Whole-body MRI compared with standard pathways for staging metastatic disease in lung and colorectal cancer: the Streamline diagnostic accuracy studies

Stuart A Taylor,[1]* Susan Mallett,[2] Anne Miles,[3] Stephen Morris,[4] Laura Quinn,[2] Caroline S Clarke,[5] Sandy Beare,[6] John Bridgewater,[7] Vicky Goh,[8] Sam Janes,[9] Dow-Mu Koh,[10] Alison Morton,[11] Neal Navani,[9] Alfred Oliver,[11] Anwar Padhani,[12] Shonit Punwani,[1] Andrea Rockall[13] and Steve Halligan[1] on behalf of the Streamline investigators

[1]Centre for Medical Imaging, University College London, London, UK
[2]Institute of Applied Health Research, NIHR Birmingham Biomedical Research Centre, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[3]Department of Psychological Sciences, Birkbeck, University of London, London, UK
[4]Applied Health Research, University College London, London, UK
[5]Research Department of Primary Care and Population Health, and Priment Clinical Trials Unit, University College London, London, UK
[6]Cancer Research UK & UCL Cancer Trials Centre, University College London, London, UK
[7]UCL Cancer Institute, University College London, London, UK
[8]Department of Cancer Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[9]Lungs for Living Research Centre, UCL Respiratory, University College London, London, UK
[10]Department of Radiology, The Royal Marsden Hospital, Sutton, UK
[11]c/o Centre for Medical Imaging, University College London, London, UK
[12]Mount Vernon Centre for Cancer Treatment, Mount Vernon Hospital, Northwood, UK
[13]Imaging Department, Imperial College Healthcare NHS Trust, London, UK

*Corresponding author stuart.taylor@ucl.ac.uk

Declared competing interests of authors: Stuart A Taylor reports personal fees from Robarts Clinical Trials Inc. (London, ON, Canada) outside the submitted work. Andrea Rockall reports personal fees from Guerbet (Paris, France) outside the submitted work. Vicky Goh and Anwar Padhani report grants from Siemens AG (Erlangen, Germany) outside the submitted work. Steve Halligan reports non-financial support from iCad Inc. (Nashua, NH, USA) outside the submitted work and was a member of the Health Technology Assessment (HTA) commissioning board (2008–14). Stephen Morris declares past membership of the National Institute for Health Research (NIHR) HTA Clinical Evaluation and Trials Board (2007–9), HTA Commissioning Board (2009–13), Public Health Research Board (2011–17), Health Services and Delivery Research (HSDR) Commissioning Board (2014–16) and HSDR Board (2014–18), and current membership of the NIHR Programme Grants for Applied Research expert subpanel (2015–present) and HSDR Evidence Synthesis Sub Board (2016–present).

Disclaimer: This report contains transcripts of interviews conducted in the course of the research and contains language that may offend some readers.

Published December 2019
DOI: 10.3310/hta23660

Scientific summary

The Streamline diagnostic accuracy studies

Health Technology Assessment 2019; Vol. 23: No. 66
DOI: 10.3310/hta23660

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

Scientific summary

Background

Colorectal and lung cancer are the leading causes of cancer-related deaths in the UK. Accurate staging is fundamental to the treatment and, in particular, the detection of metastatic disease. Cross-sectional imaging forms the bedrock of cancer staging in the NHS, notably computed tomography, positron emission tomography and magnetic resonance imaging. Ultrasonography, plain radiography and scintigraphy are also widely used. For both cancers, the National Institute for Health and Care Excellence provides guidance on staging pathways, detailing the sequential deployment of various tests before staging is deemed complete and treatment can begin. This stepwise deployment of cross-sectional imaging is resource intensive, exposes patients to ionising radiation and increases their anxiety. Whole-body magnetic resonance imaging has been proposed as an alternative to multimodality staging pathways and can image the body in 1 hour or less using magnetic resonance imaging scanners available throughout the NHS. It does not impart ionising radiation and could reduce the time and number of tests required before staging is complete and treatment can begin. Meta-analyses suggest that whole-body magnetic resonance imaging is equivalent to or may surpass the accuracy of standard tests for metastatic disease detection, although the primary literature is of questionable quality and prospective multicentre comparative data are lacking. High-quality evidence is needed to guide implementation.

