Renal Cell Carcinoma after Previous Treatment. Technology Appraisal Guidance [TA432]*. London: NICE; 2017.
33. National Institute for Health and Care Excellence (NICE). *Nivolumab for Previously Treated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA417]*. London: NICE; 2016.
34. 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. URL: www.nice.org.uk/guidance/gid-ta11186/documents/assessment-report-3 (accessed 12 December 2023).
35. 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**:1313–20.
36. Innovation and Value Initiative (IVI). *IVI-NSCLC Value Model. Part of the Open-Source Value Project*. Alexandria, VA: IVI. URL: https://thevalueinitiative.org/ivi-nsclc-value-model/ (accessed 9 January 2023).
37. Choueiri T, Powles T, Burotto M, Escudier B, Bourlon MT, Zurawski B, *et al*. Nivolumab plus cabozantinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med* 2021;**384**:829–41.
38. ESMO. *EMA Recommends Extension of Therapeutic Indications for Cabozantinib and Nivolumab*. Lugano: European Society for Medical Oncology; 2021. URL: www.esmo.org/oncology-news/ema-recommends-extension-of-therapeutic-indications-for-cabozantinib-and-nivolumab (accessed 20 March 2023).

39. Medicines & Healthcare Products Regulatory Agency (MHRA). *Summary of Product Characteristics. Cabometyx (Cabozantinib)* London: MHRA; 2021. URL: https://mhraproducts4853.blob.core.windows.net/docs/7869464cb3603f66ff25a0f1a4285519f6bef8d7 (accessed 12 February 2023).

40. Centre for Reviews and Dissemination. *Systematic Reviews. CRD's Guidance for Undertaking Reviews in Health Care*. York: University of York; 2009.

41. National Institute for Health and Care Excellence (NICE). *NICE Health Technology Evaluations: The Manual. Process and Methods [PMG36]* (Contract No.: 10 Oct 2024). London: NICE; 2022..

42. 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**:35–49.

43. 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 (Baltimore)* 2022;**101**:e30333.

44. 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**:712–23.

45. National Institute for Health and Care Excellence (NICE). *NICE Real-World Evidence Framework: Corporate Document [ECD9]*. London: NICE; 2022.

46. Scottish Intercollegiate Guidelines Network (SIGN). *Search Filters*. URL: www.sign.ac.uk/what-we-do/methodology/search-filters/ (accessed 9 January 2023).

47. Ayiku L, Levay P, Hudson T, Craven J, Barrett E, Finnegan A, Adams R. 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**:200–16.

48. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan–Meier curves. *Stata J* 2017;**17**:786–802.

49. 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.

50. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**:1–158.

51. Dias S, Ades A, Welton N, Jansen J, Sutton A. *Network Meta-Analysis for Decision Making*. Hoboken, NJ: John Wiley & Sons; 2018.

52. 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**:839–61.

53. Wiksten A, Hawkins N, Piepho HP, Gsteiger S. Nonproportional hazards in network meta-analysis: efficient strategies for model building and analysis. *Value Health* 2020;**23**:918–27.

54. 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**:984–98.

55. 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; 2022. URL: https://roche.github.io/gemtcPlus/ (accessed 6 February 2023).

56. Collett D. *Modelling Survival Data in Medical Research*. 3rd edn. New York, NY: Chapman and Hall/CRC; 2014.

57. National Institute for Health and Care Excellence (NICE). *Avelumab with Axitinib for Untreated Advanced Renal Cell Carcinoma. Technology Appraisal Guidance [TA645]*. (Contract No.: 13 Sep 2024). London: NICE; 2020.

58. *National Cancer Registration and Analysis Service (NCRAS) [Internet]*. London: National Disease Registration Service; 2023.
59. *SACT Systemic Anti-Cancer Therapy. Chemotherapy Dataset [Internet]*. London: Public Health England; 2022.
60. *Clinical Practice Research Datalink (CPRD) [Internet]*. London: The Medicines and Healthcare Products Regulatory Agency; 2023.
61. *Hospital Episode Statistics (HES) [Internet]*. London: NHS Digital; 2023.
62. Marchioni M, Kriegmair M, Heck M, Amiel T, Porpiglia F, Ceccucci E, *et al*.; EAU-Young Academic Urologist Kidney Cancer Group. 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**:256–63.
63. *International mRCC Database Consortium (imdc) [Internet]*. Calgary: University of Calgary; 2021. URL: www.imdconline.com (accessed 16 March 2023)
64. Maroun R, Mitrofan L, Benjamin L, Nachbaur G, Maunoury F, Le Jeunne P, Durand-Zaleski I. 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**:214.
65. Challapalli A, Ratnayake G, McGrane J, Frazer R, Gupta S, Parslow DS, *et al*. 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**:S1215.
66. Frazer R, McGrane JM, Challapalli A, Ratnayake G, Malik J, Forde C, *et al*. Real-world patterns of treatment and response in metastatic renal cell carcinoma: a multicentre UK-wide review with UK Renal Oncology Collaborative (UK ROC). *ESMO Real World Data Digit Oncol* 2024;**3**:100027.
67. McGrane J, Frazer R, Challapalli A, Ratnayake G, Boh Z, Clayton A, *et al*. Real world, multicentre patterns of treatment and survival in metastatic renal cell carcinoma with the UK Renal Oncology Collaborative (UK ROC): is it time to look favourably on first-line immunotherapy containing combinations in all IMDC groups? *Cancer Med* 2024;**13**:e7327.
68. Lee D, Melendez-Torres GJ, Challapalli A, Frazer R, McGrane J, Bahl A. Efficacy of cabozantinib and sunitinib for the treatment of intermediate/poor risk renal cell carcinoma based upon UK real-world data. *ESMO Real World Data Digit Oncol* 2024;**6**:100087.
69. 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**:e588–97.
70. 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**:301–301.
71. 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**:S701.
72. 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**:v389.
73. 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**:670.
74. Olsson-Brown AC, Baxter M, Dobeson C, Feeney L, Lee R, Maynard A, *et al*.; The UK National Oncology Trainee Collaboration for Healthcare Research (NOTCH). 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**:7065–7065.

75. Kidney Cancer UK. *Kidney Cancer UK Quality Performance Audit of Kidney Cancer Services in England.* Cambridge: Kidney Cancer UK; 2022. URL: www.kcuk.org.uk/wp-content/uploads/2022/08/Kidney-Cancer-UK-Accord-AUDIT-REPORT-VFINAL.pdf (accessed 8 January 2023).
76. 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* 2019;**12**:84.
77. 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**:628.
78. 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**:626.
79. IQVIA. *Hospital Audit Data 2022*. Durham, NC: IQVIA; 2022.
80. National Institute for Health and Care Excellence (NICE). *Cabozantinib with Nivolumab for Untreated Advanced Renal Cell Carcinoma [TA964]*. London: NICE; 2024.
81. 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**:378–88.
82. 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**:1931–9.
83. Grünwald V, Hilser T, Meiler J, Goebell PJ, Ivanyi P, Strauss A, *et al*. A prospectively randomized phase-II trial of axitinib versus everolimus as second-line therapy in metastatic renal cell carcinoma (BERAT study). *Oncol Res Treat* 2022;**45**:272–80.
84. 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**:612–24.
85. 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.
86. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, *et al*.; CheckMate 025 Investigators. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015;**373**:1803–13.
87. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, *et al*.; CheckMate 214 Investigators. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018;**378**:1277–90.
88. Motzer R, Alekseev B, Rha SY, Porta C, Eto M, Powles T, *et al*.; CLEAR Trial Investigators. Lenvatinib plus pembrolizumab or everolimus for advanced renal cell carcinoma. *N Engl J Med* 2021;**384**:1289–300.
89. 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**:722–31.
90. 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**:e374–85.
91. 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**:866–74.

92. 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**:1287–94.

93. 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**:1103–15.

94. Choueiri TK, Escudier B, Powles T, Mainwaring PN, Rini BI, Donskov F, *et al*.; METEOR Investigators. Cabozantinib versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015;**373**:1814–23.

95. 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**:1473–82.

96. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, *et al*.; RECORD-1 Study Group. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008;**372**:449–56.

97. 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**:2765–72.

98. 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**:837–47.

99. 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.

100. 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**:695–703.

101. 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**:3791–9.

102. 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**:95–104.

103. 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**:1061–8.

104. 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**:2085–97.

105. 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**:4156–67.

106. 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**:100–1.

107. 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**:4502.

108. Ipsen. *Company Submission for 'Treatments for Renal Cell Carcinoma [ID6186]: Pathways Pilot Appraisal'*. London: Ipsen Pharma Ltd; 2023.

109. 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**:101210.

110. 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**:4546.

111. Cella D, Yount S, Brucker PS, Du H, Bukowski R, Vogelzang N, Bro WP. Development and validation of a scale to measure disease-related symptoms of kidney cancer. *Value Health* 2007;**10**:285–93.

112. 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**:1413–8.

113. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998;**16**:139–44.

114. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outc* 2007;**5**:70.

115. 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**:757–64.

116. de Groot S, Redekop WK, Versteegh MM, Sleijfer S, Oosterwijk E, Kiemeney LALM, Uyl-de Groot CA. Health-related quality of life and its determinants in patients with metastatic renal cell carcinoma. *Qual Life Res* 2018;**27**:115–24.

117. 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**:491–6.

118. 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**:1119–28.

119. 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**:2634–41.

120. 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**:124.

121. 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.

122. 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**:5662–70.

123. 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**:859–69.

124. 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.

125. 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**:S300–1.

126. 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**:S63.

127. 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**:S78.

128. 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, Washington, DC, 15–18 May 2022.

129. National Institute for Health and Care Excellence (NICE). *Renal Cell Carcinoma Pathways Pilot [ID6186]. Committee Papers*. London: NICE; 2023.

130. 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**:e18331.

131. 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**:1235–42.

132. 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**:417.e1–10.

133. 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**:S184–5.

134. 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**:1155–65.

135. 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.

136. 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**:e0175920.

137. 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**:S692–3.

138. 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**:329–329.

139. 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**:4502–4502.

140. 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.

141. 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**:292–303.

142. 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**:297–310.

143. 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**:323–323.

144. 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**:427–39.

145. 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**:1102.

146. Hansen RN, Hackshaw MD, Nagar SP, Arondekar B, Deen KC, Sullivan SD, Ramsey SD. Health care costs among renal cancer patients using pazopanib and sunitinib. *J Manag Care Spec Pharm* 2015;**21**:37–44, 44a.

147. *NHS Reference Costs, 2011–12*. London: Department of Health; 2012.

148. Office for National Statistics (ONS). *CPI Weights 06: Health. Consumer Price Indices for Health 1988–2014*. 2023. URL: www.ons.gov.uk/economy/inflationandpriceindices/timeseries/chzw/mm23 (accessed 29 March 2023).

149. *NHS Reference Costs 2014–15*. London: Department of Health; 2015.

150. Curtis L, Burns A. *Unit Costs of Health and Social Care 2015*. Canterbury: Personal Social Services Research Unit, University of Kent; 2015.

151. ClinicalTrials.gov. *A Study of Cabozantinib (XL184) vs Everolimus in Subjects with Metastatic Renal Cell Carcinoma (METEOR)*. ClinicalTrials.gov Identifier: NCT01865747. 2017. URL: https://clinicaltrials.gov/ct2/show/NCT01865747 (accessed 27 March 2023).

152. Georghiou T, Bardsley M. *Exploring the Cost of Care at the End of Life*. London: Nuffield Trust; 2014. URL: www.nuffieldtrust.org.uk/files/2017-01/end-of-life-care-web-final.pdf (accessed 21 February 2023).

153. Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, Stein K. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. *Health Technol Assess* 2010;**14**:1–184, iii.

154. 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**:603–603.

155. 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**:4527.

156. 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. URL: www.clinicaltrialsregister.eu/ctr-search/search?query=2011-005939-78 (accessed 28 June 2023).

157. 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.

158. 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. URL: 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 (accessed 14 March 2023).

159. ClinicalTrials.gov. *Lenvatinib with Everolimus versus Cabozantinib for Second-Line or Third-Line Treatment of Metastatic Renal Cell Cancer.* ClinicalTrials.gov Identifier: NCT05012371. 2021. URL: https://clinicaltrials.gov/ct2/show/NCT05012371 (accessed 27 March 2023).

160. 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. URL: https://clinicaltrials.gov/ct2/show/NCT03075423 (accessed 27 March 2023).

161. 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. URL: https://clinicaltrials.gov/ct2/show/NCT03595124 (accessed 27 March 2023).

162. ClinicalTrials.gov. *Cabozantinib or Sunitinib Malate in Treating Participants with Metastatic Variant Histology Renal Cell Carcinoma.* ClinicalTrials.gov Identifier: NCT03541902. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03541902 (accessed 27 March 2023).

163. ClinicalTrials.gov. *REduced Frequency ImmuNE Checkpoint Inhibition in Cancers (REFINE).* ClinicalTrials.gov Identifier: NCT04913025. 2021. URL: https://clinicaltrials.gov/ct2/show/NCT04913025 (accessed 31 March 2023).

164. ClinicalTrials.gov. *A Study of Subcutaneous Nivolumab Monotherapy with or without Recombinant Human Hyaluronidase PH20 (rHuPH20).* ClinicalTrials.gov Identifier: NCT03656718. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03656718 (accessed 31 March 2023).

165. 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.

166. 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**:1–124.

167. University of York, Centre for Health Economics. *Structured Expert Elicitation Resources (STEER).* URL: www.york.ac.uk/che/economic-evaluation/steer/ (accessed 4 December 2025).

168. 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.

169. Tenold M, Ravi P, Kumar M, Bowman A, Hammers H, Choueiri TK, Lara PN. Current approaches to the treatment of advanced or metastatic renal cell carcinoma. *ASCO Educat Book* 2020;**40**:187–96.

170. 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**:358.

171. 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**:479–86.

172. 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**:244.

173. 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**:139–48.

174. 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**:488.

175. 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**:485–96.

176. 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.

177. 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**:50–9.

178. 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**:5105.

179. 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.

180. 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**:385–9.

181. Branchoux S, Sofeu CL, Kurt M, Gaudin AF, Italiano A, Rondeau V, Bellera C. 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.

182. 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**:2831–49.

183. 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.

184. 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.

185. Zhu J, Tang RS. A proper statistical inference framework to compare clinical trial and real-world progression-free survival data. *Stat Med* 2022;**41**:5738–52.

186. 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.

187. 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.

188. 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.

189. 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.

190. 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.

191. 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.

192. 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.

193. Motzer RJ, Powles T, Burotto M, Escudier B, Bourlon MT, Shah AY, *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.

194. 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**:340–52.

195. 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.

196. 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**:575–9.

197. 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**:765–76.

198. Zorginstituut Nederland. *Richtlijn kosteneffectiviteitsmodellen in R [Cost-Effectiveness Models Directive in R]*. Diemen: Zorginstituut Nederland [Dutch Healthcare Institute]; 2022. URL: www.zorginstituutnederland.nl/publicaties/publicatie/2022/12/15/richtlijn-kosteneffectiviteitsmodellen-in-r (accessed 6 November 2024).

199. Shiny. *Easy Web Applications in R*. Boston, MA: Posit Software; 2023.

200. European Medicines Agency (EMA). *Cabometyx. Summary of Product Characteristics*. Amsterdam: European Medicines Agency; 2022. URL: www.ema.europa.eu/en/documents/product-information/cabometyx-epar-product-information_en.pdf (accessed 12 February 2023).

201. European Medicines Agency (EMA). *Votrient. Summary of Product Characteristics*. Amsterdam: EMA; 2021. URL: www.ema.europa.eu/en/documents/product-information/votrient-epar-product-information_en.pdf (accessed 15 November 2022).

202. European Medicines Agency (EMA). *Fotivda. Summary of Product Characteristics*. Amsterdam: EMA; 2022. URL: www.ema.europa.eu/en/documents/product-information/fotivda-epar-product-information_en.pdf (accessed 15 February 2023).

203. European Medicines Agency (EMA). *Sutent. Summary of Product Characteristics*. Amsterdam: EMA; 2021. URL: www.ema.europa.eu/en/documents/product-information/sutent-epar-product-information_en.pdf (accessed 17 November 2022).

204. European Medicines Agency (EMA). *Opdivo. Summary of Product Characteristics*. Amsterdam: EMA; 2022. URL: www.ema.europa.eu/en/documents/product-information/opdivo-epar-product-information_en.pdf (accessed 12 April 2023).

205. European Medicines Agency (EMA). *Kisplyx. Summary of Product Characteristics*. Amsterdam: EMA; 2022. URL: www.ema.europa.eu/documents/product-information/kisplyx-epar-product-information_en.pdf (accessed 29 January 2023).

206. European Medicines Agency (EMA). *Keytruda. Summary of Product Characteristics*. Amsterdam: EMA; 2022. URL: www.ema.europa.eu/en/documents/product-information/keytruda-epar-product-information_en.pdf (accessed 29 January 2023).

207. 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.

208. Burnham K, Anderson D. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. 2nd edn. New York; London: Springer; 2002.

209. Raftery A. Bayesian model selection in social research. *Sociol Methodol* 1995;**25**:111–63.

210. Lee D, Ahmad Z, Farmer C, Barnish MS, Lovell A, Melendez-Torres GJ. Slipping away: slippage in hazard ratios over datacuts and its impact on immuno-oncology combination economic evaluations. *Value Health* 2025;**28**:260–8.

211. The Uromigos. *ASCO 2023 Emergency Podcast IPINIVO vs VEGFPD1 in Kidney Cancer. Podcast. Episode 249. Spotify for Podcasters*. 2023. URL: https://podcasters.spotify.com/pod/show/the-uromigos/episodes/1st-ASCO23-emergency-podcast-IPINIVO-vs-VEGFPD1-in-kidney-cancer-e25rmji (accessed 14 August 2024).

212. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 2002;**359**:1686–9.

213. Office for National Statistics (ONS). *National Life Tables: England and Wales*. 2021 [updated 23 September 2021]. URL: www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables (accessed 19 Jan 2023).

214. 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**:CD013798.

215. 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.

216. 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**:708–15.

217. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18.

218. 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.

219. 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.

220. 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**:1–278.

221. *British National Formulary (BNF)*. (Contract No.: 31 March 2023). London: BMJ & Royal Pharmaceutical Society; 2023.

222. *Drugs and Pharmaceutical Electronic Market Information Tool (eMIT)*. London: Department of Health and Social Care; 2023. URL: www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit (accessed 2 Apr 2023).

223. NHS England. *National Schedule of NHS Costs 2021/22. V3*. 2023 [updated 19 May 2023]. URL: www.england.nhs.uk/publication/2021-22-national-cost-collection-data-publication/ (accessed 25 June 2023).

224. 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**:781.

225. 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**:542–9.

226. 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**:211–5.

227. Duchenne UK. *Project HERCULES*. 2024. URL: www.duchenneuk.org/project-hercules/ (accessed 19 February 2024).

228. Brown J, Palmer A, Bisgaard P, Chan W, Pedula K, Russell A. The Mt. Hood challenge: cross-testing two diabetes simulation models. *Diabetes Res Clin Pract* 2000;**50**:S57–64.

229. European Commission. *New EU Rules on Health Technology Assessment Open Up a New Era for Patient Access to Innovation.* Press Release. Brussels: Directorate-General for Communication, European Commission; 2025. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_226 (accessed 4 February 2025).

230. Beneluxa. *Initiative on Pharmaceutical Policy [Homepage]*. Amsterdam, The Netherlands: Beneluxa; 2025.

231. Smith R, Samyshkin Y, Mohammed W, Lamrock F, Ward T, Smith J, *et al*. assertHE: an R package to improve quality assurance of HTA models [version 1; peer review: 1 approved with reservations]. *Wellcome Open Res* 2024;**9**:701.

232. Lord J, Willis S, Eatock J, Tappenden P, Trapero-Bertran M, Miners A, *et al*. Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technol Assess* 2013;**17**:v–vi, 1.

233. Hawkins N, Sculpher M, Claxton K. *A Method for Identifying Optimum Treatment Sequences in Chronic Diseases*. 27th Annual Meeting of the Society for Medical Decision Making, 2005.

234. Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. *Value Health* 2012;**15**:1127–36.

235. Smith RA, Schneider PP, Mohammed W. Living HTA: automating health economic evaluation with R. *Wellcome Open Res* 2022;**7**:194.

236. *OpenSAFELY [Internet]*. Oxford: University of Oxford; 2024. URL: www.opensafely.org/ (accessed 6 March 2024).

237. Eckermann S, Willan AR. Presenting evidence and summary measures to best inform societal decisions when comparing multiple strategies. *PharmacoEconomics* 2011;**29**:563–77.

238. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. 'Time traveling is just too dangerous' but some methods are worth revisiting: the advantages of expected loss curves over cost-effectiveness acceptability curves and Frontier. *Value Health* 2019;**22**:611–8.

239. National Institute for Health and Care Excellence (NICE). *Kidney Cancer in Development (GID-NG10398)*. URL: www.nice.org.uk/guidance/indevelopment/gid-ng10398 (accessed June 2023).

240. Latimer NR, Abrams KR. *TSD 16: Adjusting Survival Time Estimates in the Presence of Treatment Switching*. Sheffield: School of Health and Related Research, University of Sheffield; 2014.

Appendix 1 Literature searches and Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram

Search strategies

This section contains example search strategies for each published evidence type. Full search strategies are provided in appendices to the final report submitted to NICE and available on the NICE website (NICE.org.uk).

