Axitinib, cabozantinib, everolimus, nivolumab, sunitinib and best supportive care in previously treated renal cell carcinoma: a systematic review and economic evaluation

Steve J Edwards,* Victoria Wakefield, Peter Cain, Charlotta Karner, Kayleigh Kew, Mariana Bacelar, Natalie Masento and Fatima Salih

BMJ Technology Assessment Group, BMJ, London, UK

*Corresponding author sedwards@bmj.com

Declared competing interests of authors: The *British Medical Journal* (BMJ) Technology Assessment Group (TAG) and editorial team of the BMJ work independently of one another. The views and opinions expressed in this report are those of the BMJ-TAG. No competing interests were declared that affect the impartiality of this report.

Published January 2018
DOI: 10.3310/hta22060

Scientific summary

Treatment comparisons for previously treated renal cell carcinoma

Health Technology Assessment 2018; Vol. 22: No. 6
DOI: 10.3310/hta22060

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Treatments for advanced or metastatic renal cell carcinoma (amRCC) aim to prevent or slow further spread by targeting pathways that support tumour growth. Patients who do not tolerate first-line treatments or whose disease has progressed may require subsequent therapy. There is a lack of evidence about the relative clinical effectiveness and cost-effectiveness to guide decisions between treatments, several of which have recently been approved for use in the NHS for second-line, and above, treatments.

Objectives

The objectives of this systematic review are to:

- evaluate the clinical effectiveness and cost-effectiveness of axitinib (Inlyta®, Pfizer Inc., NY, USA), best supportive care (BSC), cabozantinib (Cabometyx®, Ipsen, Slough, UK), everolimus (Afinitor®, Novartis, Basel, Switzerland), nivolumab (Opdivo®, Bristol-Myers Squibb, NY, USA), and sunitinib (Sutent®, Pfizer, Inc., NY, USA) for treated amRCC in line with their respective marketing authorisations
- identify key areas for further primary and secondary research.

The review focuses on patients who have received prior vascular endothelial growth factor (VEGF)-targeted therapy and not prior cytokines to reflect treatment sequences in UK clinical practice.

Methods

A systematic review was undertaken to compare the clinical effectiveness of treatments using mixed-treatment comparison (MTC). Randomised controlled trial (RCT) data were preferred but did not link all treatments in the network and so non-RCTs were sought to link in axitinib and sunitinib for the primary outcomes. Studies comparing treatments of interest with sorafenib (Nexavar®, Bayer, Leverkusen, Germany) were also included to connect the network. Eligible studies compared two or more treatments of interest for people with amRCC previously treated with VEGF-targeted therapy. Placebo has been used as a surrogate for BSC. Primary outcomes were overall survival (OS) and progression-free survival (PFS). Secondary outcomes were objective response rate (ORR), adverse events of treatment and health-related quality of life (HRQoL).

The databases MEDLINE, EMBASE and The Cochrane Library were searched from inception to January and June 2016 for RCTs and non-RCTs, respectively. Additional searches were conducted of reference lists of included studies and systematic reviews, conference abstracts and trial registries for ongoing studies. Two or more reviewers sifted the searches, reviewed full papers, abstracted study data and performed critical appraisals.

Fixed-effects MTCs using Bayesian Markov chain Monte Carlo simulation were conducted for OS, PFS and ORR. Primary analyses were limited to RCT data and hazard ratio (HR) (OS and PFS) or odds ratio (OR) (the ORR), with associated 95% credible intervals (CrIs), were used as summary statistics. Non-RCTs and studies rated as being at a high risk of bias were included in sensitivity analyses (SAs). Subgroup analyses to explore the effect of prior therapies and baseline prognostic scores were also carried out for OS and PFS. HRQoL and adverse event (AE) data were summarised narratively owing to inconsistencies in data reporting.

An additional systematic review was undertaken to search for published cost-effectiveness analyses, costing studies and quality-of-life studies in patients with amRCC. A review of National Institute for Health and Care Excellence (NICE) technology appraisals (TAs) of second-line treatments for amRCC was also undertaken.

A cost–utility analysis comparing axitinib, BSC, cabozantinib, everolimus and nivolumab was performed by developing a partitioned survival model in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA). The model consisted of three health states: pre progression, post progression and death. Outcomes were calculated at each 2-weekly cycle up to a time horizon of 30 years. The perspective was reflective of the NHS in England.