Objectives

The primary objective of the Streamline trials was to compare the per-patient diagnostic accuracy for metastatic disease of staging pathways utilising initial whole-body magnetic resonance imaging (whole-body magnetic resonance imaging staging pathway) with those following current National Institute for Health and Care Excellence recommendations (standard pathway). We conducted two parallel but separate trials in colorectal cancer (Streamline C) and non-small-cell lung cancer (Streamline L). Additional objectives were a comparison between alternative pathways for the time and number of tests taken before staging completion and the nature of the first major treatment decision, accuracy for local tumour staging, cost-effectiveness, patient experience, interobserver variability of whole-body magnetic resonance imaging interpretation and diagnostic accuracy of time-efficient whole-body magnetic resonance imaging sequence protocols.

Design

We performed two parallel multicentre prospective cohort trials (Streamline C and Streamline L). For both trials, participants were recruited from 16 representative NHS teaching and general hospitals in England. Patients were eligible for Streamline C if they had histologically proven or suspected colorectal cancer (defined as a mass highly suspicious for cancer on endoscopy or imaging) and were referred for staging investigations. Polyp cancers were excluded. Patients were eligible for Streamline L if they had histologically proven non-small-cell lung cancer or this was suspected on chest computed tomography with sufficient confidence to trigger staging investigations/biopsy and deemed radically treatable (stage IIIB or less). Patients with unequivocal metastatic or N3 disease on diagnostic chest computed tomography were excluded. For both trials, patients were excluded if they were aged < 18 years, were pregnant or had contraindications to magnetic resonance imaging or if the final diagnosis was not lung or colorectal cancer as appropriate. Consecutive (i.e. unselected) patients underwent whole-body magnetic resonance imaging (including T1-weighted imaging pre and post gadolinium enhancement and T2- and diffusion-weighted imaging sequences) in addition to all standard staging investigations. Standard investigations were interpreted locally by the usual clinical care team blinded to the whole-body magnetic resonance imaging. Whole-body

magnetic resonance imaging was interpreted by 19 (Streamline C) and 16 (Streamline L) radiologists blinded to all standard staging investigations and clinical data other than the location of the primary tumour. Radiologists recorded the local stage of the primary tumour together with the presence, organ site and size of any metastatic deposits. The whole-body magnetic resonance imaging images and report were withheld from the multidisciplinary team meeting until all standard investigations had completed such that the first major treatment decision could be made (which was recorded). The whole-body magnetic resonance imaging was then revealed to the multidisciplinary team, which stated if additional tests would have been generated had it been the initial staging investigation. The multidisciplinary team then recorded its treatment decision, based on the whole-body magnetic resonance imaging and the results of generated additional tests (if any), the whole-body magnetic resonance imaging staging pathway, and the final treatment decision based on all available tests. We used the construct reference standard paradigm (multidisciplinary panel diagnosis). Participants' clinical courses were followed for 12 months. For each participant, the panel considered the results of all staging and follow-up imaging investigations and all additional information including surgical findings, histopathology and clinical course. Based on all available data, the panel adjudicated on the Tumour Node Metastasis Classification of Malignant Tumors stage of the participant at the time of recruitment and indicated the optimal retrospective treatment decision. For Streamline C, we estimated that a sample size of 290 participants would give 80% power to detect a 10% sensitivity difference between the whole-body magnetic resonance imaging pathway and standard pathways, assuming a 40% prevalence of metastasis. For Streamline L we estimated that a sample size of 200 participants would give 80% power to detect a 24% sensitivity difference between pathways, assuming a 25% prevalence of metastasis. Binary comparisons (sensitivity, specificity, treatment decision agreement) were calculated using paired proportions (population marginal). We calculated the time and number of tests taken to complete staging for each pathway (by adding times for staging tests [from request to performance] to median treatment decision multidisciplinary team wait times). The median difference in time and number of staging tests between pathways was compared for each participant with 95% confidence interval from 2.5 and 97.5 centiles of 1999 bootstrap samples with replacement. We compared the nature of the first major treatment decision with that made by the multidisciplinary team and also with the retrospective optimal treatment decision made by the consensus reference panel. We also tested interobserver variability in whole-body magnetic resonance imaging interpretation and the impact of whole-body magnetic resonance imaging sequences on radiologist accuracy (T1- and diffusion-weighted sequences vs. T1-, diffusion- and T2-weighted sequences vs. T1-, diffusion- and T2-weighted sequences and contrast-enhanced T1-weighted sequences), investigated participant experience staging pathways using interviews, questionnaires and a discrete choice experiment, and performed a cost-effectiveness analysis.