Systematic reviews and meta-analyses of RCTs

Ovid MEDLINE(R) ALL <1946–19 December 2022>

Search date: 19 December 2022

#	Search terms	Hits
1	exp renal cell carcinoma/	38,967
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	79,433
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,496
4	or/1-3	85,754
5	exp Kidney Neoplasms/co, dt, pc, th [Complications, Drug Therapy, Prevention and Control, Therapy]	24,002
6	exp antineoplastic agents/	1,224,683
7	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ti.	2,918,163
8	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ab./freq=2	4,058,024
9	exp nivolumab/	4780
10	(nivolumab or "anti-PD-1 human monoclonal antibody MDX-1106" or "Opdivo" or "Opdivo Injection" or "NIVO" or "BMS-936558" or "MDX-1106" or "ONO-4538").mp.	9104
11	exp Ipilimumab/	2762
12	(ipilimumab or "anti-cytotoxic T-lymphocyte-associated antigen-4 monoclonal antibody" or "MOAB CTLA-4" or "monoclonal antibody CTLA-4" or Yervoy or "MDX-CTLA-4" or "BMS-734016" or "MDX-010").mp.	5188
13	(pembrolizumab or keytruda or "MK-3475" or "SCH 900475").mp.	8075
14	(lenvatinib or kisplyx or E7080 or "E?7080").mp.	1797
15	(avelumab or bavencio or MSB0010718 or "MSB?0010718C").mp.	847
16	exp axitinib/	689
17	(axitinib or Inlyta or "AG-013736").mp.	1402
18	(cabozantinib or cometriq or cabometyx or XL184).mp.	1459
19	exp sunitinib/	4073
20	(sunitinib or Sutent or "SU11248" or "SU011248" or "SU11248").mp.	7243

21	(pazopanib or Votrient or “GW786034B”).mp.	2218
22	(tivozanib or Fotivda or AV951 or “AV?951”).mp.	150
23	exp everolimus/	5540
24	(everolimus or Zortress or Certican or Afinitor or Votubia or “RAD 001” or RAD001 or SDZ-RAD or SDZRAD or SDZ RAD).mp.	8786
25	(Belzutifan or Welireg or MK-6482 or PT2977).mp.	53
26	or/5-25	6,153,895
27	(systematic review or meta-analysis).pt.	294,997
28	meta-analysis/or systematic review/or systematic reviews as topic/or meta-analysis as topic/or “meta analysis (topic)”/or “systematic review (topic)”/or exp technology assessment, biomedical/or network meta-analysis/	332,150
29	((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf.	296,051
30	((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,kf.	14,743
31	((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kf.	36,779
32	(data synthes* or data extraction* or data abstraction*).ti,ab,kf.	37,881
33	(handsearch* or hand search*).ti,ab,kf.	10,835
34	(mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kf.	33,973
35	(met analy* or metanaly* or technology assessment* or HTA or HTAs or technology overview* or technology appraisal*).ti,ab,kf.	11,663
36	(meta regression* or metaregression*).ti,ab,kf.	13,549
37	(meta-analy* or metaanaly* or systematic review* or biomedical technology assessment* or bio-medical technology assessment*).mp,hw.	438,050
38	(MEDLINE or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw.	319,211
39	(cochrane or (health adj2 technology assessment) or evidence report).jw.	21,080
40	(comparative adj3 (efficacy or effectiveness)).ti,ab,kf.	16,821
41	(outcomes research or relative effectiveness).ti,ab,kf.	10,926
42	((indirect or indirect treatment or mixed-treatment or bayesian) adj3 comparison*).ti,ab,kf.	4168
43	(meta-analysis or systematic review).mp.	410,085
44	(multi* adj3 treatment adj3 comparison*).ti,ab,kf.	285
45	(mixed adj3 treatment adj3 (meta-analy* or metaanaly*)).ti,ab,kf.	177
46	umbrella review*.ti,ab,kf.	1226
47	(multi* adj2 paramet* adj2 evidence adj2 synthesis).ti,ab,kf.	13
48	(multiparamet* adj2 evidence adj2 synthesis).ti,ab,kf.	18
49	(multi-paramet* adj2 evidence adj2 synthesis).ti,ab,kf.	11
50	or/27-49	644,080
51	(“Case Reports” or Comment or Editorial or “Historical article” or Letter).pt. or “case report”.ti.	4,587,898
52	4 and 26 and 50	1486
53	52 not 51	1394
54	limit 53 to yr=“2018 -Current”	628

Randomised controlled trials

Clinical effectiveness searches: RCT update (top-up)

Database(s): Ovid MEDLINE(R) ALL 1946–24 January 2023

Search date: 24 January 2023

#	Search terms	Hits
1	exp renal cell carcinoma/	39,158
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	80,072
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,958
4	or/1-3	86,420
5	(advanced or unresect* or un?resect* or nonresect* or (non adj2 resect*) or inopera* or unopera* or metastas* or metastat* or "end stage" or "late-stage" or "late stage" or terminal or "stage 3" or "stage iii" or "stage three" or "stage 4" or "stage iv" or "stage four").ti,ab.	1,641,447
6	4 and 5	31,916
7	exp Kidney Neoplasms/co, dt, pc, th [Complications, Drug Therapy, Prevention and Control, Therapy]	24,060
8	exp antineoplastic agents/	1,227,896
9	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ti.	2,938,616
10	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ab./freq=2	4,096,120
11	exp nivolumab/	4852
12	(nivolumab or "anti-PD-1 human monoclonal antibody MDX-1106" or "Opdivo" or "Opdivo Injection" or "NIVO" or "BMS-936558" or "MDX-1106" or "ONO-4538").mp.	9273
13	exp Ipilimumab/	2785
14	(ipilimumab or "anti-cytotoxic T-lymphocyte-associated antigen-4 monoclonal antibody" or "MOAB CTLA-4" or "monoclonal antibody CTLA-4" or Yervoy or "MDX-CTLA-4" or "BMS-734016" or "MDX-010").mp.	5252
15	(pembrolizumab or keytruda or "MK-3475" or "SCH 900475").mp.	8260
16	(lenvatinib or kisplyx or E7080 or "E?7080").mp.	1862
17	(avelumab or bavencio or MSB0010718 or "MSB?0010718C").mp.	866
18	exp axitinib/	693
19	(axitinib or Inlyta or "AG-013736").mp.	1419
20	(cabozantinib or cometriq or cabometyx or XL184).mp.	1492
21	exp sunitinib/	4080
22	(sunitinib or Sutent or "SU11248" or "SU011248" or "SU11248").mp.	7304
23	(pazopanib or Votrient or "GW786034B").mp.	2242
24	(tivozanib or Fotivda or AV951 or "AV?951").mp.	151
25	exp everolimus/	5557

26	(everolimus or Zortress or Certican or Afinitor or Votubia or "RAD 001" or RAD001 or SDZ-RAD or SDZRAD or SDZ RAD).mp.	8834
27	(Belzutifan or Welireg or MK-6482 or PT2977).mp.	59
28	or/7-27	6,200,040
29	randomized controlled trial.pt.	585,212
30	controlled clinical trial.pt.	95,167
31	randomized.ab.	591,414
32	placebo.ab.	235,411
33	clinical trials as topic.sh.	200,787
34	randomly.ab.	401,088
35	trial.ti.	278,624
36	or/29-35	1,501,489
37	exp animals/not humans.sh.	5,086,917
38	36 not 37	1,381,740
39	6 and 28 and 38	2481
40	limit 39 to yr="2021 -Current"	242

Economic evaluations

Database(s): Ovid MEDLINE(R) ALL 1946–9 January 2023

Search date: 9 January 2023

#	Searches	Hits
1	exp renal cell carcinoma/	39,067
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	79,756
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,731
4	or/1-3	86,085
5	(advanced or unresect* or un?resect* or nonresect* or (non adj2 resect*) or inopera* or unopera* or metastas* or metastat* or "end stage" or "late-stage" or "late stage" or terminal or "stage 3" or "stage iii" or "stage three" or "stage 4" or "stage iv" or "stage four").ti,ab.	1,634,422
6	4 and 5	31,811
7	exp Kidney Neoplasms/co, dt, pc, th [Complications, Drug Therapy, Prevention and Control, Therapy]	24,032
8	exp antineoplastic agents/	1,226,142
9	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ti.	2,927,184
10	(efficacy or effectiveness or treatment* or therap* or management or chemotherap* or adjuvant or antineoplastic).ab./freq=2	4,075,456
11	exp nivolumab/	4822

12	(nivolumab or "anti-PD-1 human monoclonal antibody MDX-1106" or "Opdivo" or "Opdivo Injection" or "NIVO" or "BMS-936558" or "MDX-1106" or "ONO-4538").mp.	9197
13	exp Ipilimumab/	2772
14	(ipilimumab or "anti-cytotoxic T-lymphocyte-associated antigen-4 monoclonal antibody" or "MOAB CTLA-4" or "monoclonal antibody CTLA-4" or Yervoy or "MDX-CTLA-4" or "BMS-734016" or "MDX-010").mp.	5215
15	(pembrolizumab or keytruda or "MK-3475" or "SCH 900475").mp.	8170
16	(lenvatinib or kisplyx or E7080 or "E?7080").mp.	1829
17	(avelumab or bavencio or MSB0010718 or "MSB?0010718C").mp.	861
18	exp axitinib/	691
19	(axitinib or Inlyta or "AG-013736").mp.	1414
20	(cabozantinib or cometriq or cabometyx or XL184).mp.	1474
21	exp sunitinib/	4077
22	(sunitinib or Sutent or "SU11248" or "SU011248" or "SU11248").mp.	7276
23	(pazopanib or Votrient or "GW786034B").mp.	2233
24	(tivozanib or Fotivda or AV951 or "AV?951").mp.	152
25	exp everolimus/	5549
26	(everolimus or Zortress or Certican or Afinitor or Votubia or "RAD 001" or RAD001 or SDZ-RAD or SDZRAD or SDZ RAD).mp.	8808
27	(Belzutifan or Welireg or MK-6482 or PT2977).mp.	55
28	or/7-27	6,174,505
29	Economics/	27,484
30	"costs and cost analysis"/	51,061
31	Cost allocation/	2017
32	Cost–benefit analysis/	91,428
33	Cost control/	21,659
34	Cost savings/	12,669
35	Cost of illness/	31,192
36	Cost sharing/	2713
37	"deductibles and coinsurance"/	1846
38	Medical savings accounts/	547
39	Healthcare costs/	43,742
40	Direct service costs/	1217
41	Drug costs/	17,301
42	Employer health costs/	1097
43	Hospital costs/	11,907
44	Health expenditures/	23,560
45	Capital expenditures/	2001

46	Value of life/	5797
47	exp economics, hospital/	25,665
48	exp economics, medical/	14,376
49	Economics, nursing/	4013
50	Economics, pharmaceutical/	3092
51	exp "fees and charges"/	31,278
52	exp budgets/	14,065
53	(low adj cost).mp.	82,135
54	(high adj cost).mp.	18,878
55	(health?care adj cost$).mp.	15,660
56	(fiscal or funding or financial or finance).tw.	188,804
57	(cost adj estimate$).mp.	2676
58	(cost adj variable).mp.	50
59	(unit adj cost$).mp.	3031
60	(economic$ or pharmacoeconomic$ or price$ or pricing).tw.	389,987
61	or/29-60	897,051
62	(editorial or letter or case report or clinical conference or review).pt.	4,916,431
63	exp "systematic review"/or exp meta analysis/	296,555
64	(systematic or meta* or "mixed treatment comparison" or "indirect treatment comparison").ti,ab.	3,349,855
65	62 not (63 or 64)	4,302,209
66	(6 and 28 and 61) not 65	305
67	limit 66 to yr="2009 -Current"	271

Utility studies

Database(s): Ovid MEDLINE(R) ALL 1946–9 January 2023

Search date: 9 January 2023

#	Searches	Hits
1	exp renal cell carcinoma/	39,067
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	79,756
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,731
4	or/1-3	86,085
5	(advanced or unresect* or un?resect* or nonresect* or (non adj2 resect*) or inopera* or unopera* or metastas* or metastat* or "end stage" or "late-stage" or "late stage" or terminal or "stage 3" or "stage iii" or "stage three" or "stage 4" or "stage iv" or "stage four").ti,ab.	1,634,422
6	4 and 5	31,811
7	"Value of Life"/	5797

8	Quality of Life/	257,015
9	quality of life.ti,kf.	110,630
10	((instrument or instruments) adj3 quality of life).ab.	3834
11	Quality-Adjusted Life Years/	15,318
12	quality adjusted life.ti,ab,kf.	16,684
13	(qaly* or qald* or qale* or qtime* or life year or life years).ti,ab,kf.	26,843
14	disability adjusted life.ti,ab,kf.	4934
15	daly*.ti,ab,kf.	4456
16	(sf36 or sf 36 or short form 36 or shortform 36 or short form36 or shortform36 or sf thirtysix or sfthirtysix or sfthirty six or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab,kf.	29,912
17	(sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or shortform6 or short form6).ti,ab,kf.	2555
18	(sf8 or sf 8 or sf eight or sfeight or shortform 8 or shortform 8 or shortform8 or short form8 or shortform eight or short form eight).ti,ab,kf.	604
19	(sf12 or sf 12 or short form 12 or shortform 12 or short form12 or shortform12 or sf twelve or sftwelve or shortform twelve or short form twelve).ti,ab,kf.	7393
20	(sf16 or sf 16 or short form 16 or shortform 16 or short form16 or shortform16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).ti,ab,kf.	39
21	(sf20 or sf 20 or short form 20 or shortform 20 or short form20 or shortform20 or sf twenty or sftwenty or shortform twenty or short form twenty).ti,ab,kf.	448
22	(hql or hqol or h qol or hrqol or hr qol).ti,ab,kf.	22,951
23	(hye or hyes).ti,ab,kf.	76
24	(health* adj2 year* adj2 equivalent*).ti,ab,kf.	48
25	(pqol or qls).ti,ab,kf.	450
26	(quality of wellbeing or quality of well being or index of wellbeing or index of well being or qwb).ti,ab,kf.	692
27	nottingham health profile*.ti,ab,kf.	1222
28	sickness impact profile.ti,ab,kf.	1091
29	exp health status indicators/	340,260
30	(health adj3 (utilit* or status)).ti,ab,kf.	88,742
31	(utilit* adj3 (valu* or measur* or health or life or estimat* or elicit* or disease or score* or weight)).ti,ab,kf.	15,264
32	(preference* adj3 (valu* or measur* or health or life or estimat* or elicit* or disease or score* or instrument or instruments)).ti,ab,kf.	13,811
33	disutilit*.ti,ab,kf.	593
34	rosser.ti,ab,kf.	107
35	willingness to pay.ti,ab,kf.	8121
36	standard gamble*.ti,ab,kf.	906
37	(time trade off or time tradeoff).ti,ab,kf.	1616
38	tto.ti,ab,kf.	1350
39	(hui or hui1 or hui2 or hui3).ti,ab,kf.	1892
40	(eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual).ti,ab,kf.	21,519

41	duke health profile.ti,ab,kf.	92
42	functional status questionnaire.ti,ab,kf.	129
43	dartmouth coop functional health assessment*.ti,ab,kf.	13
44	or/7-43	730,445
45	6 and 44	659
46	(editorial or letter or case report or clinical conference or review).pt.	4,916,431
47	exp "systematic review"/or exp meta analysis/	296,555
48	(systematic or meta* or "mixed treatment comparison" or "indirect treatment comparison").ti,ab.	3,349,855
49	46 not (47 or 48)	4,302,209
50	45 not 49	632
51	exp animals/not humans.sh.	5,080,261
52	50 not 51	630
53	limit 52 to yr="2009 -Current"	497

Studies containing UK costs

Database(s): Ovid MEDLINE(R) ALL 1946–9 January 2023

Search date: 9 January 2023

#	Searches	Hits
1	exp renal cell carcinoma/	39,067
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	79,756
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,731
4	or/1-3	86,085
5	(advanced or unresect* or un?resect* or nonresect* or (non adj2 resect*) or inopera* or unopera* or metastas* or metastat* or "end stage" or "late-stage" or "late stage" or terminal or "stage 3" or "stage iii" or "stage three" or "stage 4" or "stage iv" or "stage four").ti,ab.	1,634,422
6	4 and 5	31,811
7	(cost? adj2 (illness or disease or sickness)).tw.	4713
8	(burden? adj2 (illness or disease? or condition? or economic*)).tw.	52,154
9	("quality-adjusted life years" or "quality adjusted life years" or QALY?).tw.	16,193
10	Quality-adjusted life years/	15,318
11	"cost of illness"/	31,192
12	Health expenditures/	23,560
13	(out-of-pocket adj2 (payment? or expenditure? or cost? or spending or expense?)).tw.	6449

14	(expenditure? adj3 (health or direct or indirect)).tw.	10,563
15	((adjusted or quality-adjusted) adj2 year?).tw.	27,647
16	or/7-15	137,065
17	exp United Kingdom/	387,636
18	(national health service* or nhs*).ti,ab,in.	259,084
19	(english not ((published or publication* or translat* or written or language* or speak* or literature or citation*) adj5 english)).ti,ab.	47,472
20	(gb or "g.b." or britain* or (british* not "british columbia") or uk or "u.k." or united kingdom* or (england* not "new england") or northern ireland* or northern irish* or scotland* or scottish* or ((wales or "south wales") not "new south wales") or welsh*).ti,ab,jw,in.	2,385,721
21	(bath or "bath's" or ((birmingham not alabama*) or ("birmingham's" not alabama*) or bradford or "bradford's" or brighton or "brighton's" or bristol or "bristol's" or carlisle* or "carlisle's" or (cambridge not (massachusetts* or boston* or harvard*)) or ("cambridge's" not (massachusetts* or boston* or harvard*)) or (canterbury not zealand*) or ("canterbury's" not zealand*) or chelmsford or "chelmsford's" or chester or "chester's" or chichester or "chichester's" or coventry or "coventry's" or derby or "derby's" or (durham not (carolina* or nc)) or ("durham's" not (carolina* or nc)) or ely or "ely's" or exeter or "exeter's" or gloucester or "gloucester's" or hereford or "hereford's" or hull or "hull's" or lancaster or "lancaster's" or leeds* or leicester or "leicester's" or (lincoln not nebraska*) or ("lincoln's" not nebraska*) or (liverpool not (new south wales* or nsw)) or ("liverpool's" not (new south wales* or nsw)) or ((london not (ontario* or ont or toronto*)) or ("london's" not (ontario* or ont or toronto*)) or manchester or "manchester's" or (newcastle not (new south wales* or nsw)) or ("newcastle's" not (new south wales* or nsw)) or norwich or "norwich's" or nottingham or "nottingham's" or oxford or "oxford's" or peterborough or "peterborough's" or plymouth or "plymouth's" or portsmouth or "portsmouth's" or preston or "preston's" or ripon or "ripon's" or salford or "salford's" or salisbury or "salisbury's" or sheffield or "sheffield's" or southampton or "southampton's" or st albans or stoke or "stoke's" or sunderland or "sunderland's" or truro or "truro's" or wakefield or "wakefield's" or wells or westminster or "westminster's" or winchester or "winchester's" or wolverhampton or "wolverhampton's" or (worcester not (massachusetts* or boston* or harvard*)) or ("worcester's" not (massachusetts* or boston* or harvard*)) or (york not ("new york*" or ny or ontario* or ont or toronto*)) or ("york's" not ("new york*" or ny or ontario* or ont or toronto*))))).ti,ab,in.	1,690,052
22	(bangor or "bangor's" or cardiff or "cardiff's" or newport or "newport's" or st asaph or "st asaph's" or st davids or swansea or "swansea's").ti,ab,in.	67,819
23	(aberdeen or "aberdeen's" or dundee or "dundee's" or edinburgh or "edinburgh's" or glasgow or "glasgow's" or inverness or (perth not australia*) or ("perth's" not australia*) or stirling or "stirling's").ti,ab,in.	249,038
24	(armagh or "armagh's" or belfast or "belfast's" or lisburn or "lisburn's" or londonderry or "londonderry's" or derry or "derry's" or newry or "newry's").ti,ab,in.	32,543
25	or/17-24	2,994,651
26	(exp africa/or exp americas/or exp antarctic regions/or exp arctic regions/or exp asia/or exp australia/or exp oceania/) not (exp United Kingdom/or europe/)	3,272,772
27	25 not 26	2,836,173
28	6 and 16 and 27	37
29	limit 28 to yr="2017 -Current"	20

General economic studies (costs, resource use, utilities, economic evaluations)

International Network of Agencies for Health Technology Assessment

Search date: 10 January 2023

(((“renal cell cancer” or RCC or “renal cell carcinoma” or “kidney cancer” or “kidney carcinoma” or “clear cell” or “non clear cell” or hypermephroma or “hypernephroid carcinoma”) OR (“Carcinoma, Renal Cell”[mhe]) OR (renal AND (carcinoma or cancer or tumor or tumour or neoplasm or adenocarcinoma)) OR ((kidney AND (carcinoma or cancer or tumor or tumour or neoplasm or adenocarcinoma))) AND (economic* OR cost*) FROM 2009 TO 2023

= 137 hits

Observational studies (to identify sources of real-world evidence)

Database(s): Ovid MEDLINE(R) ALL 1946–18 January 2023

Search date: 18 January 2023

#	Searches	Hits
1	exp renal cell carcinoma/	39,106
2	((renal or kidney) adj3 (carcinoma or cancer* or cancer* or tumor* or tumour* or neoplas* or adenocarcinoma*)).ti,ab.	79,866
3	("renal cell cancer" or RCC or "renal cell carcinoma" or "kidney cancer" or "kidney carcinoma" or "clear?cell" or "non?clear?cell" or hypermephroma or "hypernephroid carcinoma").ti,ab.	50,806
4	or/1-3	86,203
5	epidemiologic studies/	9242
6	exp case control studies/	1,383,274
7	exp cohort studies/	2,436,199
8	case control.tw.	149,642
9	(cohort adj (study or studies)).tw.	298,113
10	Cohort analy$.tw.	11,161
11	(Follow up adj (study or studies)).tw.	55,254
12	(observational adj (study or studies)).tw.	152,540
13	Longitudinal.tw.	309,912
14	Retrospective.tw.	710,258
15	Cross sectional.tw.	487,001
16	Cross-sectional studies/	453,088
17	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16	3,683,297
18	exp United Kingdom/	387,773
19	(national health service* or nhs*).ti,ab,in.	259,935
20	(english not ((published or publication* or translat* or written or language* or speak* or literature or citation*) adj5 english)).ti,ab.	47,619
21	(gb or "g.b." or britain* or (british* not "british columbia") or uk or "u.k." or united kingdom* or (england* not "new england") or northern ireland* or northern irish* or scotland* or scottish* or ((wales or "south wales") not "new south wales") or welsh*).ti,ab,jw,in.	2,390,072

22	(bath or "bath's" or ((birmingham not alabama*) or ("birmingham's" not alabama*) or bradford or "bradford's" or brighton or "brighton's" or bristol or "bristol's" or carlisle* or "carlisle's" or (cambridge not (massachusetts* or boston* or harvard*)) or ("cambridge's" not (massachusetts* or boston* or harvard*)) or (canterbury not zealand*) or ("canterbury's" not zealand*) or chelmsford or "chelmsford's" or chester or "chester's" or chichester or "chichester's" or coventry or "coventry's" or derby or "derby's" or (durham not (carolina* or nc)) or ("durham's" not (carolina* or nc)) or ely or "ely's" or exeter or "exeter's" or gloucester or "gloucester's" or hereford or "hereford's" or hull or "hull's" or lancaster or "lancaster's" or leeds* or leicester or "leicester's" or (lincoln not nebraska*) or ("lincoln's" not nebraska*) or (liverpool not (new south wales* or nsw)) or ("liverpool's" not (new south wales* or nsw)) or ((london not (ontario* or ont or toronto*)) or ("london's" not (ontario* or ont or toronto*)) or manchester or "manchester's" or (newcastle not (new south wales* or nsw)) or ("newcastle's" not (new south wales* or nsw)) or norwich or "norwich's" or nottingham or "nottingham's" or oxford or "oxford's" or peterborough or "peterborough's" or plymouth or "plymouth's" or portsmouth or "portsmouth's" or preston or "preston's" or ripon or "ripon's" or salford or "salford's" or salisbury or "salisbury's" or sheffield or "sheffield's" or southampton or "southampton's" or st albans or stoke or "stoke's" or sunderland or "sunderland's" or truro or "truro's" or wakefield or "wakefield's" or wells or westminster or "westminster's" or winchester or "winchester's" or wolverhampton or "wolverhampton's" or (worcester not (massachusetts* or boston* or harvard*)) or ("worcester's" not (massachusetts* or boston* or harvard*)) or (york not ("new york*" or ny or ontario* or ont or toronto*)) or ("york's" not ("new york*" or ny or ontario* or ont or toronto*))))).ti,ab,in.	1,693,813
23	(bangor or "bangor's" or cardiff or "cardiff's" or newport or "newport's" or st asaph or "st asaph's" or st davids or swansea or "swansea's").ti,ab,in.	67,988
24	(aberdeen or "aberdeen's" or dundee or "dundee's" or edinburgh or "edinburgh's" or glasgow or "glasgow's" or inverness or (perth not australia*) or ("perth's" not australia*) or stirling or "stirling's").ti,ab,in.	249,588
25	(armagh or "armagh's" or belfast or "belfast's" or lisburn or "lisburn's" or londonderry or "londonderry's" or derry or "derry's" or newry or "newry's").ti,ab,in.	32,629
26	or/18-25	2,999,945
27	(exp africa/or exp americas/or exp antarctic regions/or exp arctic regions/or exp asia/or exp australia/or exp oceania/) not (exp United Kingdom/or europe/)	3,275,806
28	26 not 27	2,841,192
29	4 and 17 and 28	1251

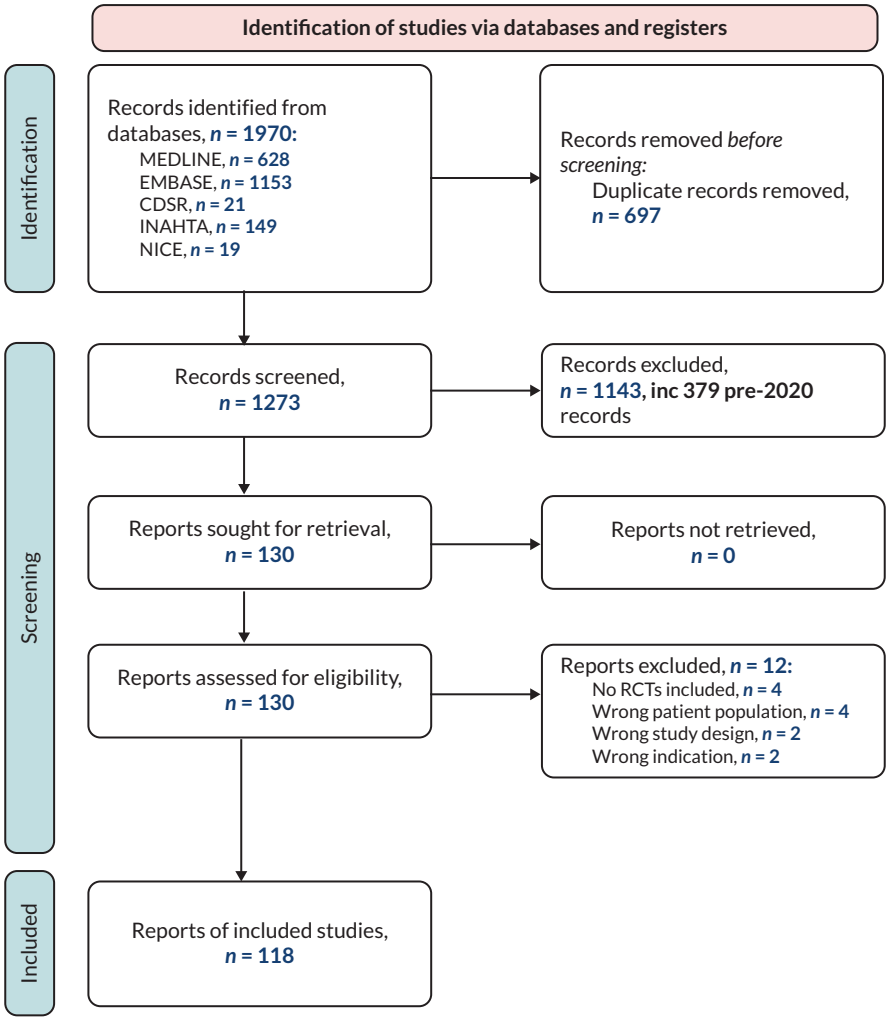


FIGURE 18 Systematic reviews and meta-analyses literature review PRISMA. CDSR, Cochrane Database of Systematic Reviews; INAHTA, International Network of Agencies for Health Technology Assessment.