To estimate the expected proportion of patients in each health state at each cycle, parametric survival curves were fitted to digitised Kaplan–Meier data taken from published plots for PFS and OS from the CheckMate 025 trial (Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, *et al.* Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015;**373**:1803–13). This enabled the proportions to be calculated for the nivolumab and everolimus groups, while HRs from the MTC were applied to estimate the proportions for each of the remaining treatments.

A range of scenario analyses were performed, as were probabilistic and deterministic SAs.

Results

Twelve studies ($n = 4144$) met the inclusion criteria: four RCTs (one double-blind RCT and three open-label RCTs) and eight non-RCTs (six retrospective cohort studies and two crossover RCTs in which only second-phase data were relevant). Populations were predominantly male and white, and the mean age was generally between 60 and 70 years. When reported, most patients had stage 3 or 4 clear-cell renal cell carcinoma (RCC) and reasonably good baseline performance status.

The primary PFS analysis, based on two RCTs (RECORD-1 and METEOR), included cabozantinib, everolimus and BSC and showed statistically significant benefits for cabozantinib and everolimus compared with BSC (HR 0.17, 95% CrI 0.12 to 0.24; and HR 0.33, 95% CrI 0.25 to 0.43, respectively), and for cabozantinib compared with everolimus (HR 0.51, 95% CrI 0.41 to 0.63).

A SA for PFS connected axitinib and sunitinib by including five non-RCTs and a third RCT; this analysis showed statistically significant benefits of all active treatments compared with BSC (everolimus HR 0.33, 95% CrI 0.25 to 0.43; cabozantinib HR 0.17, 95% CrI 0.12 to 0.24; axitinib HR 0.31, 95% CrI 0.21 to 0.44; and sunitinib HR 0.27, 95% CrI 0.17 to 0.40). Cabozantinib showed a statistically significant benefit compared with all other treatments: everolimus (HR 0.51, 95% CrI 0.41 to 0.63), sunitinib (HR 0.63, 95% CrI 0.44 to 0.95), axitinib (HR 0.54, 95% CrI 0.40 to 0.76) and BSC (HR 0.17, 95% CrI 0.12 to 0.24). None of the differences in PFS between sunitinib, everolimus and axitinib was statistically significant. Cabozantinib was found to have a 99% probability of being the most effective treatment for improving PFS. Data were not available to provide a robust estimate of PFS for nivolumab compared with other treatments.

The primary OS analysis, based solely on RCT data, included cabozantinib, everolimus, nivolumab and BSC, and did not show statistically significant benefits for any treatment compared with BSC. This is likely to be due to uncertainty in the efficacy of BSC caused by RECORD-1 (Cella D, Michaelson MD, Bushmakin AG, Cappelleri JC, Charbonneau C, Kim ST, *et al.* Health-related quality of life in patients with metastatic renal cell carcinoma treated with sunitinib vs interferon-alpha in a phase III trial: final results and geographical analysis. *Br J Cancer* 2010;**102**:658–64) requiring crossover adjustment for confounding. All mean estimates were in favour of the active treatments. Cabozantinib and nivolumab led to longer OS compared with everolimus (HR 0.66, 95% CrI 0.53 to 0.82; and HR 0.73, 95% CrI 0.60 to 0.89, respectively); however, the difference between nivolumab and cabozantinib was not statistically significant (HR 1.12, 95% CrI 0.82 to 1.49). Cabozantinib was associated with the highest probability of being the best treatment for this outcome (72%).

The results of the SA for OS, including data to compare treatments with axitinib, were in keeping with those of the primary analysis. Everolimus, cabozantinib and nivolumab showed longer OS compared with axitinib (HR 0.74, 95% CrI 0.56 to 0.99; HR 0.48, 95% CrI 0.34 to 0.71; and HR 0.54, 95% CrI 0.38 to 0.77, respectively). Data were not available to provide an OS estimate for sunitinib compared with the other treatments and there was statistically significant inconsistency in the network for this SA.