Results

Streamline C

A total of 299 participants completed the trial; 68 (23%) had metastasis. The whole-body magnetic resonance imaging pathway had 67% (95% confidence interval 56% to 78%) sensitivity for participants with metastasis, not significantly different from standard pathways [63% (95% confidence interval 51% to 74%)], a 4% (95% confidence interval –5% to 13%; $p = 0.508$) difference in sensitivity. Specificity was not significantly different [whole-body magnetic resonance imaging pathway 95% (95% confidence interval 92% to 97%) vs. standard pathways 93% (95% confidence interval 90% to 96%)]. The whole-body magnetic resonance imaging staging pathway had 86% (95% confidence interval 74% to 94%) sensitivity for participants in whom the largest metastasis was $\geq$ 1 cm in size, which was not significantly different from that of standard staging pathways [82% (95% confidence interval 69% to 91%)]. As a stand-alone investigation, whole-body magnetic resonance imaging had a comparable sensitivity of 70% (95% confidence interval 59% to 80%) to the standard staging pathway, but a significantly lower specificity [86% (95% confidence interval 81% to 90%) vs. 94% (95% confidence interval 90% to 96%), respectively]. The whole-body magnetic resonance imaging staging pathway had 54% agreement for T stage compared with 60% for the standard pathway, a difference of 6% (95% confidence interval 0% to 12%). There was no significant difference in agreement between the whole-body magnetic resonance imaging and standard

pathways for N staging (58% vs. 56%, respectively), a difference of 2% (95% confidence interval –4% to 7%). Agreement with the final multidisciplinary team treatment decision was 96% and 95% for whole-body magnetic resonance imaging and standard pathways, respectively, and 68% with the retrospective optimal treatment decision for both pathways. Time to complete staging was significantly shorter for whole-body magnetic resonance imaging [8 days (95% confidence interval 6 to 9 days) vs. 13 days (95% confidence interval 11 to 15 days)], a 5-day (95% confidence interval 3- to 7-day) difference. Across the cohort, a total of 558 and 320 investigations were performed as part of standard pathways and the whole-body magnetic resonance imaging staging pathway, respectively (whole-body magnetic resonance imaging would have generated an additional 21 tests). The median number of tests in the whole-body magnetic resonance imaging pathway was significantly lower [1 test (95% confidence interval 1 to 1 test) vs. 2 tests (95% confidence interval 2 to 2 tests)], a difference of 1 test (95% confidence interval 1 to 1 test).

The mean test costs per participant (bootstrapped 95% confidence intervals) were £285 (95% confidence interval £260 to £310) for standard pathways and £216 (95% confidence interval £211 to £221) for the whole-body magnetic resonance imaging pathway.