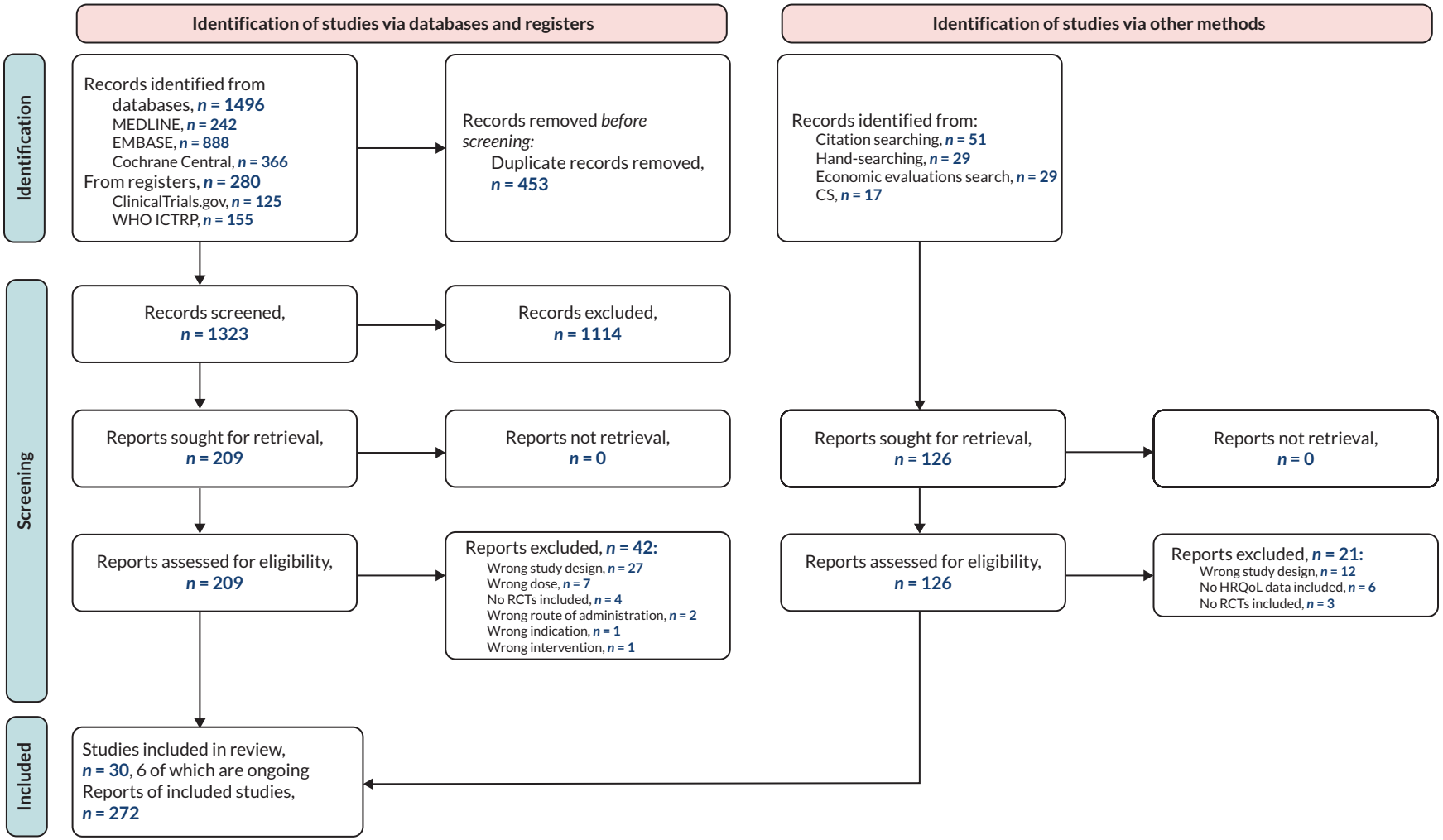


FIGURE 19 Randomised controlled trials literature review PRISMA. WHO ICTRP, World Health Organization International Clinical Trials Registry Platform.

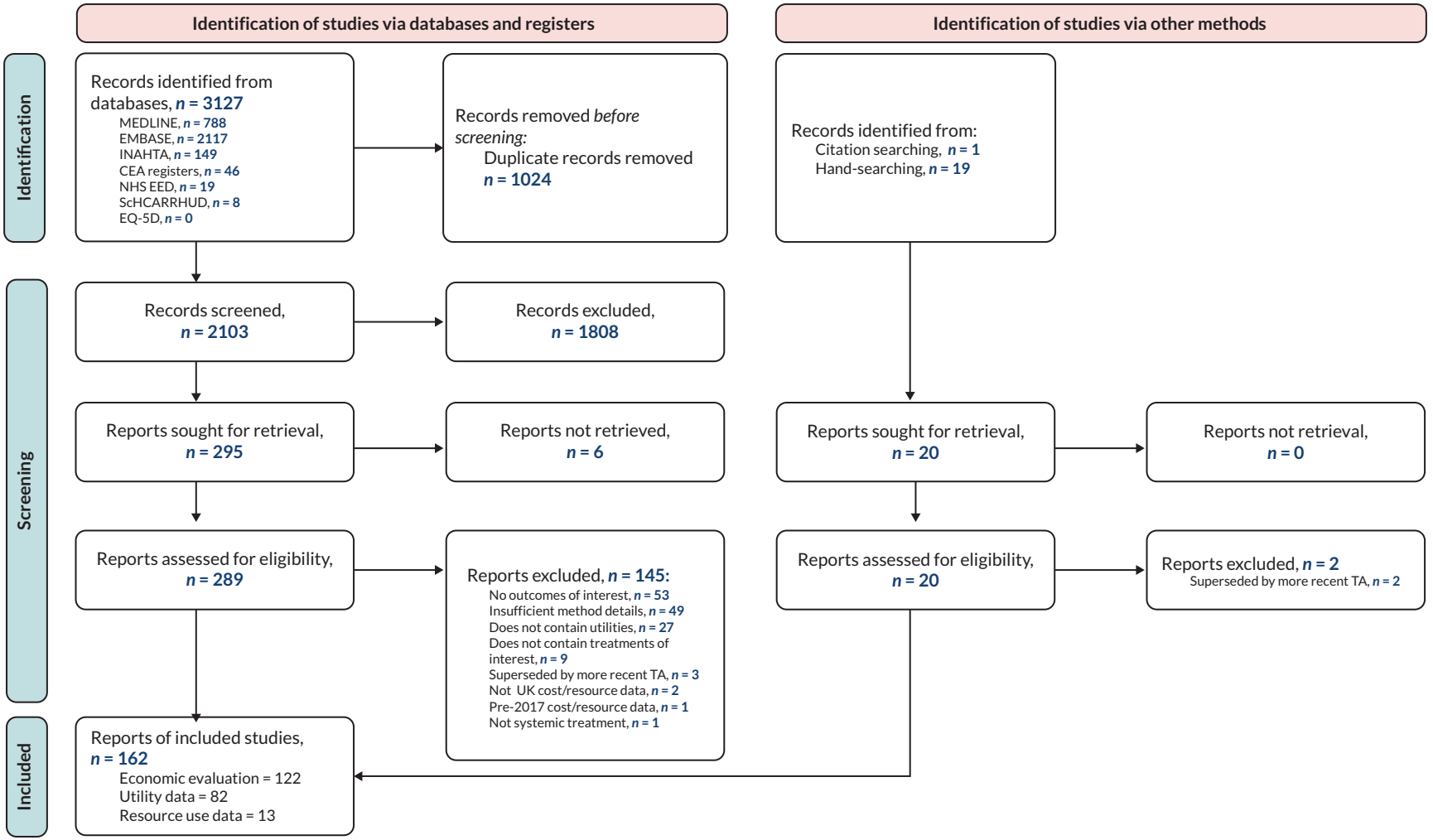


FIGURE 20 Economic literature review PRISMA. Note: A number of studies qualified for more than one of the economic reviews and therefore the total across each of the three reviews (122 + 82 + 13) sums to more than the number of reports included (*n* = 162).CEA, cost-effectiveness analysis; INAHTA, International Network of Agencies for Health Technology Assessment; NHS EED, The NHS Economic Evaluation Database; ScHCARRHUD, School of Health and Related Research Health Utilities Database.

Appendix 2 Included randomised controlled trial characteristics and results

TABLE 34 Study design characteristics of included trials

Trial name	Line	Comparison	Design (blinding)	Study sponsor	Continent: country	Number of centres (number of 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[a]	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, the Netherlands, Poland, Spain, Sweden, Taiwan (Province of China), Turkey, UK	175 (6)	October 2014–February 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, the Netherlands, Poland, Russian Federation, Spain, Switzerland, UK	200 (8)	October 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)	August 2008–September 2011	34.1 months	May 2012
CROSS-J-RCC	1L	Suni vs. sora	Crossover (open label)	Industry and non-industry	Asia: Japan	39 (0)	February 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 of how many centres)	March 2016–December 2017	34.1 months	April 2020
SWITCH	1L	Suni vs. sora	Crossover (open label)	Industry and non-industry	Europe: Germany, Austria, the Netherlands	72 (0)	February 2009–December 2011	15 months	January 2014

continued

TABLE 34 Study design characteristics of included trials (*continued*)

Trial name	Line	Comparison	Design (blinding)	Study sponsor	Continent: country	Number of centres (number of UK centres)	Enrolment period	Final follow-up	Date of last datacut
SWITCH II	1L	Pazo vs. sora	Crossover (open label)	Industry and non-industry	Europe: Germany, Austria, the Netherlands	67 (0)	June 2012–September 2016	NR; KM > 45 months	November 2016
TIVO-1[a]	1L and 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)	February 2010–August 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 (Province of China), UK	175 (11)	15 September 2008–23 July 2010	37 months	November 2011
BERAT[a]	2L	TKI (axi/suni) vs. evero	Crossover (open label)	Industry	Europe: Germany	5 (0)	November 2012–August 2016	NR' KM curve up to 800 days	January 2020
CheckMate 025[a]	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)	October 2012–March 2014	72 months	NR
METEOR	2L+	Cabo vs. evero	Parallel (open label)	Industry	Mixed: multiple	173 (11)	August 2013–November 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 follow-up	December 2014
RECORD-1[a]	2L+	Evero vs. placebo	Parallel (double blind)	Industry	Mixed: Australia, Canada, Europe, Japan, USA	86 (NR)	November 2006–November 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–August 2017	NR; KM up to 48 months	May 2021

TABLE 34 Study design characteristics of included trials (*continued*)

Trial name	Line	Comparison	Design (blinding)	Study sponsor	Continent: country	Number of centres (number of UK centres)	Enrolment period	Final follow-up	Date of last datacut
Deprioritised									
ASPEN	1L	Suni vs. evero	Parallel (open label)[a]	Industry and non-industry	Mixed: USA, Canada, UK	17 (6)	23 September 2010–28 October 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 June 2017–18 July 2019	Median 42.1 months (40.5–45.2)	NR
ESPN[a]	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)	June 2010–April 2011	4.5 years	December 2014
RECORD-3[a]	1L	Suni vs. evero	Crossover (open label)	Industry	Mixed: USA, Argentina, Australia, Brazil, Canada, Denmark, France, Germany, Hong Kong, Italy, South Korea, Mexico, the Netherlands, Peru, Spain, Taiwan (Province of China), Thailand, Turkey, UK	83 (3)	October 2009–June 2011	Median 3.7 years	May 2015
SWOG 1500	1L[a]	Cabo vs. suni	Parallel (open label)	Non-industry	North America: USA, Canada	65 (0)	April 2016–December 2019	NR; KM to 40 months	October 2020
SUNNIFORECAST	1L	Nivo + ipi vs. SoC	Parallel (open label)	Industry	Europe: Belgium, Czech Republic, France, Germany, the Netherlands, Spain, UK	30 (2)	November 2017–ongoing	NR	NR
VEG105192[a]	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)	April 2006–April 2007	Unclear	March 2010

a Crossover to the comparator permitted following progression.

TABLE 35 Population characteristics of included trials

		Key inclusion and exclusion criteria						Baseline characteristics					
Trial name	*N* (UK pts)	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 autoimmune disease and 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 glucocorticoid or immunosuppressants 7 days before randomisation	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
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

TABLE 35 Population characteristics of included trials (*continued*)

		Key inclusion and exclusion criteria						Baseline characteristics					
Trial name	*N* (UK pts)	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 nephre-ctomy
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
2L+													
AXIS	723 (NR)	≥ 18	CC	–	1[a]	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
METEOR	658 (26)	≥ 18	CC	–	≥ 1 TKI	KPS ≥ 70	Disease progression during or within 6 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, IL-2, or IFNα permitted Exclusion: untreated CNS mets	61, 27–85	100/NR	91	35	29; 56; 14	97

continued

TABLE 35 Population characteristics of included trials (*continued*)

		Key inclusion and exclusion criteria						Baseline characteristics					
Trial name	*N* (UK pts)	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 nephre-ctomy
TIVO-3	350 (NR)	≥ 18	CC	–	2 or 3[b]	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
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[c]	–	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
SUNNIFORECAST	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; sarcomatoid features NR	NR	NR	NR; NR; NR	NR

TABLE 35 Population characteristics of included trials (*continued*)

		Key inclusion and exclusion criteria						Baseline characteristics					
Trial name	*N* (UK pts)	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 nephre-ctomy
VEG105192	435 (NR)	≥ 18	CC	–	0 or 1[d]	0 or 1	Exclusion: CNS mets	NR for whole sample	100/NR	83.2	27.4	39; 54; 3	88.5

bev, bevacizumab; CC, clear cell; KPS, Karnofsky Performance Status; MET, mesenchymal-epithelial transition; mets, metastasis; nCC, non-clear cell; Pts, patients.

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

b One of which included a VEGFR TKI other than tivo or sora.

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

d 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).

TABLE 36 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	N/A	Cabo vs. suni	Cabo (orally): 60 mg OD Suni (orally): 50 mg OD for 4 weeks then 2-week break per cycle	NR	N/A	Int 60.8 Control 61.5
CheckMate 214	N/A	Nivo + ipi vs. suni	Nivo (IV): 3 mg/kg bodyweight over 60-minute period/3 weeks for four doses and then at a dose of 3 mg/kg bodyweight every 2 weeks Ipi (IV): 1 mg/kg bodyweight over a period of 30 minutes/3 weeks for four doses Suni (orally): 50 mg OD for 4 weeks, 2 weeks off per cycle Nivo or ipi dose reductions not allowed. Dose delays for AEs were permitted in both groups	Nivo induction: 79;[a] Nivo maintenance: (confidential information has been removed) Ipi: 79[a]	Treated beyond progression: Nivo + ipi *n* = 157 (29%), Suni *n* = 129 (24%)	Int 53.5 Control 66.5
CheckMate 9ER	N/A	Cabo + nivo vs. suni	Nivo (IV): 240 mg every 2 weeks and cabo(orally) 40 mg OD Suni (orally): 50 mg OD for 4 weeks, then 2-week break in 6-week cycle	Nivo: (confidential information has been removed) Cabo: (confidential information has been removed) Suni: NR	Nivo stopped after 2 years (from the first dose)	Int 25.1 Control 40.5
CLEAR	N/A	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 weeks on/2 weeks off) Dose reduction and interruptions: investigators decide the probability of the event being related to one or both drugs; lenv dose reduction to 14, 10 and 8 mg/day. Dose reductions < 8 mg/day must be discussed with sponsor	Median pem + lenv Len: 69.6% Median number of pem infusions per patient 22 (range, 1–39) 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
COMPARZ	N/A	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 AEs	NR	N/A	Int NR Control NR

TABLE 36 Intervention characteristics of included trials trials (*continued*)

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery, etc.)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
CROSS-J-RCC	N/A	Suni vs. sora	Suni (orally): 50 mg OD (4 weeks on/2 weeks off) Suni dose reductions to 37. 5 mg then 25 mg/day schedule 4/2. Dose reduction < 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	N/A	Ave + axi vs. suni	Ave + axi: ave (IV) 10 mg/kg every 2 weeks and axi (orally) 5 mg BID Suni (orally): 50 mg OD (4 weeks on/2 weeks off)	Ave 91.5%; Axi 89.4%; Sun 83.9% (all median)	N/A	Int 46.2 Control 60.6
SWITCH	N/A	Suni vs. sora	Suni (orally): 50 mg OD, 4 weeks on 2 weeks off; dose reductions permitted	NR	N/A	Int 57% crossed over Control 42% crossed over
SWITCH II	N/A	Pazo vs. sora	Pazo (orally) 800 mg OD, dose reductions permitted	NR	N/A	Int 64.0 Control 58.5
TIVO-1	N/A	Tivo vs. sora	Tivo (orally) 1.5 mg OD for 3 weeks followed by 1 week 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 150/90 mm Hg or the patient was receiving antihypertensive 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	Median 99% for axi and 92% for sora	Patients were treated until progression of disease (RECIST version 1.017), occurrence of unacceptable toxic effects, death, or withdrawal of patient consent	Int 54.4 Control 56.6
BERAT	TKI NR; IO NR	TKI (axi/suni) vs. evero	Axi: 5 mg BID starting dose Suni: 50 mg OD, 4–2 regimen Evero: 10 mg 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 intravenous (IV) infusion every 2 weeks 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

continued

TABLE 36 Intervention characteristics of included trials trials (*continued*)

Trial name	% Prior TKI; % prior IO (systemic)	Comparison	Treatment details (include dose, delivery, etc.)	RDI	Treatment stopping rules	Any subsequent systemic tx (% of ITT)
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/day + BSC Matching placebo plus BSC	NR	N/A	Int NR Control 79.9
TIVO-3	TKI 100%; IO/TKI tivo 27%, sora 25%	Tivo vs. sora	Tivo (orally): 1.5 mg OD in 4-week cycles comprising 21 days on treatment followed by 7 days off treatment. Dose reduction to 1.0 mg OD allowed for patients with treatment-related AEs ≥ grade 3. Dose interruptions allowed for persistent AEs	NR	N/A	Int 64.6 Control 58.5
Deprioritised trials						
ASPEN	TKI N/A; IO N/A	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 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-sday 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 weeks for 4 doses then IV nivo 240 mg every 2 weeks Nivo (IV): 240 mg every 2 weeks Suni (orally) 50 OD for 4 weeks every 6 weeks; pazo (orally) 800 mg OD continuously	NR	NR	Nivo: 62 Nivo + Ipi: 57.4 TKI: 50
ESPN	TKI N/A; IO N/A	Evero vs. suni	Evero 10 mg/day orally 4 weeks on and 2 weeks off; suni 50 mg/day orally 4 weeks on and 2 weeks 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-week 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 weeks 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

TABLE 36 Intervention characteristics of included trials trials (*continued*)

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 weeks on, 2 weeks off)	Evero 98%, suni 87%	N/A	Int 55 Control 51
SWOG 1500	N/A	Cabo vs. suni	Cabo (orally): 60 mg OD, dose reductions permitted Suni (orally) 50 mg 4 weeks on, 2 weeks 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 weeks for four doses followed by nivo fixed dose 240 mg IV every 2 weeks or fixed dose 480 mg IV every 4 weeks	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 NR)	NR	N/A	Int 30.3 Control 65.5

BID, twice daily; N/A, not applicable; OD, once daily; SoC, standard of care; TRAE, treatment-related adverse event.
a 79% reported to receive all four doses of nivo and ipi within the induction phase.

Notes
Dosing is only included for treatments which are part of the UK treatment pathway.

TABLE 37 Summary of domain-level risk of bias judgements, main issues per study and overall study-level risk of bias of included RCTs

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 end points, 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
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 end points, 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 end points, with methods to account for missing data unclear, some outcomes reported in the trial registry are 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

TABLE 37 Summary of domain-level risk of bias judgments, main issues per study and overall study-level risk of bias of included RCTs (*continued*)

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
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
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 end points, 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 end points 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 end points, 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

continued

TABLE 37 Summary of domain-level risk of bias judgments, main issues per study and overall study-level risk of bias of included RCTs (*continued*)

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 end points, 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 end points, 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 end points, 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 end points, with methods to account for missing data unclear, potential conflict from industry funding
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 end points, with methods to account for missing data unclear, unclear conflict as the trial was not industry-funded but some authors received industry funding

TABLE 38 Overall survival 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	5 years+	550	546	Int: 55.7 (NR); control: 38.4 (NR)	0.72 (0.62 to 0.85)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Overall	4–5 years	355	357	Int: 53.7 (48.7 to NE); control: 54.3 (40.9 to NE)	0.79 (0.63 to 0.99)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Overall	4–5 years	60	64	Int: 38.4 (NR); control: 30.9 (NR)	0.934 (0.588 to 1.485)
SWITCH	Eichelberg (2015)	Sora	Suni	Overall	1–2 years	182	183	Int: 30 (NR); control: 27.4 (NR)	0.99 (0.73 to 1.33)
SWITCH II	Retz (2019)	Sora	Pazo	Overall	3–4 years	189	188	Int: NR; control: NR	1.22 (0.91 to 1.65)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Overall	2–3 years	442	444	Int: NE (42.2, NE); control: 37.8 (31.4 to NE)	0.79 (0.64 to 0.97)
COMPARZ	Motzer (2014)	Pazo	Suni	Overall	2–3 years	557	553	Int: 28.3 (26.0 to 35.5); control: 29.1 (25.4 to 33.1)	0.92 (0.79 to 1.06)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Overall	3–4 years	323	328	Int: 49.48 (40.31 to NE); control: 35.52 (29.24 to 42.25)	0.7 (0.56 to 0.87)
TIVO-1	Motzer (2013)	Tivo	Sora	Overall	2–3 years	260	257	Int: 29.3 (NR); control: 28.8 (NR)	1.245 (0.95 to 1.62)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	Fav	5 years+	125	124	Int: 74.1 (NR); control: 68.4 (NR)	0.94 (0.65 to 1.37)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Fav	4–5 years	110	124	Int: not reached (NR); control: 59.9 (58.8 to NE)	0.94 (0.58 to 1.52)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Fav	4–5 years	12	14	Int: NR; control: NR	0.35 (0.1 to 1.2)
SWITCH	Eichelberg (2015)	Sora	Suni	Fav	1–2 years	71	82	Int: NR; control: NR	1.24 (0.61 to 2.56)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Fav	2–3 years	94	96	Int: NE (NE, NE); control: NE (39.8 to NE)	0.66 (0.36 to 1.22)
COMPARZ	Motzer (2014)	Pazo	Suni	Fav	2–3 years	151	152	Int: 42.5 (37.9 to NE); control: 43.6 (37.1 to 47.4)	0.88 (0.63 to 1.21)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Fav	3–4 years	74	72	Int: NE (40.67 to NE); control: 47.61 (43.63 to NE)	1.07 (0.63 to 1.79)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	Int/poor	5 years+	425	422	Int: 47 (NR); control: 26.6 (NR)	0.68 (0.58 to 0.81)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Int/poor	4–5 years	243	229	Int: 47.9 (40.5 to NE); control: 34.3 (26.3 to 54.3)	0.74 (0.57 to 0.96)
CROSS-J-RCC	Tomita *et al.* (2020)	Suni	Sora	Int/poor	4–5 years	45	49	Int: NR; control: NR	1.2 (0.7 to 1.95)
SWITCH	Eichelberg (2015)	Sora	Suni	Int/poor	1–2 years	108	94	Int: NR; control: NR	0.83 (0.53 to 1.31)

continued

TABLE 38 Overall survival in prioritised included trials (*continued*)

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)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Int/poor	2–3 years	343	347	Int: 42.2 (33.1 to NE); control: 37.8 (29.6 to NE)	0.79 (0.64 to 0.98)
COMPARZ	Motzer (2014)	Pazo	Suni	Int/poor	2–3 years	389	380	Int: NR; control: NR	0.891 (0.75 to 1.06)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Int/poor	3–4 years	249	256	Int: 49.5 (34.9 to NE); control: 29.2 (23.7 to 36.0)	0.65 (0.51 to 0.83)
CABOSUN	Choueiri *et al.* (2018)	Cabo	Suni	Int/poor	2–3 years	79	78	Int: 26.6 (14.6 to NE); control: 21.2 (16.3 to 27.4)	0.8 (0.53 to 1.21)
2L+									
AXIS	Motzer (2013)	Axi	Sora	Overall	3–4 years	361	362	Int: 20.1 (16.7 to 23.4); control: 19.2 (17.5 to 22.3)	0.969 (0.8 to 1.174)
BERAT	Grünwald (2022)	Evero	Axi	Overall	1–2 years	5	5	Int: 15.29 (6.0 to NE); control: 18.64 (5.9 to 32.5)	1.12 (0.27 to 4.61)
CheckMate 025	Escudier (2022)	Nivo	Evero	Overall	5 years+	410	411	Int: NR; control: NR	0.74 (0.63 to 0.86)
METEOR	Choueiri *et al.* (2016)	Cabo	Evero	Overall	1–2 years	330	328	Int: 21.4 (18.7 to NE); control: 16.5 (14.7 to 18.8)	0.66 (0.53 to 0.83)
NCT01136733	Motzer *et al.* (2015)	Lenv + evero	Evero	Overall	2–3 years	51	50	Int: 25.5 (16.4 to NE); control: 15.4 (11.8 to 19.6)	0.51 (0.3 to 0.88)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	1–2 years	277	139	Int: 14.8 (NR); control: 14.4 (NR)	0.87 (0.65 to 1.15)
TIVO-3	Rini (2022)	Tivo	Sora	Overall	1–2 years	175	175	Int: NR; control: NR	0.89 (0.7 to 1.14)

NE, not estimable; PBO, placebo.

