

What carcinoembryonic antigen level should trigger further investigation during colorectal cancer follow-up? A systematic review and secondary analysis of a randomised controlled trial

Bethany Shinkins,¹ Brian D Nicholson,¹ Tim James,²
Indika Pathiraja,¹ Sian Pugh,³ Rafael Perera,¹
John Primrose³ and David Mant^{1*}

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

²Oxford University Hospitals NHS Foundation Trust, Oxford, UK

³Medical Sciences Division, University of Southampton, Southampton, UK

*Corresponding author david.mant@phc.ox.ac.uk

Declared competing interests of authors: none

Published April 2017

DOI: 10.3310/hta21220

Scientific summary

Follow-up of the FACS RCT

Health Technology Assessment 2017; Vol. 21: No. 22

DOI: 10.3310/hta21220

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

Scientific summary

Background

Following primary surgical and adjuvant treatment for colorectal cancer, patients are routinely followed up with blood carcinoembryonic antigen (CEA) testing for 5 years. The Follow-up After Colorectal Surgery (FACS) trial showed that this follow-up is effective at detecting recurrences treatable with curative intent. However, the optimal testing interval and method for interpreting test results lack a firm evidence base. Our initial protocol was restricted to conducting a secondary analysis of CEA testing results from the FACS trial. However, initial work revealed serious limitations of existing reviews of previous research and so we also conducted a formal systematic review following Cochrane guidelines for identifying and meta-analysing studies of diagnostic accuracy.

Aim and objectives

The main aim was to determine how the CEA test result should be interpreted to inform the decision to undertake further investigation to detect treatable recurrences. Secondary objectives were to determine whether or not diagnostic accuracy could be improved by (1) taking account of the baseline CEA level and other pretest patient characteristics; (2) considering the trend in CEA levels over time; (3) considering whether recurrence occurred early or late in follow-up; and (4) changing the testing interval.

Methods for the systematic review

Search

The search details are reported on the Cochrane website [<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011134.pub2/full> (last accessed 6 November 2016)].

Studies included

Cross-sectional diagnostic test accuracy studies, cohort studies or randomised trials, conducted in primary care or hospital settings, involving adults with no detectable residual disease after curative surgery (with or without adjuvant therapy) and reporting results extractable in a 2 × 2 format (i.e. test +/- by case +/-).

Index test

The index test was the blood CEA level.

Reference standard

The reference standard was clinical diagnosis of recurrence of colorectal cancer following primary treatment confirmed by imaging, histology or clinical follow-up [for details of the reference standard see the full report on the Cochrane website: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD011134.pub2/pdf> (accessed 25 August 2016)].

Analysis

Two review authors extracted data independently and three authors independently performed a Quality Assessment of Diagnostic Accuracy Studies 2 assessment of the included studies, with subsequent discussion to reach consensus on overall judgements of risk of bias and applicability. The meta-analyses followed Cochrane guidelines for pooling test accuracy data.

Methods for the main analysis

Design

This study involved a secondary observational analysis of data from the FACS trial, a 2 × 2 pragmatic randomised factorial controlled trial comparing minimum post-surgery follow-up of colorectal cancer patients for 5 years with 3- to 6-monthly blood tests for CEA and 6- to 12-monthly computerised tomography (CT) imaging. We analysed the two arms of the trial that required CEA testing.

Statistical analysis

Receiver operating characteristic (ROC) analysis was implemented to compare the effect of making the decision to trigger further investigation in individual patient on the basis of (1) the CEA level at each test; (2) the difference between test and postoperative CEA levels (expressed both as an absolute value and as a ratio); and (3) the trend in CEA levels over time. An operational analysis of the probable impact of CEA testing if used prospectively in clinical practice was also conducted, hypothetically applying the four most commonly reported thresholds in the systematic review (2.5, 5, 7.5 and 10 µg/l) to trigger further investigation on the basis of each test carried out during the follow-up period. To investigate the diagnostic accuracy of assessing trends in serial CEA measurements within an individual rather than simply interpreting the most recent CEA measurement taken, linear regression models were fitted to the CEA values for each individual over time. The distribution of slope coefficients for individuals who did and did not experience recurrence were compared and ROC analysis was used to evaluate the diagnostic accuracy of CEA trend. All analyses were carried out using the statistical package R [see www.R-project.org/ (accessed 25 August 2016)]; the ROC analysis was carried out using the R package pROC.