The primary ORR analysis, based on three RCTs including cabozantinib, everolimus, nivolumab and BSC, showed statistically significant benefits of all treatments compared with BSC. Cabozantinib and nivolumab resulted in statistically significant improvements in ORR compared with everolimus (OR 6.67, 95% CrI 3.28 to 12.78; and OR 6.18, 95% CrI 3.75 to 9.84, respectively). The difference between nivolumab and cabozantinib was not statistically significant for ORR (OR 1.05, 95% CrI 0.41 to 2.18). CheckMate 025 (nivolumab vs. everolimus) was rated as being at a high risk of bias owing to the absence of blinding of outcome assessors for response and METEOR (cabozantinib vs. everolimus) [Choueiri TK, Escudier B, Powles T, Tannir NM, Mainwaring PN, Rini BI, *et al.* Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *Lancet Oncol* 2016;**17**:917–27] was rated as being at an unclear risk of bias for missing data, but the impact of these potential biases on the overall direction of treatment effects is unknown.

Treatments could not be compared using MTC for HRQoL as different measures and tools were used for assessments. HRQoL scores were similar between axitinib and sorafenib in AXIS (Motzer RJ, Escudier B, Tomczak P, Hutson TE, Michaelson MD, Negrier S, *et al.* Axitinib versus sorafenib as second-line treatment for advanced renal cell carcinoma: overall survival analysis and updated results from a randomised phase 3 trial. *Lancet Oncol* 2013;**14**:552–62) and results favoured nivolumab over everolimus in CheckMate 025. Results in RECORD-1 favoured BSC over everolimus, although this effect was only apparent if models were used to account for data not missing at random. METEOR results were similar for everolimus and cabozantinib. The way that AEs were defined and categorised differed across studies and, therefore, no MTC could be conducted and the narrative synthesis was limited to RCT data. The rate of grade 3/4 AEs was higher with everolimus (36.5%) than nivolumab (18.7%), and higher with cabozantinib (71.0%) than everolimus (59.9%).

In the assessment of cost-effectiveness, the base-case analysis showed that nivolumab was the most expensive treatment overall, followed closely by cabozantinib, at £107,000 and £106,000, respectively. Axitinib incurred a significantly lower cost of £46,000, while everolimus totalled £36,000. BSC had a mean overall cost of £11,000.

The most effective treatment in the base-case analysis was cabozantinib, which accrued a mean of 1.87 quality-adjusted life-years (QALYs) over the time horizon of the model. This was noticeably more effective than nivolumab, which accrued 1.60 QALYs, which was markedly better than everolimus and axitinib, which each accrued 1.31 QALYs. BSC accrued just 0.75 QALYs over the time horizon. These values were largely driven by differences in OS, which led cabozantinib to have an expected mean survival of 3.18 years compared with just 2.53 for nivolumab. Everolimus and axitinib both resulted in a mean of 2.21 life-years due to the assumption that axitinib was as effective as everolimus, and BSC had an associated survival of 1.25 years.

These results mean that everolimus dominated axitinib as it accrued the same number of QALYs but incurred a lower cost, while nivolumab was dominated by cabozantinib, which accrued more QALYs at a slightly lower cost. The incremental analysis then simplifies to a comparison between cabozantinib, everolimus and BSC, resulting in an incremental cost-effectiveness ratio (ICER) of £45,000 per QALY for everolimus compared with BSC, and £126,000 per QALY for cabozantinib compared with everolimus.

A probabilistic SA was performed with 10,000 samples, resulting in similar results of £45,000 per QALY for everolimus compared with BSC, and £123,000 per QALY for cabozantinib compared with everolimus. Deterministic SAs showed that the most sensitive parameters were the OS HR and relative dose intensity (RDI) for the active treatments. When the upper values were used for the OS HRs relative to everolimus, the

ranking changed so that everolimus became optimal at a threshold of £30,000 per QALY. The same was also true for the lower bound of the RDI for everolimus but, when the upper bound was used, BSC remained optimal and axitinib became preferable to everolimus. Axitinib also became preferable to everolimus when the upper RDI value was used for axitinib. The ranking of nivolumab and cabozantinib changed when the lower RDI value for nivolumab or the upper RDI value of cabozantinib was used, resulting in nivolumab being preferable to cabozantinib at a threshold of £30,000 per QALY.