TABLE 39 Progression-free survival 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	5 years+	550	546	Int: NR; control: NR	0.86 (0.73 to 1.01)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Overall	BICR	RECIST, FDA rule	3–4 years	323	328	Int: 16.6 (12.8 to 19.5); control: 8.4 (7.0 to 9.7)	0.59 (0.49 to 0.71)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Overall	ICR (no blinding)	RECIST, FDA rule	4–5 years	355	357	Int: 23.9 (20.8 to 27.7); control: 9.2 (6.0 to 11.0)	0.47 (0.38 to 0.57)
COMPARZ	Motzer (2013)	Pazo	Suni	Overall	BICR	RECIST, FDA rule	1–2 years	557	553	Int: 8.4 (8.3 to 10.9); control: 9.5 (8.3 to 11.1)	1.05 (0.9 to 1.22)
COMPARZ	Motzer (2013)	Pazo	Suni	Overall	IA	RECIST, FDA rule	1–2 years	557	553	Int: 10.5 (8.3 to 11.1); control: 10.2 (8.3 to 11.1)	1 (0.86 to 1.15)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Overall	IA	RECIST, censoring rules unclear	4–5 years	60	64	Int: 8.7 (5.5 to 21.1); control: 7 (6.1 to 12.2)	0.67 (0.42 to 1.08)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Overall	BICR	RECIST, FDA rule	2–3 years	442	444	Int: 13.9 (11.1 to 16.6); control: 8.5 (8.2 to 9.7)	0.67 (0.57 to 0.79)
SWITCH	Eichelberg (2015)	Sora	Suni	Overall	IA	RECIST, FDA rule	< 1 year	182	183	Int: 5.9 (NR); control: 8.5 (NR)	1.19 (0.93 to 1.53)
SWITCH II	Retz (2019)	Sora	Pazo	Overall	IA	RECIST, censoring rules unclear	3–4 years	189	188	Int: 5.6 (4.7 to 6.3); control: 9.3 (7.4 to 10.6)	1.51 (1.19 to 1.92)
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 to 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 to 2.26)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Fav	BICR	RECIST, FDA rule	3–4 years	74	72	Int: 21.42 (13.08 to 24.71); control: 13.86 (9.56 to 16.66)	0.72 (0.49 to 1.05)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Fav	ICR (no blinding)	RECIST, FDA rule	4–5 years	110	124	Int: NR; control: NR	0.5 (0.35 to 0.71)
COMPARZ	Motzer (2013)	Pazo	Suni	Fav	BICR	RECIST, FDA rule	1–2 years	151	152	Int: NR; control: NR	1.01 (0.74 to 1.37)
CROSS-J-RCC	Tomita *et al.* (2020)	Suni	Sora	Fav	IA	RECIST, censoring rules unclear	4–5 years	12	14	Int: NR; control: NR	0.245 (0.082 to 0.734)

continued

TABLE 39 Progression-free survival in prioritised included trials (*continued*)

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)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Fav	BICR	RECIST, FDA rule	2–3 years	94	96	Int: 20.7 (16.6 to 26.3); control: 13.8 (11.1 to 23.5)	0.71 (0.49 to 1.016)
SWITCH	Eichelberg (2015)	Sora	Suni	Fav	IA	RECIST, FDA rule	< 1 year	71	82	Int: NR; control: NR	1.3 (0.81 to 2.09)
TIVO-1	Motzer (2013)	Tivo	Sora	Fav	BICR	RECIST, FDA rule	NR	70	87	Int: NR; control: NR	0.59 (0.378 to 0.921)
CABOSUN	Choueiri *et al.* (2018)	Cabo	Suni	Int/poor	BICR	RECIST, FDA rule	2–3 years	79	78	Int: 8.6 (6.8 to 14.0); control: 5.3 (3.0 to 8.2)	0.48 (0.31 to 0.74)
CABOSUN	Choueiri *et al.* (2018)	Cabo	Suni	Int/poor	IA	RECIST, FDA rule	2–3 years	79	78	Int: 8.3 (6.5 to 12.4); control: 5.4 (3.4 to 8.2)	0.56 (0.37 to 0.83)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	Int/poor	BICR	RECIST, FDA rule	5 years+	425	422	Int: NR; control: NR	0.73 (0.61 to 0.87)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	Int/poor	BICR	RECIST, FDA rule	RECIST, FDA rule	249	256	Int: 15.61 (11.17 to 19.15); control: 7.05 (5.68 to 8.90)	0.56 (0.46 to 0.69)
CLEAR	Motzer (2023)	Pem + lenv	Suni	Int/poor	ICR (no blinding)	RECIST, FDA rule	4–5 years	243	229	Int: NR; control: NR	0.43 (0.34 to 0.55)
COMPARZ	Motzer (2013)	Pazo	Suni	Int/poor	BICR	RECIST, FDA rule	1–2 years	322	328	Int: NR; control: NR	0.98 (0.80 to 1.19)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	Int/poor	IA	RECIST, censoring rules unclear	4–5 years	45	49	Int: NR; control: NR	1 (0.62 to 1.63)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	Int/poor	BICR	RECIST, FDA rule	2–3 years	343	347	Int: 12.9 (11.1 to 16.6); control: 8.4 (7.9 to 10.1)	0.66 (0.55 to 0.787)
SWITCH	Eichelberg (2015)	Sora	Suni	Int/poor	IA	RECIST, FDA rule	< 1 year	108	94	Int: NR; control: NR	1.14 (0.77 to 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 to 1.062)
2L+											
AXIS	Motzer (2013)	Axi	Sora	Overall	BICR	RECIST, censoring rules unclear	3–4 years	361	362	Int: 8.3 (6.7 to 9.2); control: 5.7 (4.7 to 6.5)	0.66 (0.55 to 0.78)
BERAT	Grünwald (2022)	Evero	Axi	Overall	IA	RECIST, censoring rules unclear	1–2 years	5	5	Int: 3.7 (2.6 to 8.4); control: 2.2 (1.9 to NC)	1 (0.26 to 3.85)
CheckMate 025	Escudier (2022)	Nivo	Evero	Overall	IA	RECIST, FDA rule	5 years+	410	411	Int: NR; control: NR	0.84 (0.72 to 0.99)

TABLE 39 Progression-free survival in prioritised included trials (*continued*)

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)
METEOR	Choueiri (2016)	Cabo	Evero	Overall	BICR	RECIST, FDA rule	< 1 year	330	328	Int: 7.4 (6.6 to 9.1); control: 3.9 (3.7 to 5.1)	0.51 (0.41 to 0.62)
METEOR	Choueiri (2016)	Cabo	Evero	Overall	IA	RECIST, FDA rule	< 1 year	330	328	Int: 7.4 (6.6 to 9.1); control: 5.1 (3.9 to 5.5)	0.54 (0.44 to 0.65)
NCT01136733	Motzer (2015)	Lenv + evero	Evero	Overall	IA	RECIST, censoring rules unclear	1–2 years	51	50	Int: 14.6 (5.9 to 20.1); control: 5.5 (3.5 to 7.1)	0.4 (0.24 to 0.68)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	BICR	RECIST, FDA rule	1–2 years	277	139	Int: 4.9 (4.0 to 5.5); control: 1.9 (1.8 to 1.9)	0.33 (0.25 to 0.43)
RECORD-1	Motzer (2010)	Evero	PBO	Overall	IA	RECIST, FDA rule	1–2 years	277	139	Int: 5.5 (4.6 to 5.8); control: 1.9 (1.8 to 2.2)	0.32 (0.25 to 0.41)
TIVO-3	Atkins (2022)	Tivo	Sora	Overall	IA	RECIST, FDA rule	1–2 years	175	175	Int: NR; control: NR	0.624 (0.49 to 0.79)
TIVO-3	Rini (2020)	Tivo	Sora	Overall	BICR	RECIST, FDA rule	1–2 years	175	175	Int: 5.6 (5.29 to 7.33); control: 3.9 (3.71 to 5.55)	0.73 (0.56 to 0.94)

TABLE 40 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	< 1 year	Overall	IA	182	183	30.22	27.87	1.12 (0.71 to 1.76)
COMPARZ	Motzer (2013)	Pazo	Suni	1–2 years	Overall	BICR	557	553	30.70	24.77	1.35 (1.03 to 1.75)
COMPARZ	Motzer (2013)	Pazo	Suni	1–2 years	Overall	IA	557	553	33.39	28.93	1.23 (0.95 to 1.59)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	2–3 years	Overall	IA	442	444	59.30	31.80	3.13 (2.37 to 4.12)
SWITCH II	Retz (2019)	Sora	Pazo	3–4 years	Overall	IA	189	188	28.57	46.28	0.46 (0.30 to 0.71)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	3–4 years	Overall	BICR	323	328	56.04	28.05	3.27 (2.36 to 4.53)
CLEAR	Motzer (2023)	Pem + lenv	Suni	4–5 years	Overall	BICR	355	357	71.30	36.70	4.28 (3.12 to 5.86)
CROSS-J-RCC	Tomita *et al.* (2020)	Suni	Sora	4–5 years	Overall	Unclear	60	64	23.33	15.63	1.64 (0.67 to 4.05)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	5 years+	Overall	BICR	550	546	39	32.00	1.36 (1.06 to 1.74)
CLEAR	Grünwald (2021)	Pem + lenv	Suni	2–3 years	Fav	BICR	74	72	68.20	50.80	1.97 (1.01 to 3.86)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	2–3 years	Fav	IA	94	96	75.50	45.80	3.65 (1.97 to 6.77)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	3–4 years	Fav	BICR	74	72	67.57	45.83	(confidential information has been removed)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	5 years+	Fav	BICR	125	124	30.00	52.00	0.41 (0.24 to 0.69)
CLEAR	Grünwald (2021)	Pem + lenv	Suni	2–3 years	Int/poor	BICR	188	188	72.40	28.80	6.51 (4.15 to 10.20)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	2–3 years	Int/poor	IA	343	347	55.10	28.00	3.16 (2.30 to 4.34)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	3–4 years	Int/poor	BICR	249	256	52.61	23.05	(confidential information has been removed)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	5 years+	Int/poor	BICR	425	422	42.00	27.00	1.97 (1.47 to 2.62)
CABOSUN	Choueiri (2018)	Cabo	Suni	2–3 years	Overall/ int/poor	BICR	79	78	20.25	8.97	2.58 (1.00 to 6.67)
CABOSUN	Choueiri (2018)	Cabo	Suni	2–3 years	Overall/ int/poor	IA	79	78	32.91	11.54	3.76 (1.63 to 8.70)

TABLE 40 Response rates in prioritised included trials (*continued*)

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)
TIVO-1[a]	Motzer (2013)	Tivo	Sora	NR	Overall	BICR	260	257	33.10	23.30	1.62 (1.10 to 2.39)
TIVO-1[a]	Motzer (2013)	Tivo	Sora	NR	Overall	IA	260	257	35.40	30.70	1.23 (0.85 to 1.78)
2L+											
METEOR	Choueiri (2016)	Cabo	Evero	< 1 year	Overall	BICR	330	328	17.27	3.35	6.02 (3.09 to 11.71)
METEOR	Choueiri (2016)	Cabo	Evero	< 1 year	Overall	IA	330	328	23.64	4.27	6.94 (3.84 to 12.56)
AXIS	Rini (2011)	Axi	Sora	1–2 years	Overall	BICR	361	362	19.39	9.39	2.32 (1.50 to 3.60)
NCT01136733	Motzer (2015)	Lenv + evero	Evero	1–2 years	Overall	IA	51	50	43.14	6.00	11.89 (3.26 to 43.26)
RECORD-1	Motzer (2010)	Evero	Placebo	1–2 years	Overall	BICR	277	139	1.81	0.00	5.63 (0.31 to 102.6)
TIVO-3	Verzoni (2021)	Tivo	Sora	1–2 years	Overall	IA	175	175	23.43	11.43	2.37 (1.32 to 4.25)
AXIS	Motzer (2013)	Axi	Sora	3–4 years	Overall	IA	361	362	22.71	12.43	2.07 (1.39 to 3.08)
CheckMate 025	Motzer (2020)	Nivo	Evero	5 years+	Overall	IA	410	411	22.93	4.14	6.89 (4.03 to 11.80)
BERAT	Grünwald (2022)	Evero	Axi	NR ('short')	Overall	IA	5	5	0.00	20.00	0.27 (0.01 to 8.46)

a 1L and 2L.

TABLE 41 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	2–3 years	260	141	Overall	IA	19.4 (15.2 to 22.3)	14.5 (8.8 to 17.1)	
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4–5 years	60	64	Overall	Unclear	32.0	14.9	
CheckMate 9ER	CS	Cabo + nivo	Suni	3–4 years	181	92	Overall	BICR	22.08 (17.97 to 26.02)	16.07 (11.07 to 19.35)	
CLEAR	Motzer 2023	Pem + lenv	Suni	4–5 years	253	131	Overall	BICR	26.7 (22.8 to 34.6)	14.7 (9.4 to 18.2)	
CheckMate 214	Motzer 2022	Nivo + ipi	Suni	5 years+	550	546	Overall	BICR	Not reached (59.0 to NE)	24.8 (19.7 to 30.1)	0.49 (0.35 to 0.68)
JAVELIN Renal 101	Haanen 2023	Ave + axi	Suni	2–3 years	71	44	Fav	IA	22.6 (15.2 to 31.7)	20.8 (14.5 to 24.9)	
JAVELIN Renal 101	Haanen 2023	Ave + axi	Suni	2–3 years	189	97	Int/poor	IA	19.3 (13.9 to 22.1)	9.8 (7.0 to 15.3)	
2L+											
AXIS	Rini 2011	Axi	Sora	1–2 years	361	362	Overall	NR	11 (7.4 to NE)	10.6 (8.8 to 11.5)	
NCT01136733	Motzer 2015	Lenv + evero	Evero	1–2 years	51	50	Overall	NR	13 (3.7 to NE)	8.5 (7.5 to 9.4)	
TIVO-3	Verzoni 2021	Tivo	Sora	1–2 years	175	175	Overall	IA	20.3 (9.8 to 29.9)	9 (3.7 to 16.6)	
CheckMate 025	Motzer 2020	Nivo	Evero	5 years+	410	411	Overall	NR	18.2 (12.9 to 25.8)	14 (8.3 to 19.2)	

TABLE 42 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	3–4 years	263	288	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)
CheckMate 9ER	Company clarification response	Cabo + nivo	Suni	Fav	1L	3–4 years	60	57	NR	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)
CheckMate 9ER	Company clarification response	Cabo + nivo	Suni	Int/ poor	1L	3–4 years	203	231	NR	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)
CheckMate 214	Stakeholder submission	Nivo + ipi	Suni	Int/ poor	1L	5 years+	423	416	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	(confidential information has been removed)	NR	NR

con, control; int, intermediate/intervention; NE, non-evaluable.

TABLE 43 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	< 1 year	177	176	Overall	Mean 8.7 months (SD 8.6)	Mean 10.1 months (SD 10.2)
COMPARZ	Motzer	2013	Pazo	Suni	1–2 years	557	553	Overall	Median 8 (range 0–38)	Median 7.6 (range 0–38)
SWITCH II	Retz	2019	Sora	Pazo	3–4 years	189	188	Overall	Median 2.1 (range 0.3–21.4)	Median 5.7 (range 0.3–43.3)
CheckMate 9ER	CS	2023	Cabo + nivo	Suni	3–4 years	323	328	Overall	Median 21.8 (IQR 8.8–34.0)	Median 8.9 (IQR 2.9–20.7)
CheckMate 9ER	CS	2023	Cabo + nivo	Suni	3–4 years	323	328	Overall	(confidential information has been removed)	(confidential information has been removed)
CROSS-J-RCC	Tomita	2020	Suni	Sora	4–5 years	60	64	Overall	Median 6.7 (95% CI NR);	Median 5.9 (95% CI NR);
CheckMate 214	Motzer	2022	Nivo + ipi	Suni	5 years+	550	546	Overall	Median 7.9 (IQR 2.1 to 21.8)	Median 7.8 (IQR 3.5 to 19.6)
CLEAR	Motzer	2023	Pem + lenv	Suni	NR	355	357	Overall	Median 17 (95% CI 9.4 to 25.4)	Median 7.8 (95% CI 3.7 to 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	CS	2023	Cabo + nivo	Suni	3–4 years	74	71	Fav	(confidential information has been removed)	(confidential information has been removed)
CABOSUN	Choueiri	2018	Cabo	Suni	2–3 years	78	72	Int/poor	Median 8.39 (95% CI 5.72 to 8.39)	Median 7.09 (95% CI 5.09 to 6.68)
CheckMate 9ER	CS	2023	Cabo + nivo	Suni	3–4 years	246	249	Int/poor	(confidential information has been removed)	(confidential information has been removed)
CheckMate 214	Stakeholder submission	2023	Nivo + ipi	Suni	5 years+	423	416	Int/poor	(confidential information has been removed)	(confidential information has been removed)
2L+										
RECORD-1	Motzer	2010	Evero	Placebo	1–2 years	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	1–2 years	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	3–4 years	361	362	Overall	Mean 8.2 (SD NR, range < 0.1–33.4)	Mean 5.2 (SD NR, range 0.2–34.1)

IQR, interquartile range; SD, standard deviation.
a Time to discontinuation.

TABLE 44 Discontinuation due to AEs 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	< 1 year	182	183	Overall	18.13	28.42	0.56 (0.34 to 0.92)
CLEAR	Motzer (2021)	Pem + lenv	Suni	1–2 years	355	357	Overall	16.90	11.48	1.57 (1.02 to 2.40)
COMPARZ	Motzer (2013)	Pazo	Suni	1–2 years	557	553	Overall	24.24	20.25	1.26 (0.95 to 1.67)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	2–3 years	442	444	Overall	31.22	15.99	2.38 (1.73 to 3.30)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	2–3 years	323	328	Overall	36.84	20.43	2.27 (1.60 to 3.23)
SWITCH II	Retz (2019)	Sora	Pazo	3–4 years	189	188	Overall	32.28	23.40	1.56 (0.99 to 2.46)
SWOG 1500	Pal (2021)	Cabo	Suni	3–4 years	44	46	Overall	22.73	23.91	0.94 (0.35 to 2.49)
CROSS-J-RCC	Tomita *et al.* (2020)	Suni	Sora	4–5 years	60	64	Overall	21.67	18.75	1.20 (0.50 to 2.88)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	5 years+	550	546	Overall	34.18	19.41	2.16 (1.64 to 2.84)
TIVO-1	Motzer (2013)	Tivo	Sora	NR	260	257	Overall	7.31	7.00	1.05 (0.54 to 2.04)
CABOSUN	Choueiri *et al.* (2018)	Cabo	Suni	NR	79	78	Int/poor	20.25	20.51	0.98 (0.45 to 2.14)
2L+										
METEOR	Choueiri *et al.* (2016)	Cabo	Evero	1–2 years	330	328	Overall	12.12%	10.37%	1.19 (0.73 to 1.94)
NCT01136733	Motzer *et al.* (2015)	Lenv + evero	Evero	1–2 years	51	50	Overall	17.65	10.00	1.93 (0.60 to 6.22)
RECORD-1	Motzer (2010)	Evero	Placebo	1–2 years	277	139	Overall	13.00	1.44	10.23 (2.43 to 43.16)
TIVO-3	Zengin (2020)	Tivo	Sora	1–2 years	175	175	Overall	20.57	29.71	0.61 (0.38 to 1.00)
AXIS	Motzer (2013)	Axi	Sora	3–4 years	361	362	Overall	7.48	12.43	0.57 (0.34 to 0.94)
CheckMate 025	Motzer (2020)	Nivo	Evero	5 years+	410	411	Overall	13.90	16.06	0.84 (0.57 to 1.24)
BERAT	Grünwald (2022)	Evero	Axi	NR ('short')	5	5	Overall	0.00	0.00	–

con, control; int, intervention.

TABLE 45 Grade 3+ AEs 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	< 1 year	182	183	Overall	64.29	64.48	0.99 (0.65 to 1.52)
CLEAR	Motzer (2021)	Pem + lenv	Suni	1–2 years	355	357	Overall	81.69	68.35	2.07 (1.46 to 2.93)
COMPARZ	Motzer (2013)	Pazo	Suni	1–2 years	557	553	Overall	73.97	72.69	1.07 (0.82 to 1.39)
JAVELIN Renal 101	Haanen (2023)	Ave + axi	Suni	2–3 years	442	444	Overall	79.64	76.58	1.20 (0.87 to 1.65)
SWITCH II	Retz (2019)	Sora	Pazo	3–4 years	189	188	Overall	57.14	62.23	0.81 (0.54 to 1.22)
SWOG 1500	Pal (2021)	Cabo	Suni	3–4 years	44	46	Overall	72.73	67.39	1.29 (0.52 to 3.19)
CheckMate 9ER	CS (2023)	Cabo + nivo	Suni	3–4 years	323	328	Overall	(confidential information has been removed)	(confidential information has been removed)	1.70 (1.16 to 2.49)
CROSS-J-RCC	Tomita (2020)	Suni	Sora	4–5 years	60	64	Overall	83.33	75.00	1.67 (0.69 to 4.03)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	5 years+	550	546	Overall	67.82	76.23	0.66 (0.50 to 0.86)
TIVO-1	Motzer (2013)	Tivo	Sora	NR	260	257	Overall	61.15	69.65	0.69 (0.48 to 0.99)
CABOSUN	Choueiri (2018)	Cabo	Suni	NR	79	78	Int/poor	67.09	60.26	1.34 (0.70 to 2.58)
2L+										
METEOR	Choueiri (2016)	Cabo	Evero	1–2 years	330	328	Overall	71.21	58.84	1.73 (1.25 to 2.39)
NCT01136733	Motzer (2015)	Lenv + evero	Evero	1–2 years	51	50	Overall	70.59	50.00	2.40 (1.06 to 5.44)
CheckMate 025	Motzer (2020)	Nivo	Evero	5 years+	410	411	Overall	21.40	36.80	0.47 (0.34 to 0.64)
BERAT	Grünwald (2022)	Evero	Axi	NR ('short')	5	5	Overall	40.00	80.00	0.17 (0.01 to 2.82)

con, control; int, intervention.