Participants

Patients who had undergone curative surgery for primary colorectal cancer and who, after extensive testing (histology, imaging and a CEA level of ≤ 10 µg/l), were confirmed to have no residual disease were recruited from 39 NHS hospitals across all regions of England. The analysis was based on 582 patients from the two arms of the study that received CEA testing.

Carcinoembryonic antigen measurement

The CEA analysis was undertaken using a Siemens Centaur XP analyser (Siemens Healthcare, Erlangen, Germany) at a single laboratory with a standard quality control regime to ensure longitudinal stability. If the blood CEA level was ≥ 7 µg/l above the patient's baseline level at trial entry after repeat measurement, the general practitioner was asked to refer the patient urgently to the local hospital for further investigation. The median number of CEA measurements available for each participant was 13 [interquartile range (IQR) 10–14], with a median of six (IQR 3–9) measurements in patients who developed a recurrence and 14 (IQR 13–14) in those who did not develop a recurrence.

Diagnostic reference standard

The reference standard against which diagnostic accuracy was assessed was clinical diagnosis of recurrence of colorectal cancer as determined by the colorectal cancer multidisciplinary team at the participating hospital centre.

Results

Diagnostic accuracy of a single test

The diagnostic accuracy of CEA testing across all thresholds, estimated on the basis of all CEA tests carried out prior to diagnosis, is modest [area under the receiver operating characteristic curve (AUC) 0.74, 95% confidence interval (CI) 0.68 to 0.80]. Sensitivity is estimated as 50.0% (95% CI 40.1% to 59.9%). The median lead time gained at a recommended threshold of 5 µg/l is about 3 months, but the predictive value would be 62% assuming the same frequency of recurrence experienced in the trial, implying that about four in 10 patients without a recurrence will have at least one false alarm. The positive predictive value of

an individual test (rather than an individual patient) is even lower at 43.3% (95% CI 35.8% to 51.0%), as some patients suffer repeated false alarms. For example, the 89 false alarms triggered at a threshold of 5 µg/l were clustered in 29 individuals, 15 of whom (51.7%) would have more than one false alarm and eight of whom would have more than five false alarms. Trying to improve the sensitivity of CEA testing by reducing the threshold for further investigation has a high cost in terms of falling specificity. Although sensitivity can be increased to 63.5% (95% CI 54.2% to 72.8%) by reducing the threshold to 2.5 µg/l, there is a sevenfold increase in the number of times further investigation is triggered and, in 84% of cases, no recurrence is detected.

Adjusting for postoperative baseline carcinoembryonic antigen level

Adjusting the CEA level by an individual's baseline measurement offers no notable improvement in diagnostic accuracy. Of the 6623 CEA measurements in the database, 3881 (59%) were lower than their baseline measurement and therefore had a negative 'difference' value or a ratio value of < 1; 19 patients who developed recurrence always had a negative adjusted value (i.e. all of their CEA measurements were lower than their baseline level).

Predicting missed cases and false alarms

None of the characteristics assessed (patient age and smoking status, site and stage of the primary tumour, receipt of adjuvant therapy and delay in commencing monitoring and site of recurrence) significantly increased the likelihood of recurrence being missed. However, current smoking was significantly predictive of multiple false alarms (adjusted odds ratio 6.55, 95% CI 1.52 to 28.20; $p = 0.01$).

Diagnostic accuracy of trend

The AUC suggests that the rate of change of CEA level provides better overall discriminatory power than the single-value CEA transformations explored (AUC for positive trend 0.85, 95% CI 0.78 to 0.91). A negative trend (i.e. a reducing level of CEA post treatment) may also have diagnostic value, increasing the AUC to 0.91; however, this improvement is not statistically significant. The optimal threshold for interpreting trend changes over time, to achieve 70% sensitivity at around 90% specificity, falls from 1.75 µg/l in year 1 to 0.3 µg/l in year 5.

Diagnostic accuracy in early and late recurrence

Assessing trend performs better than single test assessment in detecting both early recurrence (recurrence in the first 2 years after treatment) and late recurrence (recurrence in years 3–5), with little difference in accuracy for each time period.

Test interval

At a single test threshold of 5 µg/l, the testing interval needs to be approximately halved in year 1 to ensure that the number of recurrences detectable at each test, and, therefore, test operational performance, remains fairly constant over time. The test interval also needs to be reduced in year 1 if action is to be taken on the basis of trend [with a 3-monthly testing interval the CEA level would not be measured with any precision until month 9, by which time 31 (29.8%) recurrences had already been diagnosed].