Streamline L

A total of 187 participants completed the trial; 52 (28%) had metastasis. The whole-body magnetic resonance imaging pathway had 50% (95% confidence interval 37% to 63%) sensitivity for participants with metastasis, not significantly different from standard pathways [54% (95% confidence interval 41% to 67%)], a 4% (95% confidence interval –7% to 15%; $p = 0.73$) difference. Specificity was not significantly different [93% (95% confidence interval 88% to 96%) vs. 95% (95% confidence interval 91% to 98%), respectively]. The whole-body magnetic resonance imaging staging pathway had 82% (95% confidence interval 64% to 92%) sensitivity for participants in whom the largest metastasis was ≥ 1 cm, which was not significantly different from that of standard staging pathways [75% (95% confidence interval 57% to 87%)]. As a stand-alone investigation, whole-body magnetic resonance imaging had comparable sensitivity [50% (95% confidence interval 37% to 63%)] to standard pathways but lower specificity [85% (95% confidence interval 78% to 90%)]. The whole-body magnetic resonance imaging staging pathway had 65% agreement for N stage compared with 75% for the standard pathway, a statistically significant difference of –10% (95% confidence interval –3% to –18%). There was no significant difference in agreement between the whole-body magnetic resonance imaging and standard pathways for T stage (54% vs. 55%, respectively), a difference of 1% (95% confidence interval –8% to 9%). Agreement with the final multidisciplinary team treatment decision was 98% and 99% for whole-body magnetic resonance imaging and standard pathways, respectively, and 83% and 82%, respectively, with the retrospective optimal treatment decision. Time to complete staging was significantly shorter for whole-body magnetic resonance imaging [13 days (95% confidence interval 12 to 14 days) vs. 19 days (95% confidence interval 17 to 21 days)], a 6-day (95% confidence interval 4- to 8-day) difference. Across the cohort a total of 302 and 232 investigations were performed as part of standard and whole-body magnetic resonance imaging staging pathways, respectively (whole-body magnetic resonance imaging would have generated an additional 45 tests). The median number of tests in the whole-body magnetic resonance imaging pathway were similar [1 test (95% confidence interval 1 to 1 test) vs. 1 test (95% confidence interval 1 to 2 tests)], a difference of 0 tests (95% confidence interval –1 to 0 tests). Mean test costs per participant (bootstrapped 95% confidence intervals) were £620 (95% confidence interval £574 to £666) for standard pathways and £317 (95% confidence interval £273 to £361) for the whole-body magnetic resonance imaging pathway.

Participant experience

In general, whole-body magnetic resonance imaging presented a greater challenge than standard scans. Key challenges were the enclosed space, noise and scan duration. Reduced participant tolerance was associated with claustrophobia, pulmonary symptoms and existing comorbidities. Coping strategies facilitated scan tolerance. Perceived whole-body magnetic resonance imaging burden was greater than for computed tomography (mean score 2.09 vs. 1.70, respectively; $p < 0.0001$) and positron emission tomography–computed tomography (2.33 vs. 2.05; $p = 0.003$). However, participants preferred the

whole-body magnetic resonance imaging-based pathway (probability: lung 0.64, colorectal 0.66) if it was equivalent in accuracy, total scan number and time to diagnosis to a standard staging pathway. Preference was stronger if whole-body magnetic resonance imaging reduced time to staging and/or number of tests.

Whole-body magnetic resonance imaging sequence selection

Streamline C

A combination of T1-weighted imaging pre and post gadolinium enhancement and T2- and diffusion-weighted sequences had significantly greater sensitivity for participants with metastatic disease than T1- and diffusion-weighted sequences alone [72% (95% confidence interval 60% to 81%) vs. 63% (95% confidence interval 51% to 73%), respectively], a difference of 9% (95% confidence interval 1% to 17%). Specificity was not significantly different [85% (95% confidence interval 80% to 89%) vs. 84% (95% confidence interval 78% to 88%), respectively]. A combination of non-contrast enhanced T1-, T2- and diffusion-weighted sequences had 70% (95% confidence interval 58% to 80%) sensitivity and specificity of 86% (81% to 90%).

Streamline L

A combination of T1-weighted imaging pre and post gadolinium enhancement and T2- and diffusion-weighted sequences had significantly greater sensitivity for participants with metastatic disease than T1- and diffusion-weighted sequences alone [52% (95% confidence interval 39% to 65%) vs. 42% (95% confidence interval 29% to 56%), respectively], difference of 10% (95% confidence interval 1% to 19%). Specificity was not significantly different [86% (95% confidence interval 79% to 81%) vs. 82% (95% confidence interval 74% to 87%), respectively]. A combination of non-contrast enhanced T1-, T2- and diffusion-weighted sequences had 48% (95% confidence interval 35% to 61%) sensitivity and 84% (95% confidence interval 76% to 89%) specificity.

