

Temporary treatment cessation compared with continuation of tyrosine kinase inhibitors for adults with renal cancer: the STAR non-inferiority RCT

Fiona Collinson,¹ Kara-Louise Royle,¹ Jayne Swain,¹ Christy Ralph,² Anthony Maraveyas,³ Tim Eisen,⁴ Paul Nathan,⁵ Robert Jones,⁶ David Meads,⁷ Tze Min Wah,⁸ Adam Martin,⁷ Janine Bestall,⁹ Christian Kelly-Morland,¹⁰ Christopher Linsley,¹ Jamie Oughton,¹ Kevin Chan,¹¹ Elisavet Theodoulou,¹² Gustavo Arias-Pinilla,¹² Amy Kwan,¹³ Luis Daverede,¹⁴ Catherine Handforth,¹² Sebastian Trainor,¹⁵ Abdulazeez Salawu,¹³ Christopher McCabe,¹⁶ Vicky Goh,¹⁷ David Buckley,¹⁸ Jenny Hewison,⁹ Walter Gregory,¹ Peter Selby,² Julia Brown¹ and Janet Brown^{12*} on behalf of all the STAR investigators

¹Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK

²Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Leeds, UK

³Academic Oncology, Faculty of Health Sciences, Hull York Medical School, Queens Centre Oncology and Haematology, Hull, UK

⁴Department of Oncology, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge, UK

⁵Department of Oncology, Mount Vernon Cancer Centre, East and North Hertfordshire NHS Trust, Hertfordshire, UK

⁶School of Cancer Sciences, University of Glasgow, Beatson West of Scotland Cancer Centre, Glasgow, UK

⁷Academic Unit of Health Economics, University of Leeds, Leeds, UK

⁸Department of Diagnostic and Interventional Radiology, Leeds Teaching Hospitals Trust, Leeds, UK

⁹Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

¹⁰Department of Cancer Imaging, King's College London, London, UK

¹¹Medical Oncology, Weston Park Cancer Hospital, Sheffield, UK

¹²Division of Clinical Medicine, University of Sheffield, Weston Park Hospital, Sheffield, UK

¹³Academic Unit of Clinical Oncology, University of Sheffield, Sheffield, UK

¹⁴Department of Clinical Oncology, Austral University Hospital, Buenos Aires, Argentina

¹⁵St James's Institute of Oncology, St James's University Hospital, Leeds, UK

¹⁶Institute of Health Economics, University of Alberta, Edmonton, Canada

¹⁷School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

¹⁸Faculty of Medicine and Health, School of Medicine, University of Leeds, Leeds, UK

*Corresponding author j.e.brown@sheffield.ac.uk

Published August 2024

DOI: 10.3310/JWTR4127

Scientific summary

Temporary treatment cessation compared with continuation of tyrosine kinase inhibitors for adults with renal cancer: the STAR non-inferiority RCT

Health Technology Assessment 2024; Vol. 28: No. 45

DOI: 10.3310/JWTR4127

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

Scientific summary

Background

There is increasing interest in using treatment breaks in oncology, to reduce toxicity without compromising efficacy. STAR was designed to determine if a tyrosine kinase inhibitor (TKI) drug-free interval strategy (DFIS) was non-inferior to a conventional continuation strategy (CCS) in the first-line treatment of advanced renal cell carcinoma (RCC).

Objectives

The overall primary objective was to determine whether a sunitinib or pazopanib DFIS is non-inferior in terms of overall survival (OS) and quality-adjusted life-year (QALY) compared to a sunitinib or pazopanib CCS in patients with locally advanced and/or metastatic clear cell RCC.

Secondary objectives included comparing a DFIS to a CCS in terms of summative progression-free interval, time to strategy failure, time to treatment failure, toxicity (common terminology criteria for adverse events v.4.0), quality of life (QoL) [Functional Assessment of Cancer Therapy – Kidney Symptom Index (FKSI-15), Functional Assessment of Cancer Therapy-G (FACT-G), EuroQol-5 Dimensions (EQ-5D™) and EQ-Visual Analogue Scale (VAS™)], cost effectiveness, progression-free survival (PFS).

Three ancillary studies were also included in the trial. The Patient Preference and Understanding Study was designed to understand the participants' experiences of taking part in the trial. The dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) substudy was designed to investigate whether early DCE-MRI could predict progressive disease (PD) at 24 weeks. The computerised tomography (CT) substudy was designed to determine CT as a potential biomarker and predictor of PD at 24 weeks.

Methods

STAR was a UK Phase II/III multicentre, open-label, parallel-group, randomised controlled non-inferiority (NI) trial. The trial recruited adults, with histologically confirmed locally advanced or metastatic clear cell, uni-dimensionally measurable, RCC who required but had not received prior systemic therapy.