TABLE 46 Health-related quality-of-life 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														
CLEAR	Motzer (2022)	Pem + lenv	Suni	All	Disease specific HRQoL	FKSI–DRS. Mean change, LS mean difference	< 1 year	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 to 2.00)
COMPARZ	Motzer (2013)	Pazo	Suni	All	Disease-specific HRQoL	FKSI total score. Difference in mean change score intervention vs. control	1–2 years	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	1–2 years	323	328	58.74 (10.57)	58.39 (9.92)	NR	NR	2.38 (1.20 to 3.56)
SWITCH II	Retz (2019)	Sora	Pazo	All	Disease-specific HRQoL	FKSI-10	3–4 years	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	5 years+	550	546	60.1	59.1	Mean: 0.36 (SD NR)	Mean: −1.51 (SD NR)	1.87 (0.95 to 2.79)
CLEAR	Motzer (2022)	Pem + lenv	Suni	All	Generic HRQoL	EQ5D-Index, mean change, LS mean difference	< 1 year	355	357	0.83 (0.19)	0.81 (0.22)	Mean: −4 (SE 0.9)	Mean: −6 (SE 1.1)	2 (0 to 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	1–2 years	323	328	0.78 (0.25)	0.73 (0.29)	NR	NR	0.04 (0.01 to 0.07)
CheckMate 214	Cella (2020)	Nivo + ipi	Suni	All	Generic HRQoL	EQ-5D VAS LS mean using MMRM	5 years+	550	546	NR	NR	NR	NR	2.4 (0.4 to 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	< 1 year	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	1–2 years	74	72	NR	NR	NR	NR	−0.44 (−2.63, 1.75)

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	Fav	Generic HRQoL	EQ5D-Index, mean change, LS mean difference	< 1 year	110	124	NR	NR	Mean: −8 (SE 1.4)	Mean: −6 (SE 1.5)	−2 (−6 to 2)
CLEAR	Motzer (2022)	Pem + lenv	Suni	Int/ poor	Disease-specific HRQoL	FKSI–DRS. Mean change, LS mean difference	< 1 year	243	229	NR	NR	Mean: −0.72 (SE 0.86)	Mean: −1.42 (SE 0.96)	0.67 (−1.25 to 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	1–2 years	249	256	NR	NR	NR	NR	3.33 (1.96 to 4.70)
CheckMate 214	Motzer (2022)	Nivo + ipi	Suni	Int/ poor	Disease-specific HRQoL	FKSI-19 LS mean change	5 years+	425	422	NR	NR	Mean: 0.9 (SD NR)	Mean: −1.75 (SD NR)	2.65 (1.60 to 3.70)
CLEAR	Motzer (2022)	Pem + lenv	Suni	Int/ poor	Generic HRQoL	EQ5D-Index, mean change, LS mean difference	< 1 year	243	229	NR	NR	Mean: −3 (SE 1.5)	Mean: −7 (SE 1.7)	5 (1 to 8)
CheckMate 214	Cella (2020)	Nivo + ipi	Suni	Int/ poor	Generic HRQoL	EQ-5D VAS LS mean using MMRM	5 years+	425	422	NR	NR	NR	NR	3.3 (1.0 to 5.6)
2L+														
METEOR	Cella (2018)	Cabo	Evero	All	Disease-specific HRQoL	FKSI-19 LS mean change	< 1 year	324	313	NR	NR	Mean: −3.483 (SD NR)	Mean: −2.214 (SD NR)	−1.269 (−1.864 to −0.675)
AXIS	Motzer (2013)	Axi	Sora	All	Disease-specific HRQoL	FKSI-15	1–2 years	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	1–2 years	361	343	30.2 (4.4)	30.8 (4.8)	NR	NR	1.6 (1.4 to 1.9)
BERAT	Grünwald (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	< 1 year	323	314	NR	NR	Mean: −0.02 (SD NR)	Mean: −0.02 (SD NR)	−0.002 (−0.018 to 0.014)
CheckMate 025	Cella (2016)	Nivo	Evero	All	Generic HRQoL	EQ-5D mean change	1–2 years	361	344	0.78 (0.24)	0.78 (0.21)	NR	NR	0.04 (0.02 to 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

BL, baseline; con, control; int, intervention; NR, not reported; SD, standard deviation; SE, standard error.

Appendix 3 Real-world evidence characteristics and results

Summary of included studies

TABLE 47 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[28]	Multicentre UK retrospective analysis; patient-level data	UK (17)	1 January 2018–23 August 2022	Metastatic (*N* = 1319)	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 *et al.* (2020)[73] Full text	Retrospective (longitudinal) cohort	England (2)	1 January 2008–31 December 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 *et al.* (2016) (RECCORD)[6]	Registry data (RECCORD). Retrospective non-interventional study	UK (7 : 5 in England; 1 in Wales and 1 in Scotland)	March 2009–November 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 *et al.* (2021)[71]	Retrospective cohort	England (NR)	1 January 2011–31 January 2020 (CAS)	Advanced (*N* = 1485)	2L+[a]	Cabo; axi	Treatment patterns; OS
Hack *et al.* (2019)[72]	Retrospective cohort	England (3)	February 2016 and April 2019	Advanced (*N* = 109)	2L+[b]	Nivo	PFS; OS
Hilser *et al.* (2023)[77] Conference abstract	Retrospective non-interventional cohort	Germany (8)	NR	mRCC (*N* = 67)	1L	Cabo + nivo	Risk scores (Heng); PFS; OS; TTP
Nathan *et al.* (2022)[70] Conference abstract	Prospective cohort	UK (4)	After 1 August 2019	Advanced (*N* = 36)	1L	Ave + axi	Risk score (IMDC); PFS; OS
Nathan *et al.* (2023)[78] (CARINA: NCT04957160) Conference abstract + poster presentation	Retrospective, non-interventional cohort using CAS	England (6)	NR	Advanced (*N* = 129) (cabo subgroup *N* = 87)	2L[c]	Any + subgroup analysis of 2L cabo	Treatment patterns; treatment discontinuation
NCRAS 2023[58]	UK Registry data (OS for mRCC collected from 2013 to 2019)	UK (England)	2013–9	Advanced and metastatic (*N* = 18,421)	1L+	Various	OS
IQVIA 2022[79]	Hospital pharmacy audit data	UK (England)	NR	RCC-treated patients	1L+	(confidential information has been removed)	Treatment patterns
Kidney Cancer UK (audit of kidney cancer services in England)[75]	Audit data	UK	1 January 2017–December 2018	Advanced and metastatic (*N* = 18,421)	1L+	Various	Postoperative 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

TABLE 47 Summary of study characteristics of included RWE (*continued*)

Study name	Study type	Country (number of centres)	Study period	Population	LOT	Interventions	Outcomes evaluated (per PICOS)
NICE TA780:[23] SACT data report	Part of TA780 committee papers	England	5 April 2019 and 30 November 2020	Advanced (*N* = 814)	2L	Any post-1L treatment with nivo + ipi	Risk scores (IMDC); treatment patterns; OS; TTNT; treatment discontinuation

CAS, Cancer Analysis System; IO, immuno-oncology; LOT, line of treatment; tem, temsirolimus; NR, not reported.

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 first-line treatment in UK clinical practice.

TABLE 48 Summary of baseline characteristics of included RWE

Study name	Intervention	LOT	*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[28]	Cabo; suni; pazo; tivo; nivo; evero; axi; ave + axi; lenv + evero; lenv + pem; cabo + nivo; nivo + ipi; nivo	1L: 687(52%); 2L: 415 (32%);[a] 3L: 168 (13%);[a] 4L 42 (3%); 5L: 7 (0.5%)	1319	Mean 64.43 (min 21, max 90; SE 0.28)	936 (71%)	NR	Clear cell: 1092 (82.8%); chromophobe: 11 (< 1%); papillary 69 (5.2%); sarcomatoid 7 (0.5%); undifferentiated 6 (< 1%); other 53 (< 1%); missing/N/A 81 (< 1%)	Fav 294 (22.3%); Int/poor 1016 (77.0%); missing 9 (< 1%)	715 (54.2)
Hawkins *et al.* (2020)[73]	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

continued

TABLE 48 Summary of baseline characteristics of included RWE (*continued*)

Study name	Intervention	LOT	*N*	Age years, median (range)	Male *n* (%)	ECOG PS *n* (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) *n* (%)	Prior nephrectomy *n* (%)
Wagstaff *et al.* (2016) (RECCORD)[6]	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%)[b]	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[a]	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[a]	NR	NR	NR	NR	NR	NR
NCRAS 2023[58]	NR	NR	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]	NR[c]
IQVIA 2022[79]	(confidential information has been removed)	1L+	NR	NR	NR	NR	NR	NR	NR
Kidney Cancer UK (audit of kidney cancer services in England)[75]	NR	1L+	18-,421	68 (58, 77)	11,818 (63.4)	NR	NR	NR	NR
NICE TA780:[23] SACT data report	Nivo + ipi	1L	814	61 (NR) < 40 to 80+ years[b]	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 *et al.* (2021)[71]	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	1045	63.0 (NR)	556 (53.2%)	0–1 : 213 (20.4%)	NR	NR	NR
Hack *et al.* (2019)[72]	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)

TABLE 48 Summary of baseline characteristics of included RWE (*continued*)

Study name	Intervention	LOT	*N*	Age years, median (range)	Male *n* (%)	ECOG PS *n* (%)	Histology (% clear cell; % sarcomatoid)	IMDC (fav; int; poor) *n* (%)	Prior nephrectomy *n* (%)
Hilser *et al.* (2023)[77]	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 *et al.* (2022)[70]	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
Nathan *et al.* (2023)[78] (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

IO, immuno-oncology; NR, not reported; SD, standard deviation.

Notes

a One additional patient was denoted as receiving second-line, third-line and fourth-line treatments, but no treatment type was specified.
b < 40 years: 15 (2%); 40–49 years: 96 (12%); 50–59 years: 257 (32%); 60–69 years: 271 (33%); 70–79 years: 167 (21%); 80+ years 8 (1%).
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).

TABLE 49 Outcomes reported in the RWE

Trial name	Risk scores	OS + prognostic variables	PFS + prognostic variables	TTP	TTNTs	Discontinuation	Tx patterns (subsequent Tx)	Health costs	HRQoL
UK RWE 2022	IMDC	X	X	X	X	X	X	X[a]	
Hawkins *et al.* (2020)[73]	MSKCC	X				X	X		
Wagstaff *et al.* (2016) (RECCORD)[6]		X		X	X	X	X		
NICE TA780:[23] SACT data report	IMDC	X			X	X	X		
IQVIA 2022							X		
NCRAS 2023[58]		X[b]							
Kidney Cancer UK (audit of kidney cancer services in England)[75]		X[c]					X		
Brown *et al.* (2021)[71]		X					X		
Hack *et al.* (2019)[72]		X	X	X[d]					
Hilser *et al.* (2023)[77]	Heng	X	X						
Nathan *et al.* (2022)[70]	IMDC	X	X			X			
Nathan *et al.* (2023)[78] (CARINA: NCT04957160)	IMDC					X	X		

Notes:
a Data on RDI reported, included as dosing used to inform drug costs.
b OS data yearly records (2013–9) for Stage 1–4 ccRCC and RCC NOS patients with confirmed or unconfirmed diagnoses.
c Reported postoperative 30-day all-cause survival in M0 kidney cancer patients who undergo RN or NSS and postoperative 12-month all-cause survival in M0 kidney cancer patients who undergo RN or NSS.
d Proportion with disease progression only.

Critical appraisal of real-world evidence

The DataSAT was completed for UK RWE (2022),[28] Hawkins *et al.* (2020),[73] RECCORD (Wagstaff *et al.*, 2016[6]) and SACT TA780.[23] Note that the research team had access to the full data set only for UK RWE (2022),[28] 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 *et al.* (2021),[71] Hack *et al.* (2019),[72] Hilser *et al.* (2023),[77] Nathan *et al.* (2022)[70] and CARINA (Nathan *et al.*, 2023)[78] were only available as conference abstracts.
- Kidney Cancer UK Audit report[75] and the NCRAS data;[58] limited access to the data set based on information within reports available online.

The DataSAT assessments for the four appraised data sets[6,23,28,73] are summarised below.

Data provenance: Data provenance refers to the documented history and origin of data, including its creation, transformation and movement throughout its life cycle. Data for three[6,28,73] 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.

By contrast, the SACT database served as a data source for one[23] of the analyses. This national database in England collects real-time information reported by NHS Trusts through electronic prescribing systems during patient care. The data set 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)[28] data set included patients from 15 UK hospitals who started first-line systemic therapy between January 2018 and August 2022. The Hawkins *et al.* (2020)[73] analysis included patients who initiated first-line systemic therapy in two specific hospitals in Cambridge and Manchester between January 2008 and December 2015. The RECCORD data set (Wagstaff *et al.*, 2016)[6] included patients who began first-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 SACT treatment. For the included analysis,[23] 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) data set,[28] but no additional documents were available beyond those in the public domain for three of the four analyses,[6,23,73] limiting further insights into the data provenance.

Data quality: Across the UK RWE (2022),[28] Hawkins *et al.* (2020),[73] RECCORD (Wagstaff *et al.*, 2016[6]) and SACT TA780[23] data sets, 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[23] data set was slightly different to the other three data sets in that it selected participants based on Blueteq® (Blueteq Ltd, Havant, UK) 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.[23] In all data sets,[6,23,28,73] the majority of items linked to defining the population, for example 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.[28,73]

Similarly for treatments received (first-line and subsequent treatments), these data were considered to be 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 data sets[6,23,28,73] 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,[23] 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 the treatment start date. For outcomes which may have included some element of clinician judgement, for example the assessment of progression, the EAG notes that there may have been some variability between centres and across studies. In most cases, the assessment was based on analysis of multiple markers, such as radiology, symptomatology, clinical investigation and therapy changes, although primarily radiological assessment was used to determine the progression. Medical records were assumed to be accurate sources for determining the 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 data sets[6,23,28,73] exhibited a reasonable data quality, with a focus on accuracy, completeness, and they 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[6,23,28,73] each included a significant number of patients, with sample sizes ranging from 514[6] to 1319.[28] All four data sets[6,23,28,73] included data from treatment and monitoring in a UK secondary care setting. In three of the four analyses,[23,28,73] the majority of patients had clear-cell histology, while one data set[6] included only patients with clear-cell histology. The majority of patients in the data sets were categorised as intermediate or poor risk[6,28,73] according to the IMDC criteria, with one data set[23] 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 data sets reflected the appropriate population.

The UK RWE (2022)[28] data set provided valuable insights into the population of RCC patients starting first-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 [*first line*: cabozantinib TA542;[27] tivozanib (TA512);[26] nivolumab + ipilimumab (TA780 via CDF for the majority of the data collection period 2019–22 TA581/TA780);[23] and avelumab + axitinib TA645 (via CDF);[57] *second-line*: nivolumab TA417;[33] cabozantinib TA463;[30] and lenvatinib + everolimus TA498[31] (refer to *Table 50*)]. The Hawkins *et al.* (2020)[73] data set 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 *et al.*, 2016[6]) 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 data sets pre-date the recommendations for many current treatment options, comparing them with the more recent UK RWE (2022)[28] data set can provide insights into the impact of newer treatments on outcomes and the treatment pathway. The SACT TA780[23] data set specifically focused on patients who received nivolumab + ipilimumab treatment during the managed access period following the NICE appraisal. The data set 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 2019–22 and 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.[6,23,28,73] In the SACT TA780 data set,[23] 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 to be sufficient to capture the specified outcomes beyond the trial period and to gather valuable insights into subsequent treatments.

Sample sizes ranged from 514[6] to 1319[28] participants. The SACT TA780 data set[23] provided a flow diagram for participants identified to participants included with reasons for not including participants. None of the analyses[6,23,28,73] conducted a sample size calculation as their primary objective was to collect descriptive information rather than test a specific research hypothesis.

Overall, the included data sets[6,23,28,73] provide relevant information from UK practice in terms of treatment patterns and efficacy outcomes (e.g. OS, PFS, TTNT, discontinuation and 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.

TABLE 50 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[24]	TA215[25]	TA219 → TA432[32]	TA333[29]	TA417[33]	TA463[30]	TA542[27]	TA498[31]	TA512[26]	TA780 (CDF review of TA581))[23]	TA645 (CDF)[57]	TA858[15]
LOT 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 *et al.*, 2016)[6]	Mar 2009–Oct 2012	Y	Y	Y									
Hawkins *et al.* (2020)[73]	1 Jan 2008–31 Dec 2015	Y	Y	Y	Y								
UK RWE 2022	1 Jan 2018–23 Aug 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y (via CDF)	
Brown *et al.* (2021)[71]	1 Jan 2011–31 Jan 2020	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	*N* (publ Sep 2020)	
SACT TA780[23]	4 Apr 2019–30 Nov 2020	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y (via CDF)	
CARINA (Nathan *et al.*, 2023)[78]	15 Jan 2015–Sept 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y (via CDF)	Y (via CDF)	

ECOG PS, Eastern Cooperative Oncology Group performance status; publ, published.

Additional outcomes from real-world evidence

Overall survival

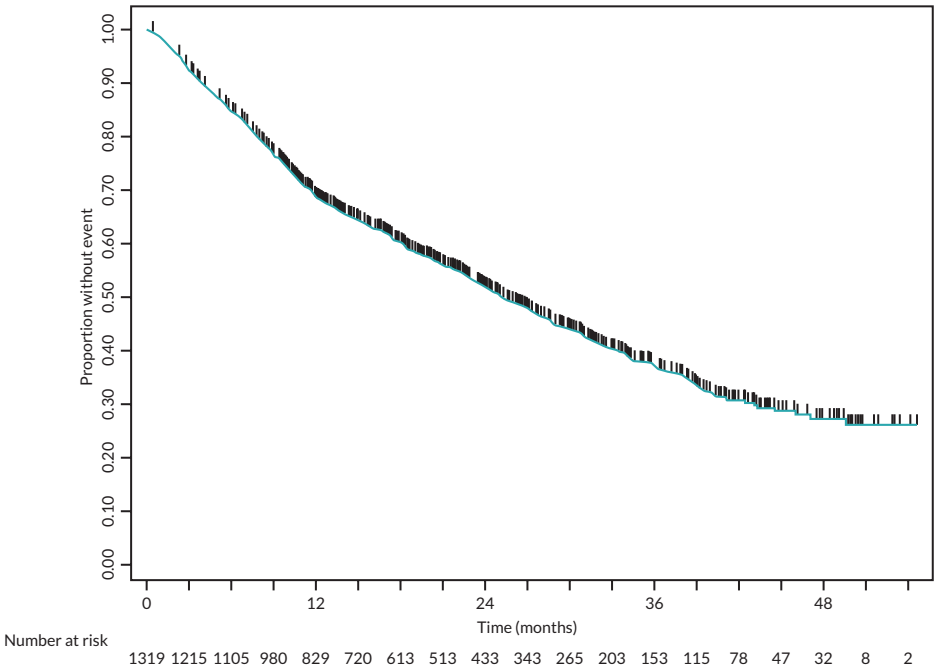


FIGURE 21 United Kingdom RWE: pooled OS at first line.

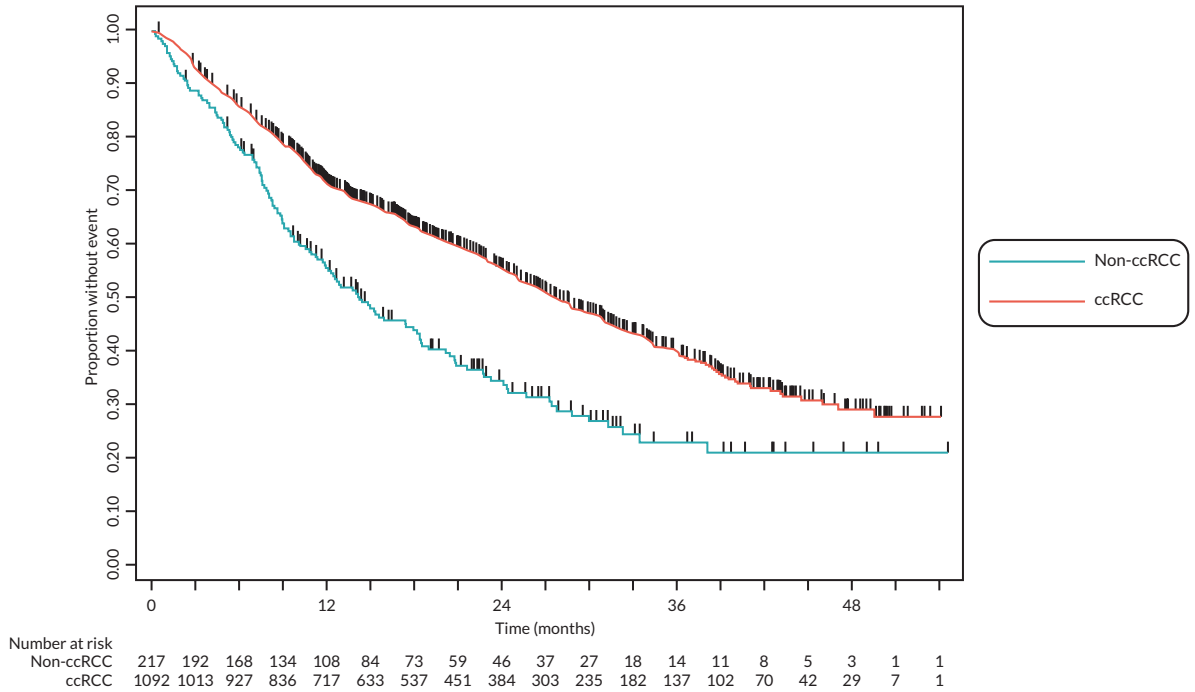


FIGURE 22 United Kingdom RWE: histology stratified OS at first line.

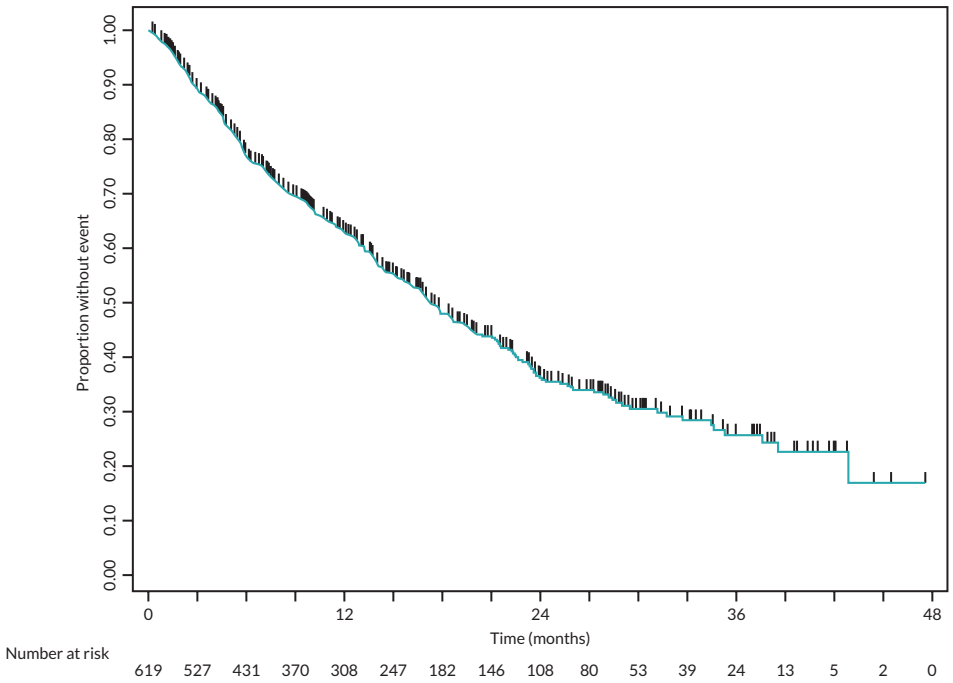


FIGURE 23 United Kingdom RWE: pooled OS at second line.

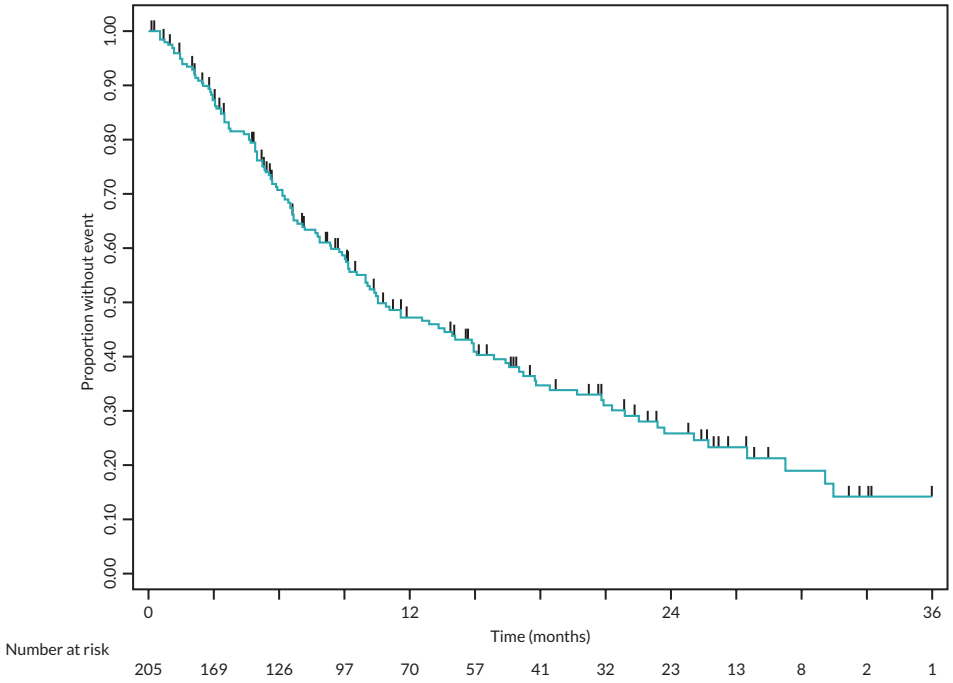


FIGURE 24 United Kingdom RWE: pooled OS at third line.