Whole-body magnetic resonance imaging interobserver agreement

Interobserver agreement for interpretation of whole-body magnetic resonance imaging against the final reference standard was moderate for Streamline C (80% overall agreement against the consensus reference standard, $\kappa = 0.6$), but poor for Streamline L (44% overall agreement against the consensus reference standard, $\kappa = -0.12$). For the primary outcome, not considering the final participant metastatic status based on the consensus reference standard, two reads agreed in 35 out of 40 (88%) whole-body magnetic resonance imaging data sets (Streamline C) and 28 out of 43 (65%) whole-body magnetic resonance imaging data sets (Streamline L).

Conclusions

For both colorectal and non-small-cell lung cancer, the whole-body magnetic resonance imaging staging pathway is as accurate as current standard staging pathways for identifying patients with metastatic disease and results in the same treatment decisions. It is more efficient, reducing time to complete staging, the number of staging tests (in colorectal cancer) and staging costs.

Implications for health care

In a NHS setting, the whole-body magnetic resonance imaging staging pathway achieves similar accuracy as standard staging pathways for detecting patients with metastatic disease in both colorectal cancer and non-small-cell lung cancer. Agreement for local T and N stage is also similar, although in non-small-cell lung cancer, the whole-body magnetic resonance imaging pathway has lower agreement and sensitivity for N stage than standard pathways. However, for both cancers, agreement between the primary treatment decisions based on the whole-body magnetic resonance imaging pathway and both a contemporaneous multidisciplinary team treatment decision and a retrospective optimal treatment decision is nearly identical

to that for decisions based on standard staging pathways. For colorectal cancer, the whole-body magnetic resonance imaging pathway reduces the number of tests required to complete staging, and for both cancers whole-body magnetic resonance imaging significantly reduces the time to complete staging. Based on NHS reference costs, the whole-body magnetic resonance imaging staging pathway is cheaper than standard pathways in both colorectal and non-small-cell lung cancer. Although patients find whole-body magnetic resonance imaging a more burdensome test than standard staging investigations, in general they prefer the whole-body magnetic resonance imaging staging pathway if it at least matches the accuracy and efficiency (time and number of tests) of standard pathways. Time-efficient whole-body magnetic resonance imaging protocols using just T1- and diffusion-weighted sequences have lower sensitivity than protocols that also include T2- and post-gadolinium-enhanced T1-weighted sequences. Interobserver variation in whole-body magnetic resonance imaging interpretation is moderate, and lower for non-small-cell lung cancer than for colorectal cancer.

Recommendations for future research

Future research should investigate:

1. the diagnostic accuracy, patient acceptability and cost-effectiveness of whole-body magnetic resonance imaging for staging and treatment follow-up of other primary cancer sites, notably breast, prostate and myeloma, compared with standard investigations
2. the diagnostic accuracy, patient acceptability and cost-effectiveness of whole-body magnetic resonance imaging for investigating patients with clinically suspected recurrence of colorectal and non-small-cell lung cancer compared with standard investigations
3. the diagnostic accuracy, patient acceptability and cost-effectiveness of whole-body magnetic resonance imaging in routine post-cancer therapy surveillance in comparison with standard investigations
4. the impact on diagnostic accuracy and cost-effectiveness of adding liver-specific contrast agents to whole-body magnetic resonance imaging protocols
5. the impact of formalised training on radiologist performance and interobserver agreement.

Trial registration

These trials are registered as ISRCTN43958015 and ISRCTN50436483.

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: 3.819

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

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 report

The research reported in this issue of the journal was funded by the HTA programme as project number 10/68/01. The contractual start date was in October 2012. The draft report began editorial review in December 2018 and was accepted for publication in May 2019. 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.

This report presents independent research funded by the National Institute for Health 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, NETSCC, 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, NETSCC, the HTA programme or the Department of Health and Social Care.

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