Participants were required to have an Eastern Cooperative Oncology Group performance status of 0–1, meet pre-specified blood parameters prior to randomisation, provide written informed consent, be able and willing to comply with the terms of the protocol and follow approved pregnancy prevention guidelines. Exclusion criteria included pulmonary or mediastinal disease, life expectancy < 6 months, previous treatment with or known contraindications or hypersensitivity to TKIs, untreated brain metastases, concurrent or previous invasive cancers which could confuse diagnosis or end points, use of contraindicated concomitant medication or substances and poorly controlled hypertension.

At trial entry, participants were randomised (1 : 1) centrally by Leeds CTRU to a CCS or a DFIS. Randomisation was stratified by Motzer prognostic group (favourable, intermediate, poor), trial site, gender, age (< 60, ≥ 60 years), disease status (metastatic, locally advanced), previous nephrectomy (yes, no), TKI (sunitinib, pazopanib). Both strategies received an initial 4 cycles of trial treatment [sunitinib: 1 cycle of treatment refers to 50 mg (starting dose) on days 1–28, repeated every 42 days; pazopanib: 1 cycle of treatment refers to 800 mg (starting dose) on days 1–42, repeated every 42 days] following which participants in the DFIS arm took up their first treatment break. Following disease progression on a treatment break, DFIS participants restarted their treatment. Additional breaks were permitted

following a further 4 cycles of treatment at the treating clinician's discretion. The trial strategy continued until death, progression while receiving treatment, unacceptable toxicity (clinical withdrawal) or participant withdrawal. Following cessation of the trial strategy, participants were followed up after 6 months and then annually thereafter until death, patient withdrawal or the end of the trial.

All trial data were collected on trial-specific case report forms (CRFs). All CRFs, apart from the questionnaires for the patient-reported outcome measures (FSKI-15, FACT-G, EQ-5D and EQ-VAS) were completed by research staff at the site.

Overall survival was defined as time from randomisation to death from any cause. For NI to be concluded, a margin of $\leq 7.5\%$ in OS was pre-specified. The survival at 2 years in the CCS arm was assumed to be 48.5%. A 7.5% NI margin led to the survival at 2 years in the DFIS arm assumed to be at least 41%, leading to a hazard ratio (HR) of 0.812. This along with a one-sided 2.5% significance level, 5% dropout rate and 80% power required 920 participants. Where 80% power would be attained when 720 events were observed.

Quality-adjusted life-years were defined as the time spent in each health state over trial and follow-up, calculated as the utility index of the EQ-5D questionnaire, which was collected frequently (6-weekly for 24 weeks, 2-weekly for 24 weeks, 6-weekly while on study, 6 months post end of trial strategy and annually thereafter). For NI to be concluded a margin of $\leq 10\%$ in mean QALYs was pre-specified. The power for the QALY comparison was derived using simulations. Under the assumptions of a 0.9 HR in favour of CCS, a sample size of 920, 5.83 years of recruitment and 3.25 years of follow-up a CCS QALY estimate of 1.56 was derived along with a power of 69.94%. The CCS QALY estimate of 1.56 in conjunction with the 10% margin equates to a NI boundary of -0.156 for the difference in mean QALYs.

It was pre-specified that NI was required in both intention-to-treat (ITT) and per-protocol (PP) analyses, in both end points, to be concluded overall.

An in-depth qualitative inductive study of 11 patients on the DFIS arm was conducted to better understand their acceptability of extended treatment breaks. A thematic analysis using a comparative and contrastive approach was employed.

An economic evaluation was also conducted which estimated the cost effectiveness of DFIS versus CCS at 2 years (within-trial analysis) and over the patients' lifetime (decision modelling analysis).

Results

Participant flow

Two thousand one hundred and ninety-seven patients were screened; of these, 920 were randomised (CCS: 461, DFIS: 459). In total, 878 (95.4%) participants ceased trial treatment prior to the end of follow-up. Radiological disease progression was the most frequent reason for treatment discontinuation at 43.7% (CCS: 48.4%, DFIS: 39.0%).

Overall, 13,147 (78.6%, $n = 16,726$) (CCS: 5764/7401, 77.9%; DFIS: 7383/9327, 79.2%) of QoL questionnaire booklets were returned on the trial. The proportions of reasons for missing booklets were similar between the two arms, where the highest reason was 'missed by site in error' (overall: 27%, CCS: 28.4%, DFIS: 25.8%). However, the majority of reasons for missing were missing (overall: 52.1%, CCS: 50.6%, DFIS: 53.4%).

Overall, 63 participants (6.8%, $n = 920$) withdrew from some aspect of the trial. This resulted in 21 (2.82%, $n = 920$) participants being formally lost to follow-up (CCS: 11, 2.4%; DFIS: 10, 2.2%). An

additional DFIS participant was lost to follow-up due to the participant moving away. However, all participants were included in the analysis.