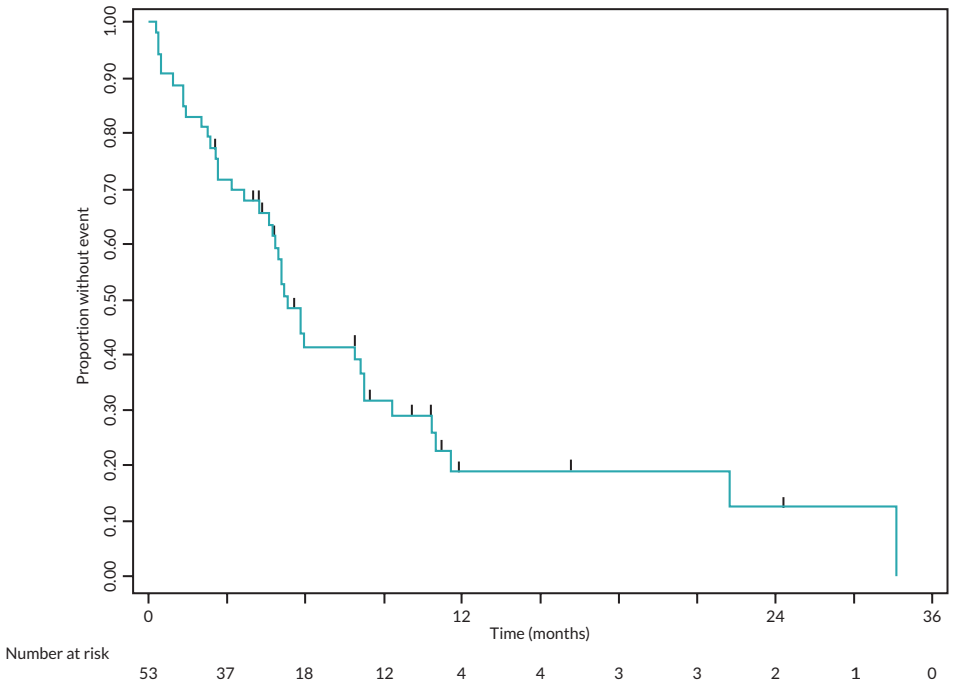
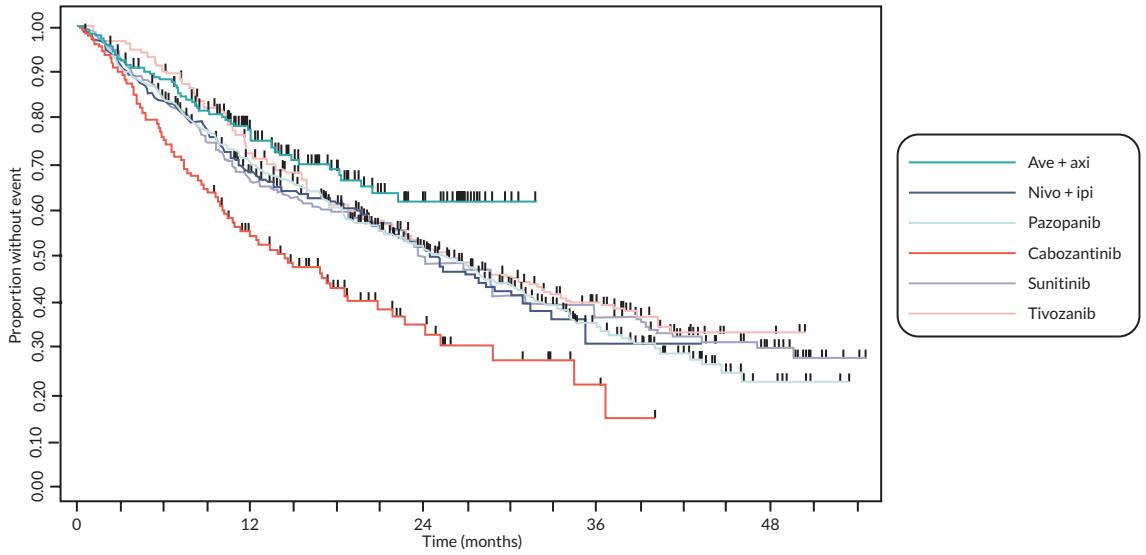


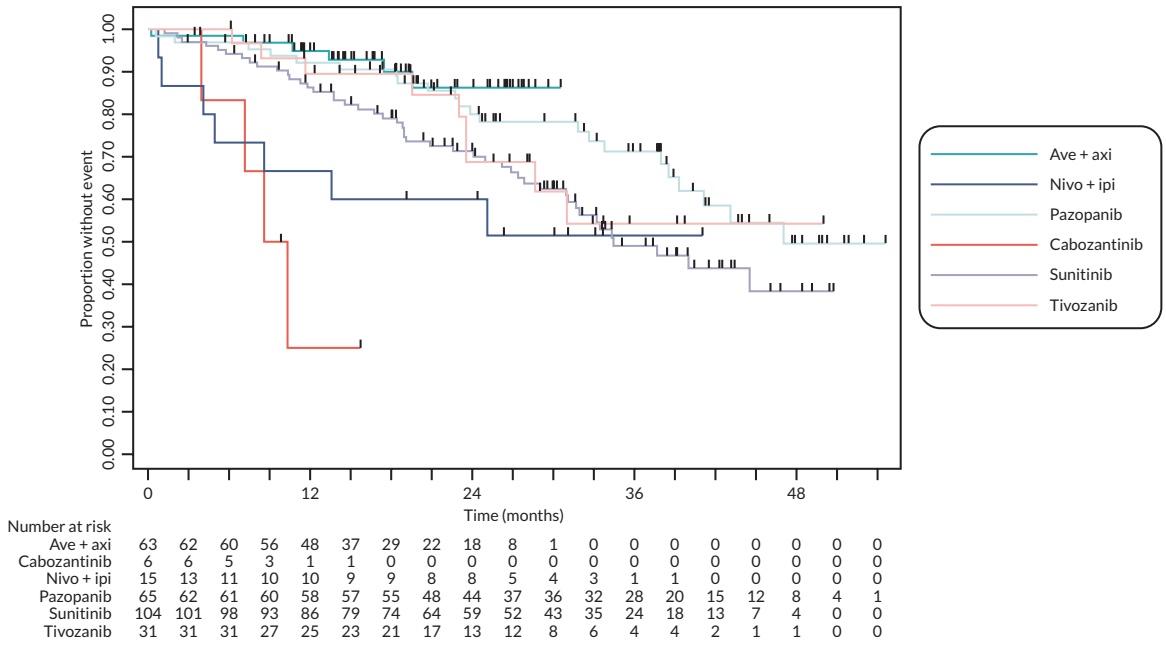
FIGURE 25 United Kingdom RWE: pooled OS at fourth line.



Number at risk

Ave + axi	168	155	144	127	98	75	59	40	29	13	2	0	0	0	0	0	0	0	0
Cabozantinib	114	103	85	73	56	46	34	24	17	10	9	6	4	2	0	0	0	0	0
Nivo + ipi	309	284	255	226	187	160	135	106	85	59	35	16	4	2	1	0	0	0	0
Pazopanib	234	216	195	172	152	142	130	117	107	92	83	72	64	49	35	24	17	6	2
Sunitinib	326	296	274	249	224	207	183	163	141	123	98	78	56	41	28	14	8	2	0
Tivozanib	104	101	94	83	70	60	50	42	34	29	21	18	14	12	8	4	4	0	0

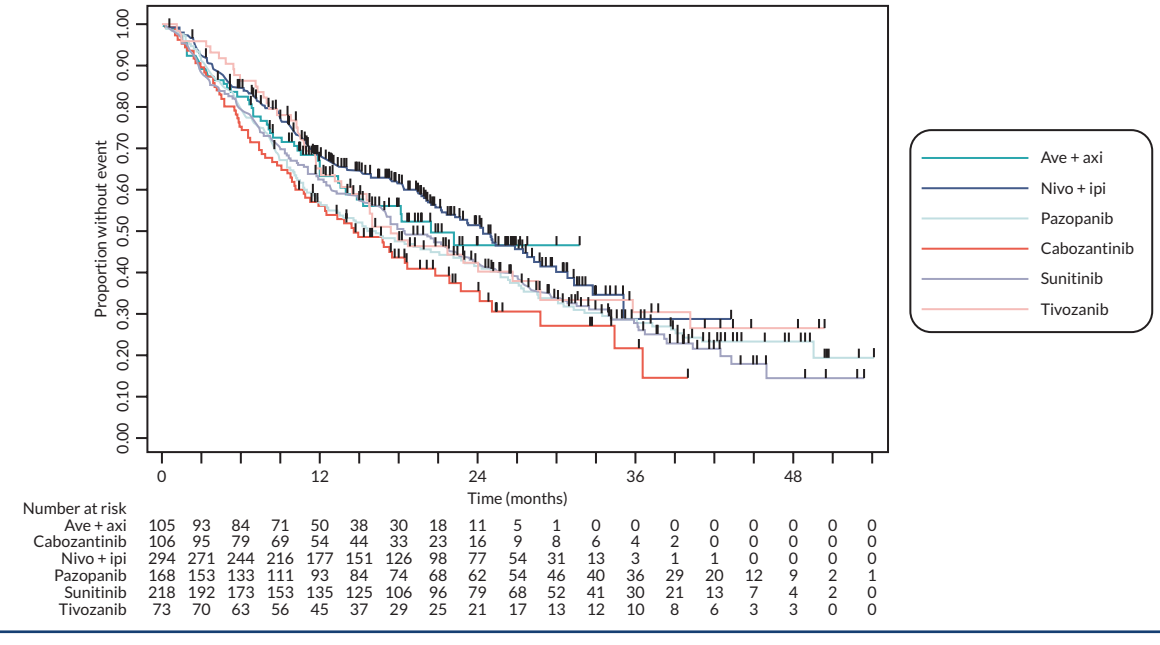
FIGURE 26 United Kingdom RWE: treatment-stratified OS at first line.



Number at risk

Ave + axi	63	62	60	56	48	37	29	22	18	8	1	0	0	0	0	0	0	0	0
Cabozantinib	6	6	5	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Nivo + ipi	15	13	11	10	10	9	9	8	8	5	4	3	1	1	0	0	0	0	0
Pazopanib	65	62	61	60	58	57	55	48	44	37	36	32	28	20	15	12	8	4	1
Sunitinib	104	101	98	93	86	79	74	64	59	52	43	35	24	18	13	7	4	0	0
Tivozanib	31	31	31	27	25	23	21	17	13	12	8	6	4	4	2	1	1	0	0

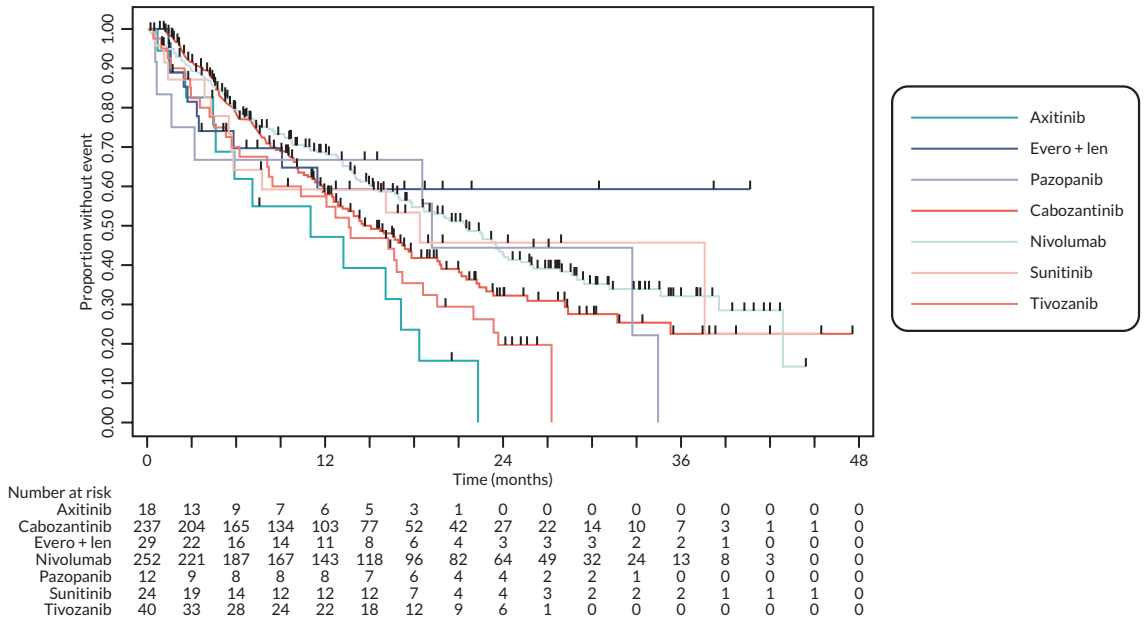
FIGURE 27 United Kingdom RWE: treatment-stratified OS at first line by IMDC favourable risk.



Number at risk

Ave + axi	105	93	84	71	50	38	30	18	11	5	1	0	0	0	0	0	0	0	0
Cabozantinib	106	95	79	69	54	44	33	23	16	9	8	6	4	2	0	0	0	0	0
Nivo + ipi	294	271	244	216	177	151	126	98	77	54	31	13	3	1	1	0	0	0	0
Pazopanib	168	153	133	111	93	84	74	68	62	54	46	40	36	29	20	12	9	2	1
Sunitinib	218	192	173	153	135	125	106	96	79	68	52	41	30	21	13	7	4	2	0
Tivozanib	73	70	63	56	45	37	29	25	21	17	13	12	10	8	6	3	3	0	0

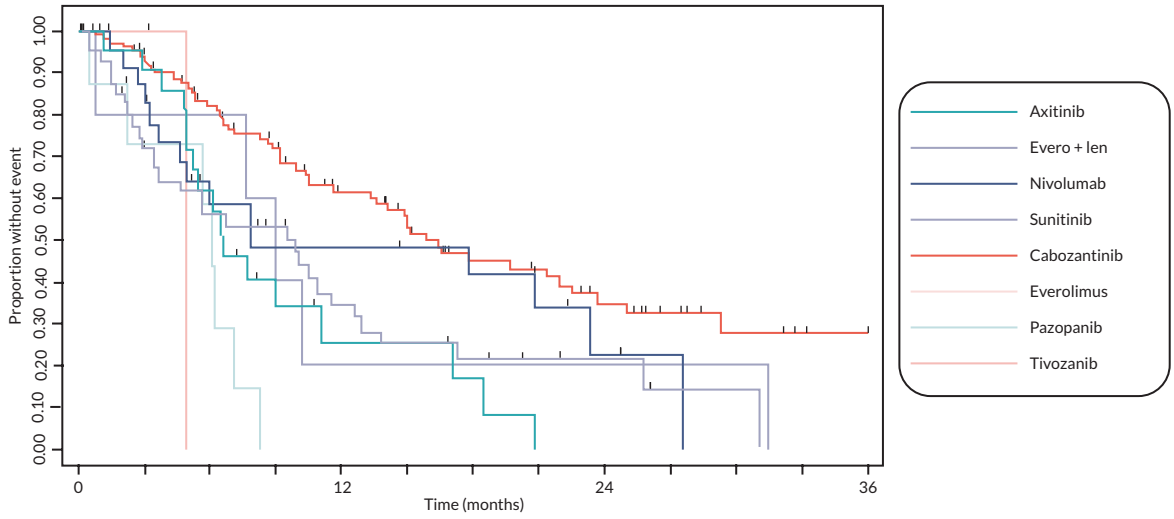
FIGURE 28 United Kingdom RWE: treatment-stratified OS at first line by IMDC intermediate/poor risk (Challapalli *et al.*).



Number at risk

Axitinib	18	13	9	7	6	5	3	1	0	0	0	0	0	0	0	0	0
Cabozantinib	237	204	165	134	103	77	52	42	27	22	14	10	7	3	1	1	0
Evero + len	29	22	16	14	11	8	6	4	3	3	3	2	2	1	0	0	0
Nivolumab	252	221	187	167	143	118	96	82	64	49	32	24	13	8	3	0	0
Pazopanib	12	9	8	8	8	7	6	4	4	2	2	1	0	0	0	0	0
Sunitinib	24	19	14	12	12	12	7	4	4	3	2	2	2	1	1	1	0
Tivozanib	40	33	28	24	22	18	12	9	6	1	0	0	0	0	0	0	0

FIGURE 29 United Kingdom RWE: treatment-stratified OS at second line (Challapalli *et al.*).



Number at risk

Ave + axi	23	19	12	6	3	3	2	0	0	0	0	0	0
Cabozantinib	102	91	73	60	45	36	24	22	16	9	5	2	1
Evero + len	24	20	11	9	9	8	7	4	2	1	0	0	0
Everolimus	8	5	4	0	0	0	0	0	0	0	0	0	0
Nivolumab	40	27	21	18	11	8	6	4	3	1	1	0	0
Pazopanib	1	1	0	0	0	0	0	0	0	0	0	0	0
Sunitinib	5	4	4	3	1	1	1	1	1	1	1	0	0
Tivozanib	1	1	0	0	0	0	0	0	0	0	0	0	0

FIGURE 30 United Kingdom UK RWE: treatment-stratified OS at third line (Challapalli *et al.*).

Progression-free survival

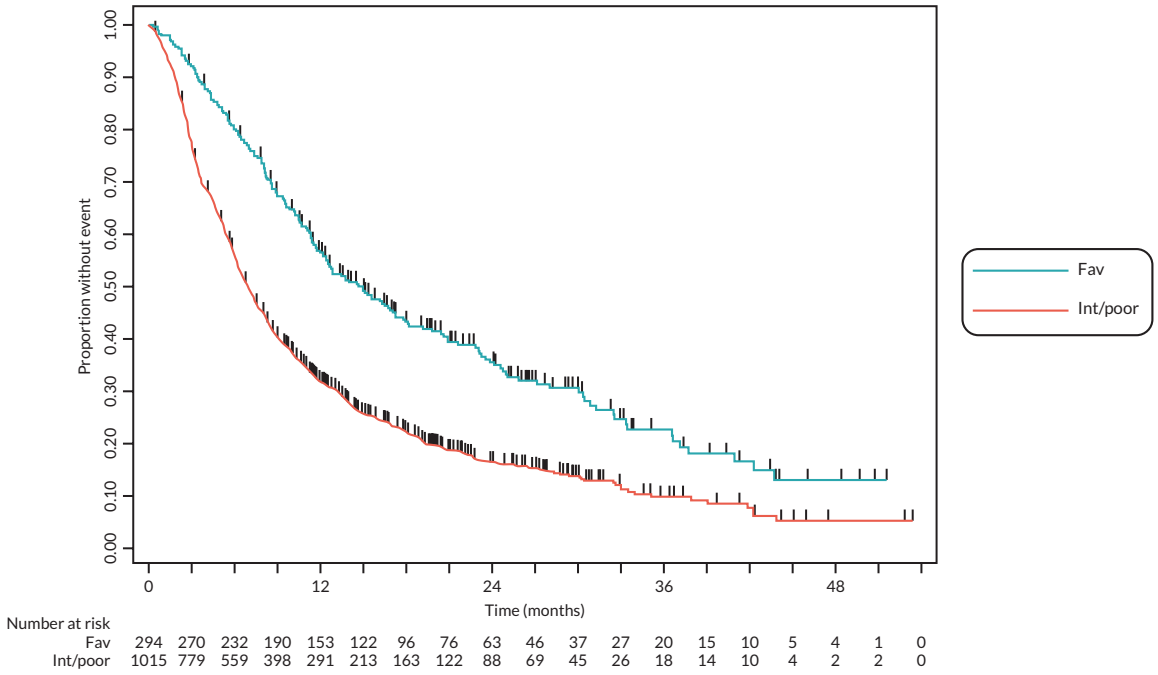


FIGURE 31 United Kingdom RWE: risk-stratified PFS at first line.

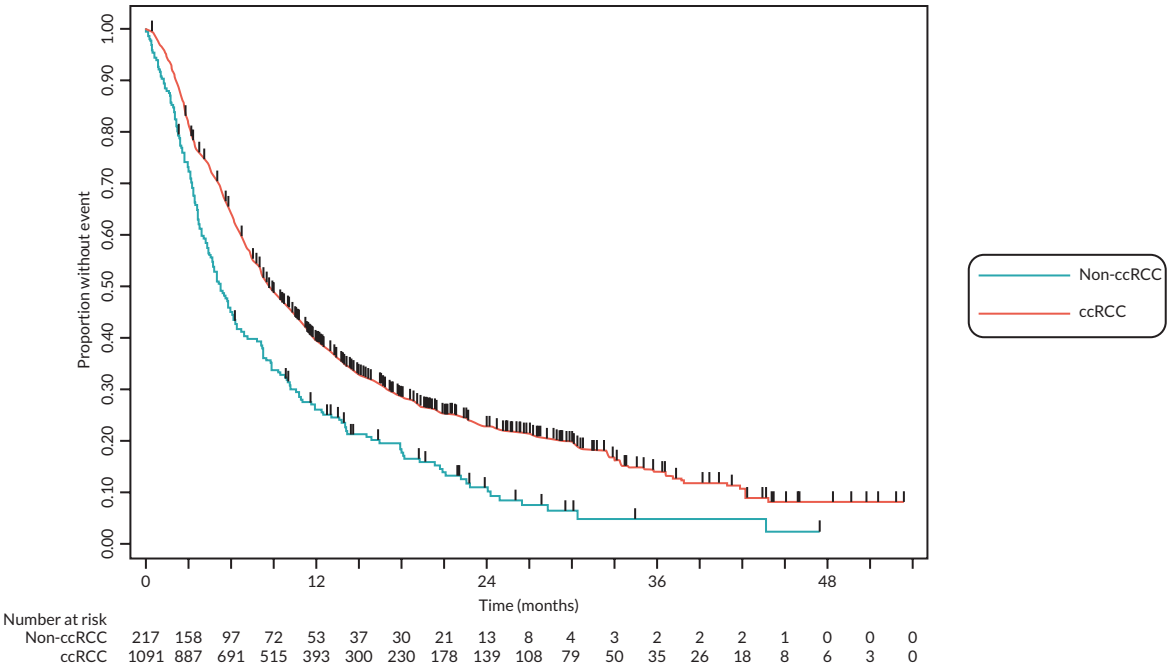
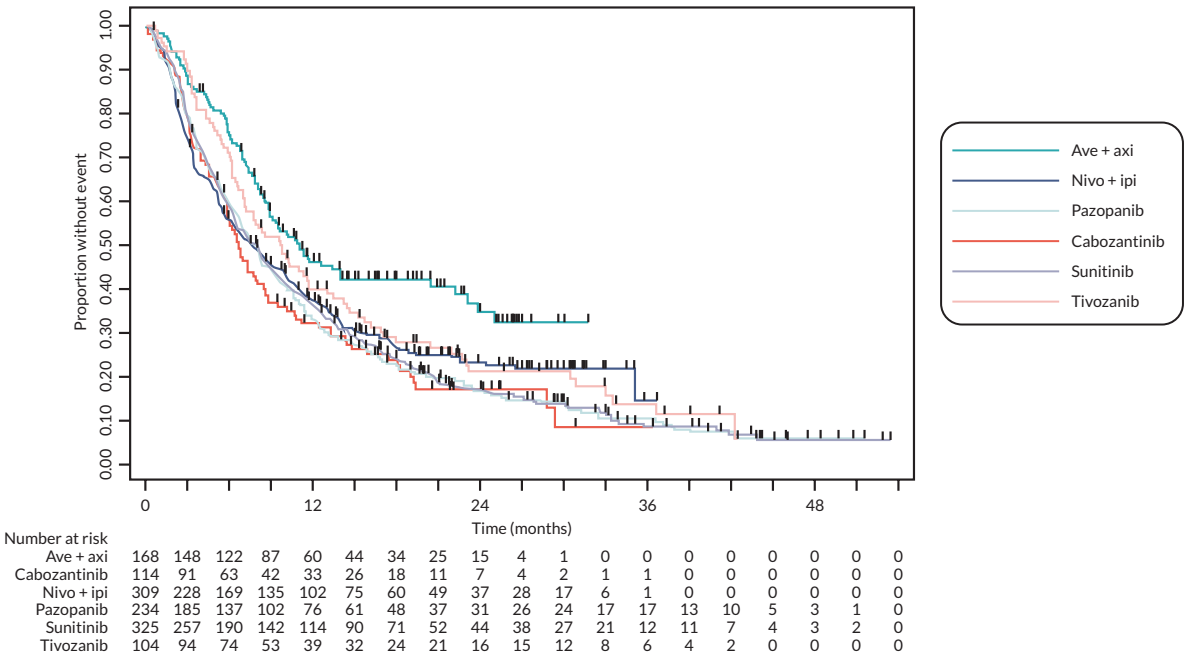


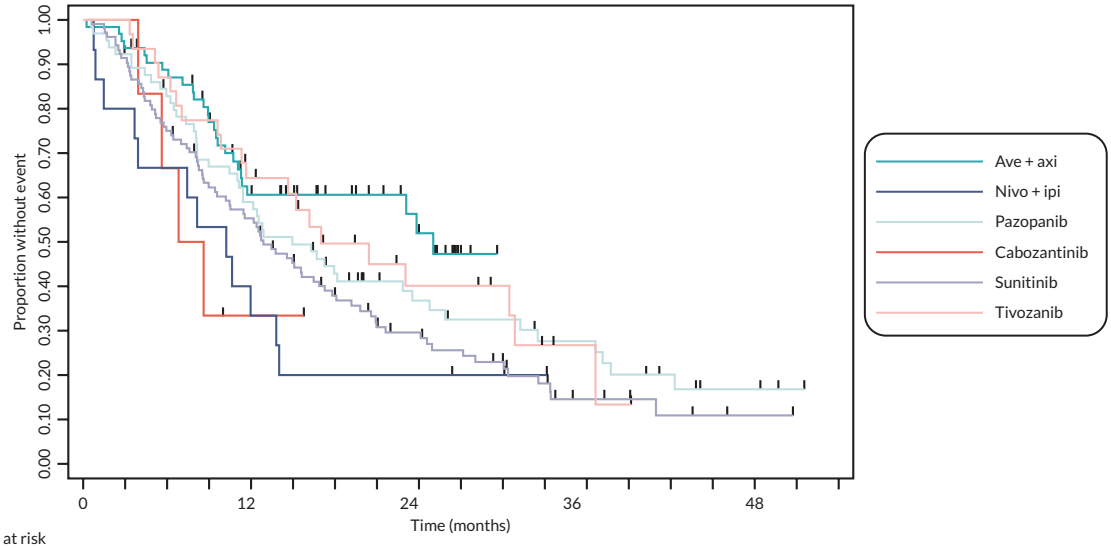
FIGURE 32 United Kingdom RWE: histology-stratified PFS at first line.