Relevance and implications

The importance of not triaging with carcinoembryonic antigen alone

Our main analysis confirms the findings of the systematic review: CEA testing alone is insufficient as a triage test for colorectal cancer recurrence. Whatever threshold is applied to interpret the CEA test result (based on a single test or trend), a significant number of patients will suffer recurrence without a detectable change in CEA levels. This underlines the importance of combining CEA testing with scheduled imaging, as recommended in most national guidelines.

The advantage of making decisions on the trend in carcinoembryonic antigen levels

The diagnostic accuracy of the trend in CEA levels, assessed by the slope (beta-coefficient) of the linear regression line, was consistently better than interpreting the results of a single test, regardless of whether the single test was adjusted for the baseline postoperative CEA level. The observation that optimal performance was achieved by taking account of a negative as well as a positive trend merits further investigation. It suggests that a slow post-treatment reduction in CEA level is itself a marker of recurrence.

The choice of carcinoembryonic antigen threshold

Both the systematic review and the main analysis highlighted the very high cost in terms of false alarms of adopting an action threshold for a single test below the 5 µg/l commonly recommended by national guidelines. The analysis of operational performance suggests that a higher threshold (of perhaps 10 µg/l) may be preferable. Even using trend analysis, the number of false alarms suggests that aiming for a sensitivity of 70% – augmenting CEA testing with a colonoscopy and one or two CT scans to detect the missed 30% of recurrences – may be the clinically preferable option. In applying the trend analysis, the main results also highlighted the importance of not applying the same action threshold throughout the 5-year follow-up period. It is important that thresholds derived from our data are checked and refined in an experimental setting before being rolled out, but, as a starting point, we suggest initiating further investigation if the rate of change in CEA level exceeds 1.7 µg/l/year in year 1, 1.4 µg/l/year in year 2, 0.8 µg/l/year in year 3, 0.5 µg/l/year in year 4 and 0.3 µg/l/year in year 5.

The choice of testing interval

The testing frequency in year 1 would need to be increased to achieve a more consistent level of operational performance over time and to allow for an earlier assessment of trend. A testing schedule of monthly for the first 3 months and then every 2 months for the rest of the year would be consistent with our findings. Adopting this increased testing frequency would be challenging in some health-care systems and would need careful planning (as it requires rapid turnaround of results, good communication with patients and access to a clinician who is able to discuss and act on the results quickly). Although the falling incidence of recurrence would suggest that testing frequency should be reduced to one test in year 5, this would have implications for the achievable lead time.

Who should and should not be followed up with carcinoembryonic antigen

The main analysis of the FACS trial shows that CEA follow-up is appropriate for all patients who have completed surgical and adjuvant treatment for their colorectal cancer and who have, on extensive investigation, no sign of recurrence. It also shows that patients at Dukes' stages A–C have a similar incidence of treatable recurrence and obtain similar benefit. There is no suggestion that patient age, the characteristics of the primary tumour or the recurrence site predict either missed cases from non-response or false alarms. However, the likelihood of multiple false alarms is significantly higher in smokers, suggesting that CEA is not an appropriate follow-up method for patients who continue to smoke.

Other research implications

The systematic review drew attention to the poor quality of the majority of diagnostic studies on CEA testing. Moreover, virtually all studies assessed CEA as a single diagnostic test, ignoring the fact that it is used as a monitoring test and is carried out repeatedly over time. Even the Cochrane diagnostic accuracy methodology that we used to conduct the systematic review considers only single test results, not trend over time. This issue should be addressed, not just in relation to CEA testing but in relation to all tests that are repeated over time to monitor disease progression. The need to refine the suggested cut-off points for monitoring CEA trend (by conducting pilot implementation studies before large-scale roll-out) has already been mentioned. In conducting these studies, it might be preferable to use a reference standard based on recurrence treatable with curative intent rather than any recurrence.

Study registration

This study is registered as PROSPERO CRD42015019327 and ISRCTN93652154.

Funding

The main FACS trial and this substudy were funded by the National Institute for Health Research Health Technology Assessment programme.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.058

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI 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@nhr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nhr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nhr.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.nhr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/136/81. The contractual start date was in March 2013. The draft report began editorial review in March 2016 and was accepted for publication in August 2016. 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. 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.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Shinkins *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

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

Health Technology Assessment Editor-in-Chief

Professor Hywel Williams Director, HTA Programme, UK and Foundation Professor and Co-Director of the Centre of Evidence-Based Dermatology, University of Nottingham, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Health and Wellbeing Research Group, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

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