Numbers analysed

Nine hundred and nineteen participants (CCS: 461, DFIS: 458) were included in the ITT population as one DFIS participant did not have RCC. Similarly, 871 participants (CCS: 453, DFIS: 418) were included in the PP population, 916 participants (CCS: 485, DFIS: 431) were included in the safety population, 869 participants (CCS: 438, DFIS: 431) were included in the EuroQol-5 Dimensions, three-level version (EQ-5D-3L) population, 856 participants (CCS: 425, DFIS: 431) were included in the FACT-G population and 882 participants (CCS: 436, DFIS: 446) were included in the FKSI population.

Baseline characteristics and treatment (on trial and in follow-up)

Key demographic and disease-related characteristics were similar across both randomisation allocation and randomised under TKI for all analysis populations. STAR participants were predominantly white (ITT – 96.3%), male (ITT – 72.7%) and aged ≥ 60 at randomisation (ITT – 73.4%).

On average, participants in both arms received a similar number of treatment cycles [ITT median (interquartile range, IQR): 5 (2, 10) vs. 4 (2, 9)]. Overall, 248 participants (56.3%, $n = 459$) in the DFIS arm continued on the trial after week 24. The majority of these (127, 51.2%) only started one treatment break. Thirty-eight (15.3%) of these had 2 treatment breaks and 68 (27.4%) took 3 or more treatment breaks. The maximum number of treatment breaks taken by 1 participant was 9, and 15 participants (6%) did not take up a treatment break despite continuing on study after week 24.

Overall, 61.6% of participants received further systemic anticancer therapy treatment during follow-up, with a higher proportion in the CCS arm (68.5%, $n = 461$) compared to the DFIS arm (54.6%, $n = 458$).

Co-primary end points

Median (IQR) follow-up for OS in the PP population was 58 months (46, 72). In total 648 (74.4%, $n = 871$) of PP participants died prior to the end of follow-up (31 December 2020) [CCS: 330 (72.8%, $n = 453$), DFIS: 318 (76.1%, $n = 418$)]. Renal cancer was related to the cause of 607 (93.7%, $n = 648$) deaths [CCS: 314 (95.2%, $n = 330$), DFIS: 293 (92.1%, $n = 318$)]. Median OS [95% confidence interval (CI)] in the CCS arm was 28 months (24, 32) and 27 months (23, 31) in the DFIS arm. At 2 years post randomisation (95% CI), 54.2% (50.8%, 57.5%) of PP participants were alive. This was similar between the two arms [CCS: 55.2% (50.5%, 59.7%), DFIS: 53.1% (48.2%, 57.8%)].

On application of a Cox proportional hazards (PH) model adjusted for the stratification factors of the trial, the HR for randomisation allocation, CCS versus DFIS (95% CI) was 0.94 (0.80 to 1.09). Comparing the lower bound of the CI to the NI boundary of 0.812. At the 2.5% significance level, we do not reject the null hypothesis that DFIS is not non-inferior to CCS in terms of OS. There is insufficient evidence to conclude NI.

However, the analyses conducted in the ITT population along with the sensitivity analysis conducted in the PP population using a piecewise hazards model and a Cox regression model where Motzer Score at randomisation was considered as two categories did conclude NI between the two arms.

The mean QALY (95% CI) for PP participants in the CCS arm was 1.73 (1.60 to 1.86) and 1.80 (1.65 to 1.95) in the DFIS arm derived by combining the results across 52 imputed data sets.

The marginal effect (95% CI) of randomisation allocation, DFIS versus CCS was 0.04 (–0.14 to 0.21). Comparing the lower bound of the CI to the NI boundary of 0.156. As the lower bound of the CI is above –0.156, we conclude that the DFIS arm is non-inferior to the CCS arm in terms of QALYs in the PP population.

This conclusion remained consistent across all the sensitivity analyses; ITT population, derived from week 24, measured up to 12, 24 and 36 months post randomisation, complete case analysis, analysis conducted using a multivariate linear regression analysis and analysis conducted under missing not at random scenario 1 (worst-case scenario).

Secondary end points

All secondary time to event end-point analysis was conducted in the ITT population. With the exception of PFS, all analyses showed a significant difference in favour of the DFIS compared to the CCS when applied in a Cox PH model adjusted for the stratification factors of the trial. As PFS measures to the first progression and does therefore not account for successful rechallenge of TKI in the DFIS arm post progression this was expected.

There was little difference in the summary statistics of the various QoL questionnaire scores and subscales (FKSI-15, FACT-G, EQ-5D-3L and EQ5D-VAS) when considered by randomisation allocation in each of their populations. A difference only appeared around week 294 when few participants remained on the CCS arm. The multilevel modelling conducted on the FKSI-15 and the FACT-G questionnaires concluded some significant differences between the DFIS and the CCS in favour of the DFIS. However, these differences were small and not consistently observed in the sensitivity analysis, which modelled the questionnaires up to regular intervals post randomisation.