Number at risk

Ave + axi	168	148	122	87	60	44	34	25	15	4	1	0	0	0	0	0	0	0	0
Cabozantinib	114	91	63	42	33	26	18	11	7	4	2	1	1	0	0	0	0	0	0
Nivo + ipi	309	228	169	135	102	75	60	49	37	28	17	6	1	0	0	0	0	0	0
Pazopanib	234	185	137	102	76	61	48	37	31	26	24	17	17	13	10	5	3	1	0
Sunitinib	325	257	190	142	114	90	71	52	44	38	27	21	12	11	7	4	3	2	0
Tivozanib	104	94	74	53	39	32	24	21	16	15	12	8	6	4	2	0	0	0	0

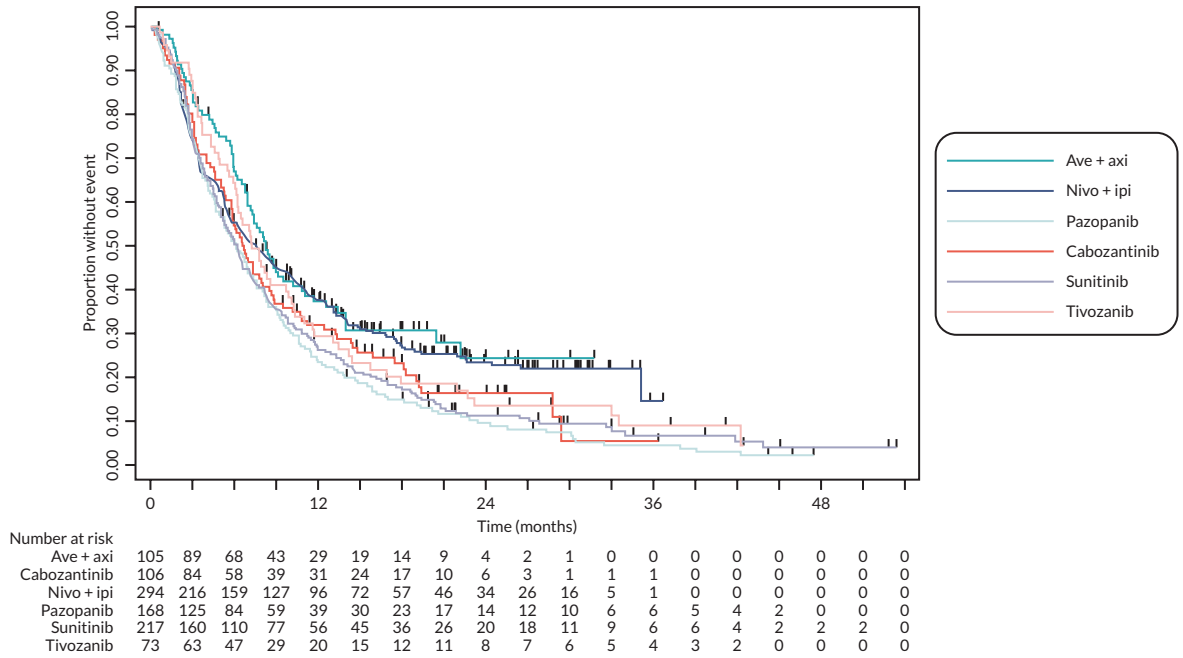
FIGURE 33 United Kingdom RWE: treatment-stratified PFS at first line.



Number at risk

Ave + axi	63	59	54	44	31	25	20	16	11	2	0	0	0	0	0	0	0	0	0
Cabozantinib	6	6	4	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Nivo + ipi	15	12	10	8	5	3	3	3	3	2	1	1	0	0	0	0	0	0	0
Pazopanib	65	59	52	42	37	31	25	20	17	14	14	11	11	8	6	3	3	1	0
Sunitinib	104	95	78	63	56	43	34	25	23	19	15	11	6	5	3	2	1	0	0
Tivozanib	31	31	27	24	19	17	12	10	8	8	6	3	2	1	0	0	0	0	0

FIGURE 34 United Kingdom RWE: treatment-stratified OS at first line by IMDC favourable risk.

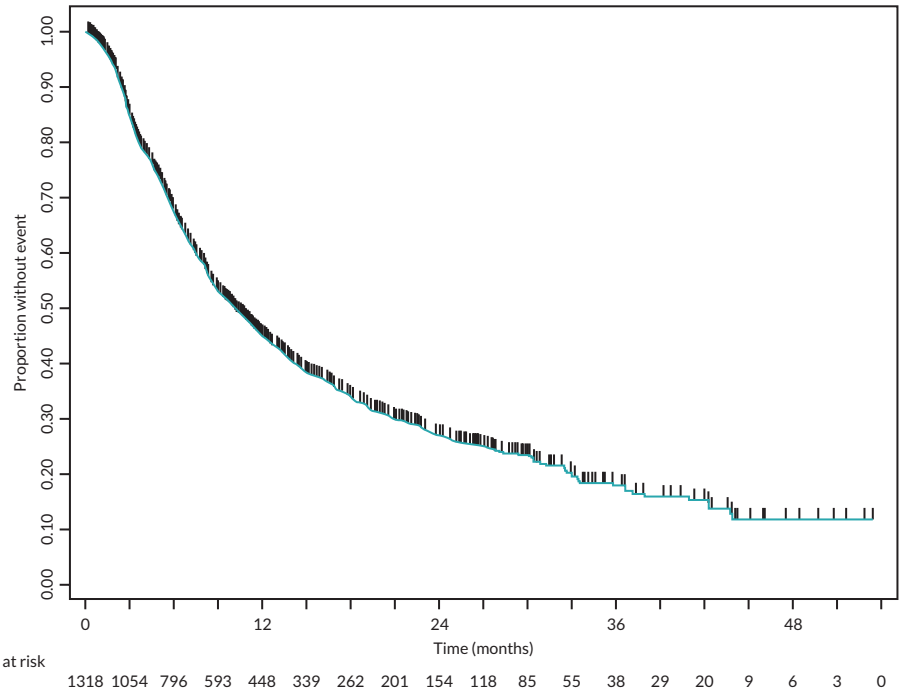


Number at risk

Ave + axi	105	89	68	43	29	19	14	9	4	2	1	0	0	0	0	0	0	0	0
Cabozantinib	106	84	58	39	31	24	17	10	6	3	1	1	1	0	0	0	0	0	0
Nivo + ipi	294	216	159	127	96	72	57	46	34	26	16	5	1	0	0	0	0	0	0
Pazopanib	168	125	84	59	39	30	23	17	14	12	10	6	6	5	4	2	0	0	0
Sunitinib	217	160	110	77	56	45	36	26	20	18	11	9	6	6	4	2	2	2	0
Tivozanib	73	63	47	29	20	15	12	11	8	7	6	5	4	3	2	0	0	0	0

FIGURE 35 United Kingdom RWE: treatment-stratified OS at first line by IMDC intermediate/poor risk.

Time to progression



Number at risk

1318	1054	796	593	448	339	262	201	154	118	85	55	38	29	20	9	6	3	0

FIGURE 36 United Kingdom RWE: pooled TTP at first line.

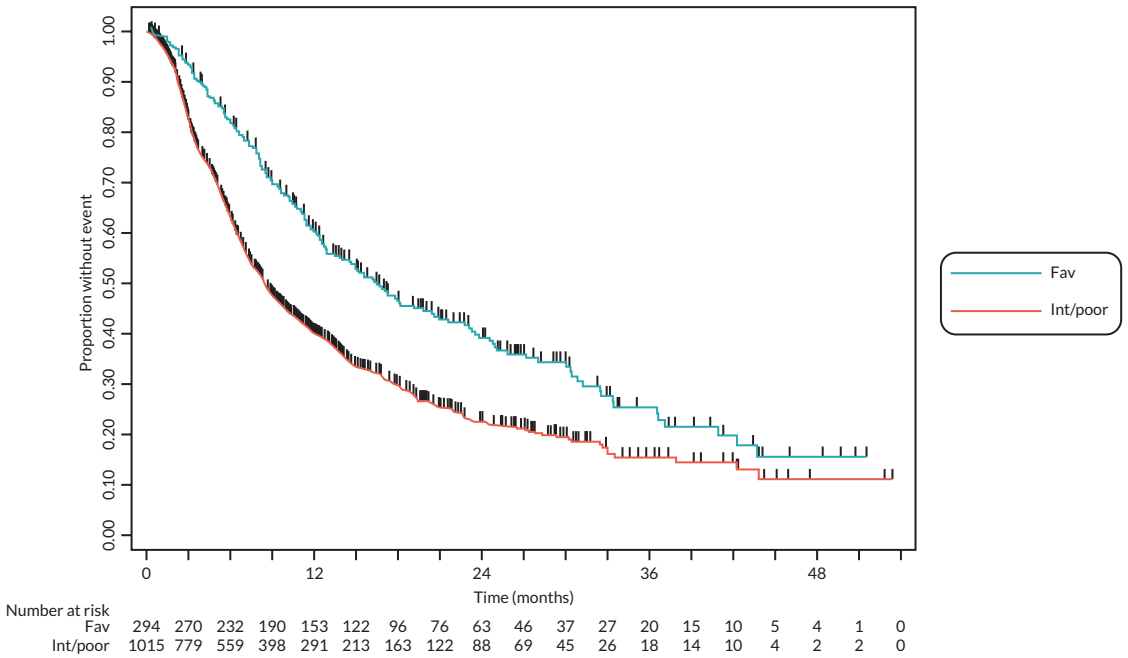


FIGURE 37 United Kingdom RWE: risk-stratified TTP at first line.

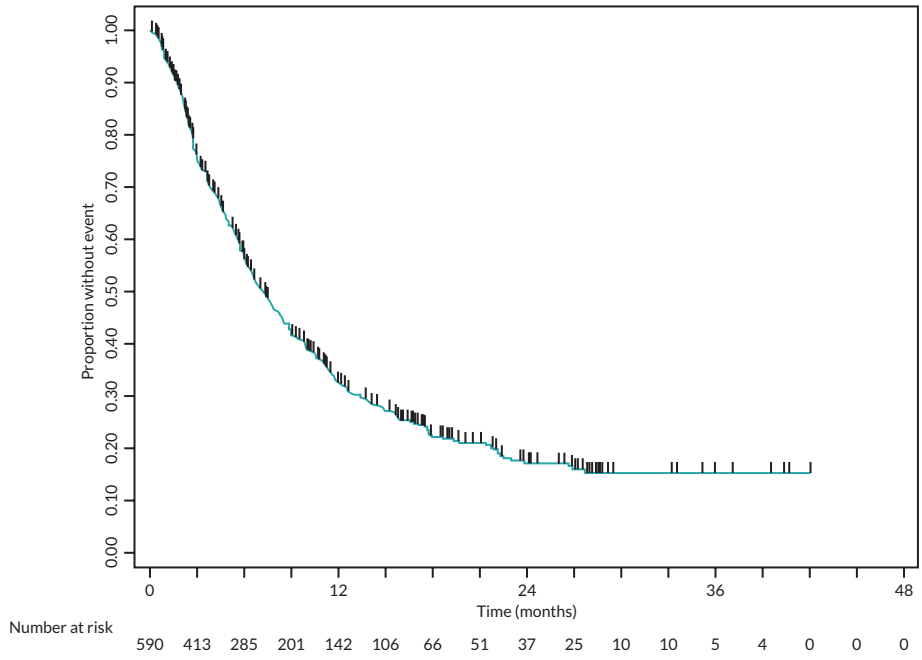


FIGURE 38 United Kingdom RWE: pooled TTP at second line.

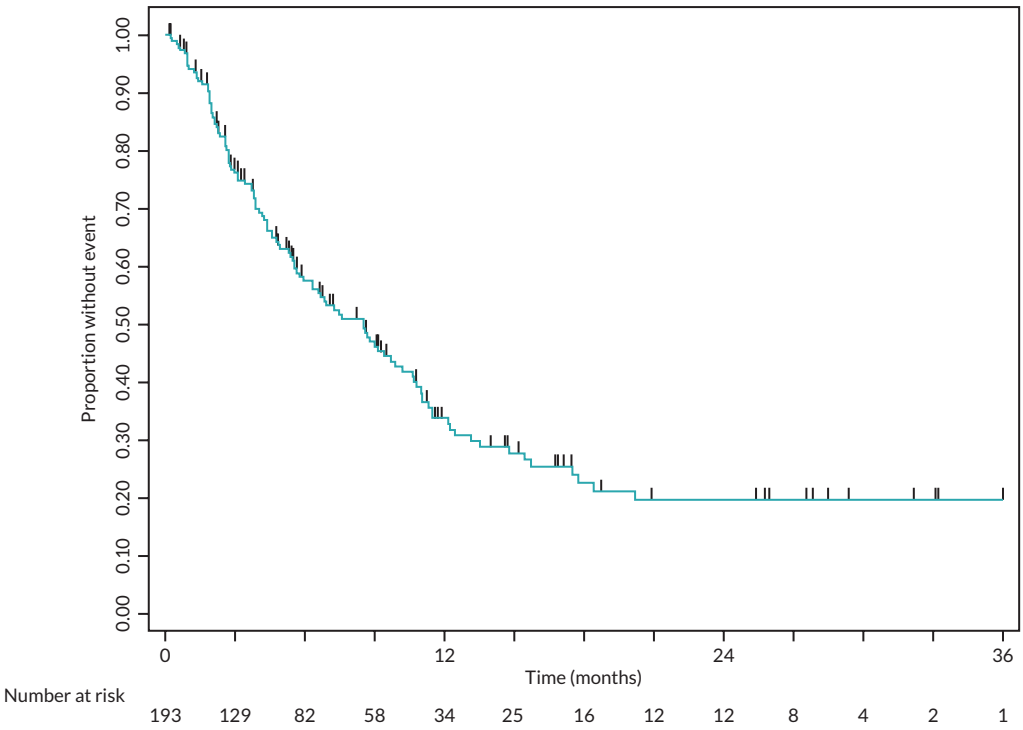


FIGURE 39 United Kingdom RWE: pooled TTP at third line.

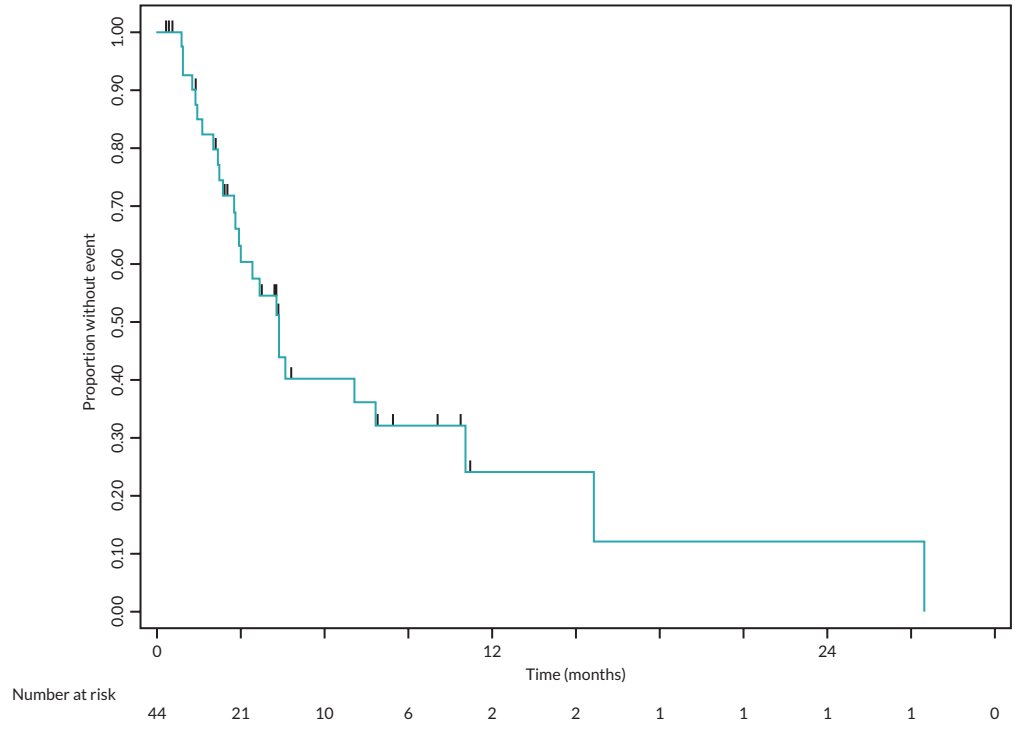


FIGURE 40 United Kingdom RWE: pooled TTP at fourth line.

Post-progression survival

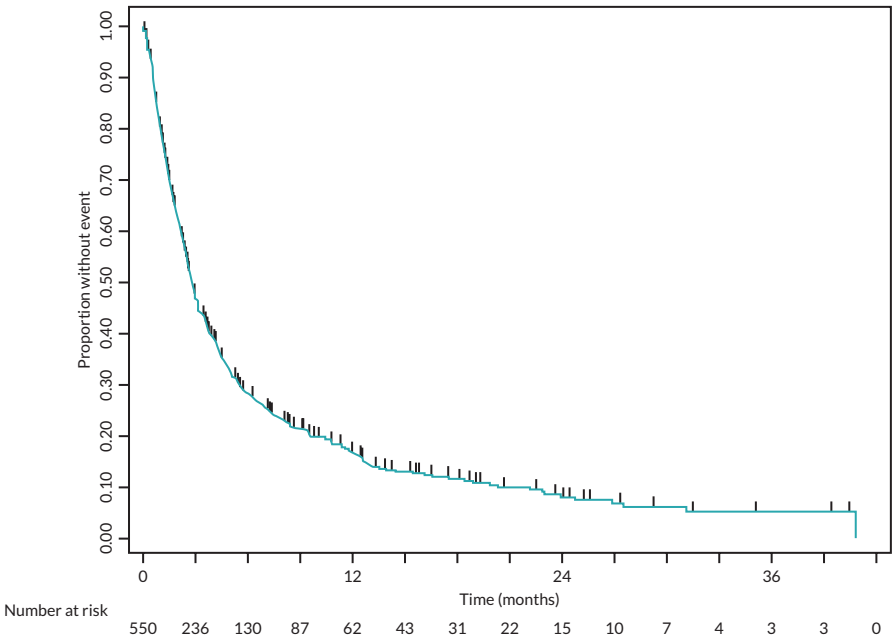


FIGURE 41 United Kingdom RWE: pooled PPS.

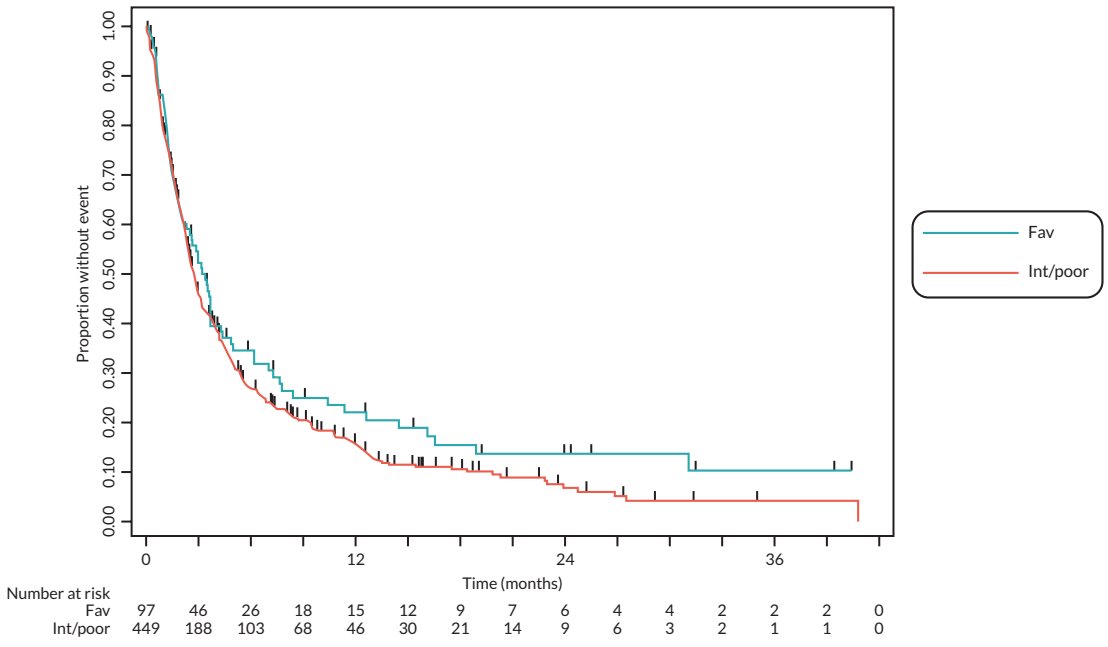
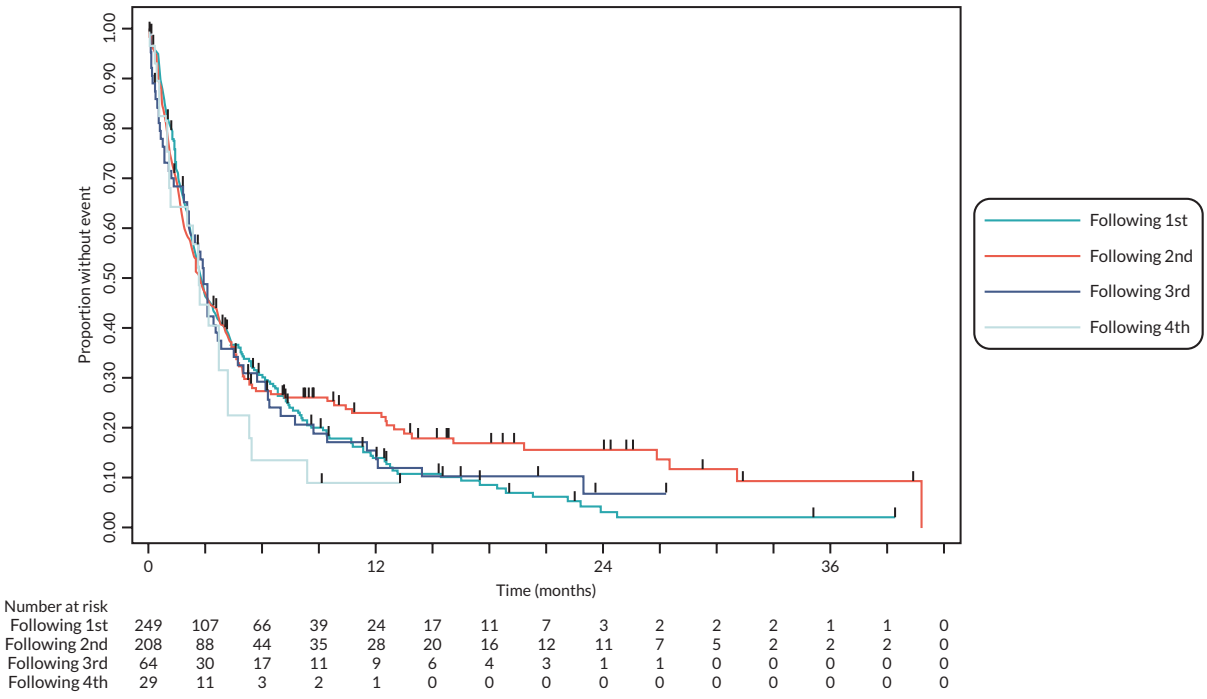


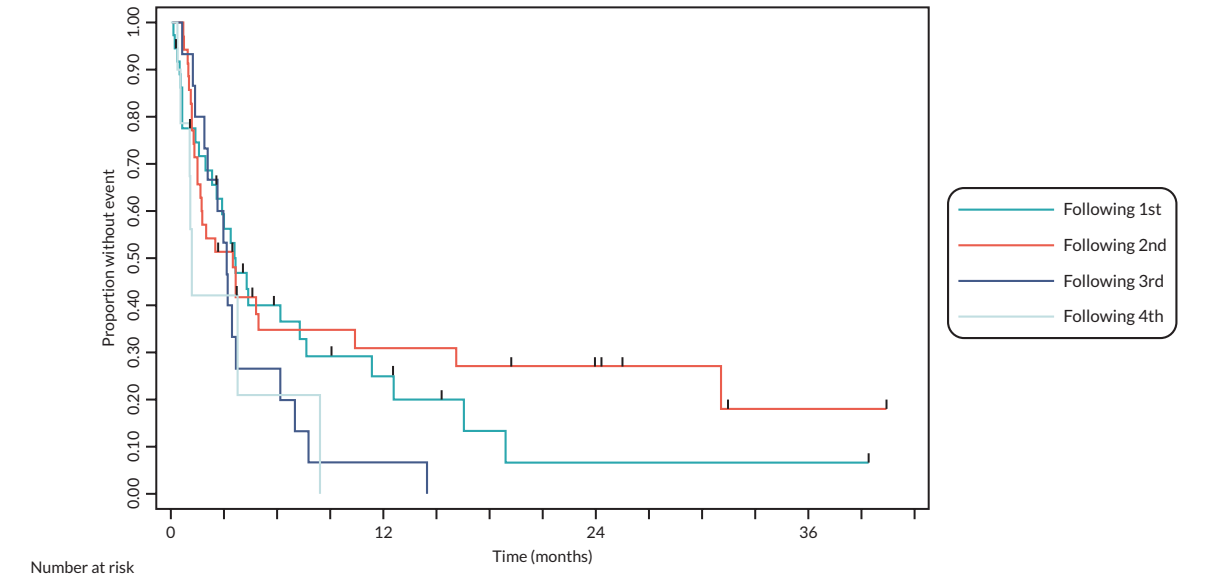
FIGURE 42 United Kingdom RWE: risk-stratified PPS at first line.