A range of scenario analyses were performed including varying the distributions applied for OS and using the HRs derived from an extended network in the MTC to include CheckMate 025 as well as the identified observational evidence. Axitinib and nivolumab were dominated in all scenarios as with the base-case analysis, and the ICERs for everolimus compared with BSC ranged from £38,000 to £46,000 per QALY, in comparison with the base-case ICER of £45,000 per QALY. For cabozantinib compared with everolimus, the ICERs ranged from £102,000 per QALY to £248,000 per QALY, in comparison with the base-case ICER of £126,000 per QALY.

Discussion

This review was conducted according to robust methods that were prespecified in a prospectively registered protocol. The primary analyses bring together high-quality evidence from RCTs for the most pertinent outcomes in this population, using MTC when possible to estimate relative treatment effects in the absence of head-to-head evidence. The inclusion criteria were widened to incorporate comparative observational evidence in SAs to substantiate the primary results and to provide estimates for all treatments of interest.

Treatment comparisons were limited by a small number of RCTs. The proportional hazards assumption did not hold for PFS in CheckMate 025, which prevented the inclusion of nivolumab; randomised evidence for axitinib was limited to a subgroup analysis of AXIS that did not connect to the other RCTs in the network; and imprecision surrounding BSC (informed by RECORD-1) led to counterintuitive results in the OS analysis. SAs incorporating non-randomised evidence provided relative effects for more treatments, but introduced inconsistency and probably bias.

Planned subgroup analyses for prior therapies and baseline prognostic score could not provide results for all treatments and there were too few studies informing the MTC to support additional analyses to explore whether or not observed inconsistencies [e.g. everolimus AE rates in METEOR (59.9%) and RECORD-1 (36.5%)] could be explained by design or between-group baseline differences.

The main limitation of the review is that the costs are based on the list prices of the drugs. There are patient access schemes (PASs) in place to provide these drugs on the NHS with a reduced price or pricing strategy. The details of these PASs are confidential and so could not be incorporated in the analysis. This limits the applicability of the results, which may not reflect current practice in the UK. A strength of the analysis is the range of models tested to fit survival models, which included flexible spline models that proved to have a very good fit to the PFS data in CheckMate 025. The analysis explored a range of scenarios to test different assumptions on the base-case results.

A range of SAs was performed including a probabilistic analysis with a large number of samples. The robustness of the results has therefore been thoroughly tested and the model was found to only be sensitive to a few key parameters: the relative OS of treatments and the RDI for each treatment. These results are not surprising as OS is an influential driver on the total QALYs by definition and the RDI has an impact on the treatment acquisition costs, which make up the majority of overall treatment costs.

Conclusions

The current evidence base to inform decisions between axitinib, cabozantinib, everolimus, nivolumab, BSC and sunitinib for previously treated amRCC is limited by the number of studies providing comparative clinical effectiveness data, and by the quality of study reporting. Analyses of PFS and OS suggest that cabozantinib is likely to be the most effective treatment, closely followed by nivolumab, and with little difference between axitinib, everolimus and sunitinib. All treatments considered in this review appear to delay disease progression and prolong survival more than providing BSC. Cabozantinib is not yet available for use in the NHS in England, although it is currently undergoing appraisal by NICE.

High-quality RCT data comparing all the available RCC treatment options are required to enable more robust estimates of efficacy, including RCTs comparing newer RCC therapies with more established treatments. Further PFS data from a RCT are also required for nivolumab to enable its inclusion in a MTC and more standardised reporting of response rates, HRQoL (e.g. EuroQol-5 Dimensions) and AEs in RCTs would facilitate direct comparisons of the RCC treatments.

The economic analysis showed that the majority of current treatments for second-line RCC are very expensive and unlikely to be cost-effective at list price. The exception to this is everolimus, which may be cost-effective at the NICE threshold of £50,000 per QALY granted to treatments that qualify as an end-of-life treatment. All drugs assessed in this analysis have confidential PASs that provide them to the NHS at a discounted price. The economic results may therefore not fully reflect the current NHS setting and should be considered with caution.

Study registration

This study is registered as PROSPERO CRD42016042384.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

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. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports 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

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

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.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 16/58/01. The contractual start date was in May 2015. The draft report began editorial review in April 2017 and was accepted for publication in September 2017. 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' report 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 report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).