The within-trial economic evaluation at 2 years indicated that DFIS was highly likely to be cost-effective. DFIS was found to provide a small QALY benefit (0.049) and substantial cost savings (£3235) over CCS. The value-for-money metric was principally driven by savings in medicine costs. There were relatively minor differences in resource use between arms. The lifetime cost-effectiveness modelling led to the same conclusion yielding QALY gains of 0.08 and cost savings of £2420 (both probabilistic values) in favour of DFIS. The probabilistic sensitivity analyses indicated that DFIS had a 96% chance of being cost saving and a 68% chance of yielding QALY gains versus CCS. The conclusions of both the trial- and model-based analyses were relatively robust to a series of supplementary and sensitivity analyses.

Safety

All participants bar three in the safety population (CCS: 1, DFIS: 2) experienced an adverse event (AE) during the course of the trial. Overall, a higher proportion of participants in the DFIS arm experienced an AE grade 3 or above (CCS: 70.7%, $n = 485$, DFIS: 76.1%, $n = 431$). This difference between arms continues when considering serious adverse events (SAEs) by arm overall and pre and post week 24 (overall – CCS: 45.8%, $n = 485$, DFIS: 54.1%, $n = 431$, pre CCS: 32.8%, $n = 485$, DFIS: 35.5%, $n = 431$, post CCS: 32.8%, $n = 265$, DFIS: 48.4%, $n = 223$). However, this may be explained through the reporting requirements. Participants on DFIS will have reported AEs and SAEs for longer due to their increased time on trial strategy. On consideration of serious AEs deemed to be related to the participants TKI [serious adverse reactions (SARs)], which were collected for the duration of the trial, the difference, in favour of CCS, is more evident prior to week 24 when both strategies are the same (overall – CCS: 18.8%, $n = 485$, DFIS: 23.4%, $n = 431$, pre CCS: 13.2%, $n = 485$, DFIS: 19.3%, $n = 431$, post CCS: 11.7%, $n = 265$, DFIS: 9.4%, $n = 223$). The majority of SAEs and SARs resulted in recovery or recovery with sequelae (SAEs: 80.1%, $n = 744$, SARs: 87.6%, $n = 226$). There were 13 suspected unexpected serious adverse reactions (CCS: 5, DFIS: 8) on the trial.

Substudies

The qualitative substudy concluded that if a DFIS approach was implemented in practice some thought should be given to how patients should be supported during the extended break to cope with and alleviate worries related to ongoing monitoring and tumour growth.

The magnetic resonance imaging (MRI) feasibility study demonstrated that DCE-MRI-derived biomarkers of tumour perfusion (perfused tumour volume, K^{trans} and extracellular volume) as potential surrogate

biomarkers to predict disease progression following TKI therapy in metastatic RCC and the technique of DCE-MRI assessment is reproducible.

The CT substudy demonstrated that assessment of enhancement as well as size change alters the categorisation of response. Use of modified Choi (mChoi) criteria may allow for earlier detection of disease progression, and more representative separation of participants with partial response versus stable disease. While published literature has suggested an association of mChoi criteria with time to progression, in terms of prediction of early progression within 6 months of treatment initiation, no association was shown in this cohort.

Conclusions

Implications for health care

The STAR trial is an exemplar trial, being one of the first to be powered on a QALY end point. It demonstrated no substantial detriment in terms of OS and QALY from a DFIS compared to a CCS. However, NI between the two arms cannot be concluded due to a lack of statistical power for the analysis of OS, as a result of fewer than expected OS events due to the changing landscape of effective treatments for relapsed advanced RCC during the trial. These results provide evidence-based reassurance for people who received National Institute for Health and Care Excellence (NICE)-approved treatment breaks in this setting during the COVID-19 pandemic. Through compliance with the protocol, treatment breaks within the trial were shown to be acceptable to patients and health professionals.

Future research implications

Further discussion is required with patient and public involvement (PPI) and NICE regarding the implications of these results after the NHS interim treatment options following the COVID-19 pandemic end. The STAR trial has provided significant learning regarding the development, design and delivery of intermittent therapy trials.

Trial registration

This trial is registered as ISRCTN06473203.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment Programme (NIHR award ref: 09/91/21) and is published in full in *Health Technology Assessment*; Vol. 28, No. 45. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.6

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 3.6 and is ranked 32nd (out of 105 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 funded by the HTA programme as award number 09/91/21. The contractual start date was in June 2011. The draft manuscript began editorial review in December 2021 and was accepted for publication in March 2023. 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.

Copyright © 2024 Collinson *et al.* This work was produced by Collinson *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

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