Number at risk

Following 1st	249	107	66	39	24	17	11	7	3	2	2	2	1	1	0
Following 2nd	208	88	44	35	28	20	16	12	11	7	5	2	2	2	0
Following 3rd	64	30	17	11	9	6	4	3	1	1	0	0	0	0	0
Following 4th	29	11	3	2	1	0	0	0	0	0	0	0	0	0	0

FIGURE 43 United Kingdom RWE: line-stratified PPS.



Number at risk

Following 1st	37	18	11	8	6	4	2	1	1	1	1	1	1	1	0
Following 2nd	35	17	10	9	8	8	7	6	5	3	3	1	1	1	0
Following 3rd	15	8	4	1	1	0	0	0	0	0	0	0	0	0	0
Following 4th	10	3	1	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 44 United Kingdom RWE: line-stratified PPS by favourable-risk group.

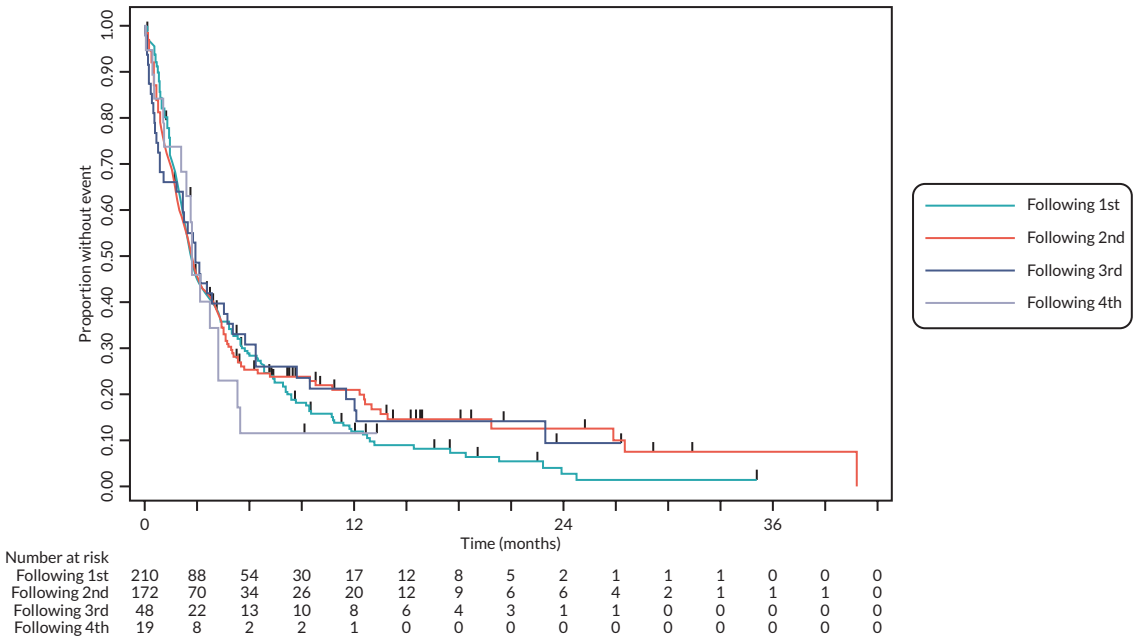


FIGURE 45 United Kingdom RWE: line-stratified PPS by intermediate-/poor-risk group.

Discontinuation

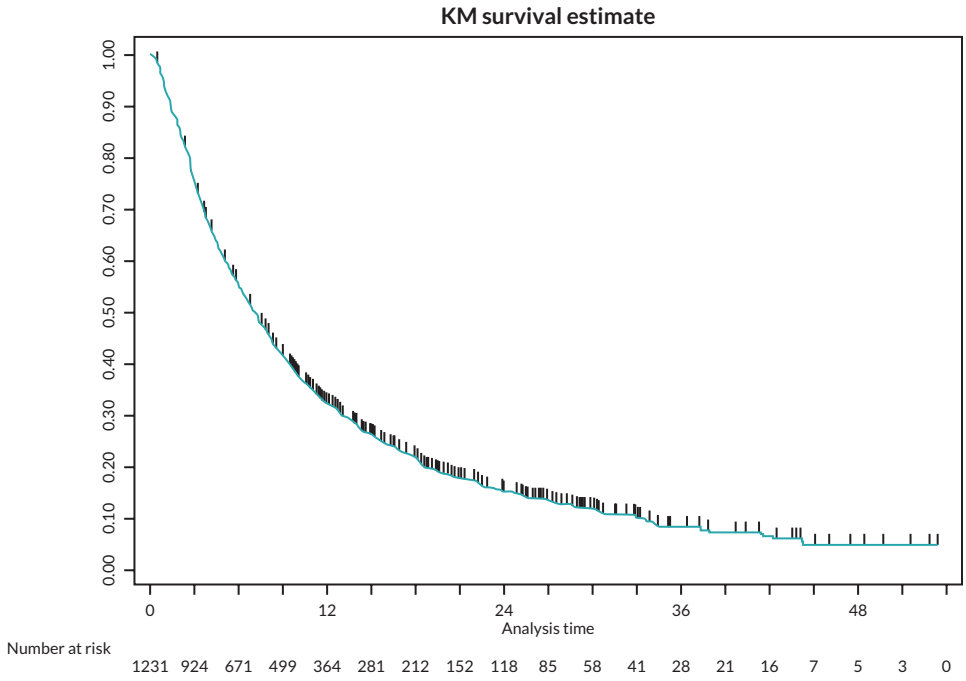


FIGURE 46 United Kingdom RWE: pooled time to treatment discontinuation at first line (Challapalli *et al.*).

TABLE 51 Treatments used from first line to fourth line across three RWE studies

	RECCORD (Wagstaff *et al.*, 2016)	Hawkins *et al.* (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
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

TABLE 51 Treatments used from first line to fourth line across three RWE studies (*continued*)

	RECCORD (Wagstaff *et al.*, 2016)	Hawkins *et al.* (2020)	UK RWE 2022
	%	*n*	%
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

belz, belzutifan.
a Other grouping included treatments not recommended by NICE or not in the treatment pathway set out in *Figure 1*: 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);[6] Hawkins *et al.* (2020);[73] UK RWE 2022.[28]

TABLE 52 Sequences described following defined first-line therapy

	SACTTA780[23]	Nathan *et al.* (2023)[78] (CARINA: NCT04957160)		Brown *et al.* (2021)[a,71]	
N	814	129		440	1045
1L treatment					
Suni	–	–	–	*N* = 186	*N* = 422
Pazo	–	–	–	*N* = 178	*N* = 500
Nivo + ipi	814 (100%)[b]	107 (82.9%)[c]	–		
Ave + axi	–	–	22 (17.1%)[c]		
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

continued

TABLE 52 Sequences described following defined first-line therapy (*continued*)

	SACTTA780[23]	Nathan *et al.* (2023)[78] (CARINA: NCT04957160)		Brown *et al.* (2021)[a,71]	
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

carbo, carboplatin; dabref, dabrafenib; IRIN, irinotecan; MDG, modified de gramont; panit, panitumumab; pem pembrolizumab; tram, trametinib.
a Total for cabo cohort *n* = 440 and total for axitinib cohort *n* = 1045. The denominator for the reported sequences was unclear from the information available in the conference abstract, and data are reported as seen.
b Study cohort was participants who had received nivolumab + ipilimumab in first line in the CDF.
c Study cohort was participants who had received a first-line combination therapy, including a checkpoint inhibitor.

TABLE 53 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[28]	1L	Ave + axi; cabo; nivo + ipi; pazo; suni; tivo	Time from start of 1L treatment to death	1319	16.8 months (15.8 to 17.6)	25.16 (23.39 to 27.47)[a]	12 months: 68.9% (66.3% to 71.3%) 24 months: 52.3% (49.3% to 55.1%) 36 months: 37.3% (33.9% to 40.7%) 48 months: 27.3% (23.2% to 31.6%)
	2L	Axi; cabo; lenv + evero; nivo; pazo; suni; tivo	Time from start of 2L treatment to death	632		17.25 (15.61 to 19.58)	12 months: 63.1% (58.9% to 66.9%) 24 months: 36.3% (31.6% to 40.9%) 36 months: 25.6% (20.3% to 31.2%)
	3L	Axi; cabo; lenv + evero; nivo; suni	Time from start of 3L treatment to death	214		10.55 (9.03 to 14.85)	12 months: 47.3% (39.7% to 54.4%) 24 months: 25.8% 918.6% to 33.6%) 36 months: 14.3% (7.2% to 23.7%)
	4L	Axi; evero	Time from start of 4L treatment to death	54		5.32 (4.63 to 8.25)	12 months: 18.8% (8.1% to 32.9%) 24 months: 12.5% (3.% to 28.6%)
Hawkins *et al.* (2020)[73]	1L	Suni; pazo; evero; other	Time from the start of 1L treatment to death	652	Mean 23.8 (22.2 to 25.4)	12.9 (NR)	12 months; 52.4% (48.6% to 56.4%) 24 months: 30.9% (27.3% to 34.9%) 36 months: 22.6% (19.3% to 26.6%) 60 months: 10.8% (8.0% to 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 months: 31.5% (25.2% to 39.5%) 24 months: 17.0% (11.8% to 24.7%) 36 months: 7.1% (3.1% to 16.5%) 60 months: 7.1% (3.1% to 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
	3L	Axi; evero; other	Time from the start of 3L treatment to death	18	Mean 26.1 (NR)	5.91 (NR)	12 months: 23.8% (10.1% to 55.9%); 24 months: 7.9% (1.3% to 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 *et al.* (2016) (RECCORD)[6]	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

continued

TABLE 53 Overall survival estimates from RWE (*continued*)

Study	LOT	Intervention	OS definition	*N*	Median follow-up (95% CI)	Median OS months (95% CI)	OS rate at:
NICE TA780:[23] 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 months: 80% (77% to 83%) 12 months: 69% (65% to 72%) 18 months: 61% (57% to 64%)
		Nivo + ipi (≥ 6 months 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 months: 88% (84% to 90%) 12 months: 76% (72% to 80%) 18 months: 69% (64% to 73%)
		Nivo + ipi (IMDC poor, score 3 or 4)		281	NR	15 (NR)	6 months: 67% (61% to 72%) 12 months: 55% (49% to 61%) 18 months: 45% (38% to 51%)
Brown *et al.* (2021)[71]	≥ 2L	Cabo	NR	816	NR	11.24 (5.65 to 27.98)[a]	NR
		Axi		1483		10.39 (4.70 to 22.03)[a]	NR
Hack *et al.* (2019)[72]	≥ 1L	Nivo	Time from the start of treatment to death	109	NR	NR	12 months: 56.88% (NR)
Hilser *et al.* (2023)[77]	1L	Cabo + nivo	NR	67	8.3 (NR)	Not reached	NR
Nathan *et al.* (2022)[70]	1L	Ave + axi	NR	36	12 (NR)	NR	12 months: 86% (74.8% to 97.4%)

NR, not reported.

a Propensity score matching (inverse probability weighting) was used to reduce baseline differences between the cohorts.

b Sensitivity analyses were also carried out for OS on a cohort with at least 6 months follow-up in SACT. To identify the cohort, CDF applications were limited from 5 April 2019 to 28 October 2020.

Note

Kidney Cancer UK audit report and the NCRAS data are reported in a separate table as OS was reported by disease stage or postoperative survival rather than by intervention.

TABLE 54 Progression-free survival estimates from RWE

Study	LOT	Intervention	Median follow-up	ToT	*N*	Median PFS mths (95% CI)	PFS rate %
UK RWE 2022[28]	1L	Suni; cabo; nivo + ipi; pazo; tivo	16.8 months (15.8 to 17.6)	7.10 (6.44 to 7.59)	1319	11.93 (10.81 to 13.86)	NR
	2L	Axi; cabo; lenv + evero; nivo; pazo; suni; tivo		Measured with PFS	604	9.89 (8.21 to 11.50)	NR
	3L	Axi; cabo; lenv + evero; nivo; suni		Measured with PFS	202	6.90 (5.52 to 9.69)	NR
	4L	Axi; evero		Measured with PFS	48	3.68 (2.23 to 4.60)	NR
Hack *et al.* (2019)[72]	2L; 3L; 4L+	Nivo	NR	NR	109	5.4 (NR)	NR
Hilser *et al.* (2023)[77]	1L	Cabo + nivo	8.3 (NR)	NR		NR	6 months 81.9%
Nathan *et al.* (2022)[70]	1L	Ave + axi	12 (NR)	NR	36	12 (NR)	NR

TABLE 55 Time to next treatment estimates from RWE

Study, year	*N*	LOT → LOT	Median time (months) to next treatment (95% CI)
UK RWE 2022[28]	1319 1L → 604 2L	1L → 2L	10.1 (9.4 to 10.8)
RECCORD Wagstaff *et al.* (2016)[6]	514 1L → 81 2L	1L → 2L	2009–10: mean 17.4 (SD 11.8) 2010–1: mean 12.3 (SD 7.1) 2011–2 cohort: mean 6.3 (SD 3.7)
SACT TA780[23]	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)

SD, standard deviation.

TABLE 56 Discontinuation estimates from RWE

Study, year	LOT	*N*	Median follow-up months (95% CI)	Discontinuations, *n* (%)	Median TTD (months) to discontinuation (95% CI)	Reason for discontinuation, *n* (%)
UK RWE 2022	1L	1319	16.8 months (15.8 to 17.6)	1049 (79.5)	Treatment duration by treatment type at 1L in appendix L of the original EAG report	Death: 101 (19.6); PD: 664 (63.3); patient choice: 23 (2.2); toxicity: 227 (21.6); other: 34 (3.2)
	2L	604		464 (76.8)	Treatment duration by treatment type at 2L in appendix L of the original EAG report	Death: 26 (5.6); PD: 323 (69.6); patient choice: 9 (1.9); toxicity: 97 (20.9); other: 9 (1.9)
	3L	202		144 (71.3)	Treatment duration by treatment type at 3L in appendix L of the original EAG report	Death: 19 (13.2); PD: 110 (76.4); patient choice: 1 (0.7); toxicity: 12 (8.3); other: 2 (1.4)
	4L	48		38 (79.2)	Treatment duration by treatment type at 4L in appendix L of the original EAG report	Death: 7 (18.4); PD: 22 (57.9); patient choice: 3 (7.9); toxicity: 4 (10.5); other: 2 (5.3)

continued

TABLE 56 Discontinuation estimates from RWE (*continued*)

Study, year	LOT	*N*	Median follow-up months (95% CI)	Discontinuations, *n* (%)	Median TTD (months) to discontinuation (95% CI)	Reason for discontinuation, *n* (%)
Hawkins *et al.* (2020)[73]	1L	652	23.8 (22.2 to 25.4)	574 (88.0)	10.5 (9.5 to 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 to 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 to 9.5)	Disease progression 11 (68.8); treatment toxicity/AE 5 (31.3); other 2 (12.5)
Wagstaff *et al.* (2016)[6]	1L	514	13.1 (12.0 to 14.1)	97 (18.9);[a] 27 (17.1)[b]	4.0 (0.2 to 5.8) (time to treatment discontinuation of a first-line drug)	NR
	2L	81		12 (14.8); 0 (0)	NR	NR
	3L	16		2 (12.5); 0 (0)	NR	NR
SACT TA780[23]	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 months 65 (14%); died on treatment 24 (5%); completed as prescribed 23 (5); patient choice 2 (< 1%); COVID 2 (< 1%)
Nathan *et al.* (2022)[70]	1L	36	NR	5	NR	Disease progression 4 (11); toxicity 1 (3)
CARINA Nathan *et al.* (2023)[77]	1L	118	NR	NR	10.2 weeks (9.1 to 17.1)	NR
	1L subgroup of cabo 2L	83	NR	NR	9.1 weeks (8.1 to 12.0)	NR
	2L	129	NR	NR	23.6 weeks (14.0 to 28.3)	NR
	2L cabo subgroup	87	NR	NR	28.1 weeks (20.1 to 37.1)	NR

a Includes *n* = 35 patients who changed to a different first-line treatment due to toxicity.
b As a percentage of patients who already experienced one dose decrease.

Appendix 4 Economic evaluations and cost and resource use studies

TABLE 57 Summary of published economic evaluations of cabo + nivo (1)

	Li (2021)	Liao *et al.* (2021)	Liu (2022)	Marciniak 2022
Analysis country	USA	USA	USA	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[a] and combinations[b]
Model structure	DES based on PFS, discontinuation and mortality due to AEs, lifetables and OS during BSC Curve selection not justified	Three-state PartSA Extrapolation methods unclear	Three-state models: state transition and PartSA Curve selection statistical and visual fit only	Three-state PartSA Curve selection statistical fit only
Source of efficacy data	CheckMate 9ER (March 2020 DBL), AXIS, TIVO-3, dovitinib vs. sora RCT[37,82,97,102]	CheckMate 9ER (March 2020 DBL)[37]	CheckMate 9ER (March 2020 DBL)[37]	CheckMate 9ER[37] (Sept 2020 DBL) NMA for comparators
Price of cabo 60 mg/nivo 240 mg	$491.30/$6849.84 (average CMS sale price)	$866.51/$8015.04 (Red Book)	$515/$7432 (average CMS sale price)	NR
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	NR
Utility sources	Cella 2018 (METEOR)[115] De Groot 2018 (PERCEPTION)[116] Wan 2019 (CheckMate 214)[117] Patel 2021 (myeloma)[118] Wu 2018 (VEG105192 trial)[103] Selection methods unclear	Wan 2017[119] Wan 2019[117] Wu 2018[120] Data not from CheckMate 9ER. Selection methods unclear	Cabo + nivo estimated from FKSI Wan 2019[117]	CheckMate 9ER
Subsequent 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	NR but appears to be payer
Base-case ICER	$508,987/QALY	$863,720/QALY	$555,663/QALY vs. $531,748/QALY[c]	Uses placeholder costs for some inputs 7.4 LYs, 5.4 QALYs for both nivo + ipi and cabo + nivo LY range, 5.1–6.2; QALY range, 3.8–4.6 for TKIs LY 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	NR

BRL, Brazilian Real; CMS, Centers for Medicare and Medicaid Services.
a TKIs included: cabo, pazo, tem, tivo, sorafenib, suni.
b Combinations: nivo + ipi, axi + ave, axi + pem, lenv + pem.
c State transition vs. PartSA.

TABLE 58 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	Three-state PartSA Extrapolation methods unclear	Three-state PartSA Curve selection statistical and visual fit only	Three-state PartSA Extrapolation methods unclear
Source of efficacy data	CheckMate 9ER Multidimensional TE NMA vs. suni	CheckMate 9ER (March 2020 DBL)	CheckMate 9ER[37] (datacut unclear) NMA for comparators
Price of cabo 60 mg/ nivo 240 mg	NR	$491.20/$3482.57	NR
Utilities	NR	PFS cabo + nivo 0.848, PFS suni 0.73, progressed 0.66 −0.157 for grade 3+ AEs	NR
Utility sources	CheckMate 9ER French value set	Li 2021, Liao *et al.* (2021)	CheckMate 9ER
Subsequent therapy	NR	CheckMate 9ER	Clinical studies, source and data NR
Perspective	All payers	Health system	NR
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	Multidimensional TE NMAs	Drug costs, utilities at progression, subsequent treatment	RDI, discount rate, drug costs

BRL, Brazilian real.

TABLE 59 Summary of cost and resource use information from published studies

	Amdahl (2017)	Edwards *et al.* (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-naive 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	1100 (COMPARZ)	Sample size of the included studies ranged from 14 to 362	1096
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 AEs costs Subsequent therapy costs	Drug and administration costs Disease management/ health-state costs Terminal care costs AEs costs
Source of resource use estimates	HCRU data sourced from post hoc analysis of COMPARZ trial.[146] Data collected included medical office visits, laboratory visits and tests, home health care, hospitalisation, 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 2011–2,[147] adjusted to 2014 prices using the Consumer Price Index for health.[148]	NHS reference costs 2014–15,[149] PSSRU 2015[150]	NHS reference costs 2014–5,[149] PSSRU 2015[150]
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[25] and, for suni, the first treatment cycle (i.e. 28 days of treatment in first 6 weeks) was provided at no cost[24]	BNF	BNF Dosing and administration schedules from relevant trials, publications or NICE TAs[33,82,151]
Source of terminal care costs	Terminal care costs not considered	Based on Nuffield Trust report 2014[152]	Based on Nuffield Trust report 2014

AG, assessment group; GBP, Great British pounds.

TABLE 60 Summary of cost and resource use information from previous NICE TAs

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 AEs, 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 2017–8	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 2017–8	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 2015–6 and 2016–7	BNF	Based on Nuffield Trust report 2014, inflated to 2016–7
TA542	2018	1L int/poor risk	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, aligned with TA512 and TA215	PSSRU 2016, NHS reference costs 2016–7	BNF	Based on Nuffield Trust report 2014, inflated to 2017
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 2015–6	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 2015–6	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	Estimated by UK clinicians	PSSRU 2015, NHS reference costs 2015–6	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[153]	PSSRU 2015, NHS reference costs 2014–5	BNF	Guest *et al.* and Coyle *et al.*

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library