

Blood CEA levels for detecting recurrent colorectal cancer (Protocol)

Nicholson BD, Shinkins B, Pathiraja I, Roberts NW, James TJ, Mallett S, Perera R, Primrose JN, Mant D



**THE COCHRANE
COLLABORATION®**

This is a reprint of a Cochrane protocol, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2014, Issue 6

<http://www.thecochranelibrary.com>

WILEY

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
BACKGROUND	1
OBJECTIVES	4
METHODS	4
ACKNOWLEDGEMENTS	6
REFERENCES	7
APPENDICES	9
CONTRIBUTIONS OF AUTHORS	13
DECLARATIONS OF INTEREST	13
SOURCES OF SUPPORT	13

[Diagnostic Test Accuracy Protocol]

Blood CEA levels for detecting recurrent colorectal cancer

Brian D Nicholson¹, Bethany Shinkins¹, Indika Pathiraja¹, Nia W Roberts², Tim J James³, Susan Mallett¹, Rafael Perera¹, John N Primrose⁴, David Mant¹

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. ²Bodleian Health Care Libraries, University of Oxford, Oxford, UK. ³Clinical Biochemistry, Oxford University Hospitals NHS Trust, Oxford, UK. ⁴Department of Surgery, University of Southampton, Southampton, UK

Contact address: Brian D Nicholson, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. brian.nicholson@phc.ox.ac.uk.

Editorial group: Cochrane Colorectal Cancer Group.

Publication status and date: New, published in Issue 6, 2014.

Citation: Nicholson BD, Shinkins B, Pathiraja I, Roberts NW, James TJ, Mallett S, Perera R, Primrose JN, Mant D. Blood CEA levels for detecting recurrent colorectal cancer. *Cochrane Database of Systematic Reviews* 2014, Issue 6. Art. No.: CD011134. DOI: 10.1002/14651858.CD011134.

Copyright © 2014 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

To determine the accuracy of single-measurement blood CEA as a triage test to prompt further investigation for CRC recurrence after curative resection.

To identify sources of between- and within-study heterogeneity to inform potential subsequent analyses.

BACKGROUND

International guidelines recommend that blood carcino-embryonic antigen (CEA) levels are measured to detect recurrent colorectal cancer (CRC) as part of an intensive follow-up regimen (Duffy 2013a; Labianca 2010; Locker 2006; NCCN 2013; NICE 2011).

These guidelines are derived from non-randomised studies, and later randomised controlled trials (RCTs) investigating the optimal follow-up strategy following curative CRC resection. Follow-up strategies have been broadly classed as intensive and minimal, but the investigative modalities included in each strategy have varied greatly, with similarities in the composition of intensive and minimal regimens between studies. Jeffery's 2007 Cochrane review (Jeffery 2007) of eight RCTs (Kjeldsen 1997; Makela 1995; Ohlsson 1995; Pietra 1998; Rodriguez 2006; Schoemaker 1998;

Secco 2002; Wattoo 2006) showed that when compared to minimal follow-up, an intensive regimen reduces five-year all-cause mortality (odds ratio (OR) 0.73 95% confidence interval (CI) 0.59 to 0.91). As direct comparison of CEA measurement versus no CEA measurement was only possible for recurrence rates using two of these RCTs (OR 0.85 95% CI 0.58 to 1.25) and data on overall mortality were only available from one trial (OR 0.57 95% CI 0.26 to 1.29), the results were not conclusive due to the small numbers. The aim of intensive follow-up is to detect asymptomatic recurrences more amenable to a resection with clear margins. Jeffery's review demonstrated no significant difference in the recurrence rate between investigative strategies, but significantly more curative surgical procedures were conducted for recurrence in the intensive group (OR 2.41 95% CI 1.63 to 3.54) (Jeffery 2007).

The between-study heterogeneity, and the fact that many of the RCTs included in the Cochrane review predate modern approaches to cancer care lead many to caution when applying the meta-analysis to modern-day practice. For others, the advances in chemotherapy, hepatic resection, and multidisciplinary CRC follow-up has led to assertions that the clinical benefits of intensive follow-up will be even greater today (Labianca 2010). Published in 2014, the FACS pragmatic factorial RCT followed 1202 participants from 39 NHS hospitals reporting that those followed up with an intensive regimen had three times the odds of detecting a recurrence amenable to curative resection, that monitoring with CEA combined with a single computed tomography (CT) scan at 12 to 18 months was equally as effective as undertaking regular CT scanning, and that concurrent CEA and CT does not improve accuracy (Primrose 2014).

The absence of a difference in cancer-specific mortality between follow-up approaches has led to suggestions that the psychological support gained from regular medical follow-up, and the associated modifications of diet, lifestyle, and chronic disease management account for the improvement seen in all-cause mortality (Tjandra 2007).

Whilst the optimal combination and frequency of clinic visits, blood tests (including CEA), endoscopy and imaging included in an intensive follow-up regimen remains unclear (Scheer 2009), there is evidence that most recurrences will occur in the first 30 months after primary tumour resection, with almost all occurring within the first five years (Guthrie 2002), that CEA measurement is the most sensitive modality for detecting early recurrent disease (especially liver metastasis) (Duffy 2013a; Tsikitis 2009), that there are an increasing number of well-tolerated effective chemotherapy regimens for recurrent CRC in older populations (Cunningham 2010; Locker 2006), and that primary care follow-up results in similar outcomes to surgical outpatient follow-up (Wattchow 2006). Economic analyses have shown intensive follow-up to be cost-effective (Renehan 2004) and that CEA is the most cost-effective way of detecting recurrent CRC in primary care detecting (Primrose 2014).

There is no consensus on the interpretation of blood CEA results, with substantial variability in clinical practice.

Target condition being diagnosed

Colorectal cancer is globally the third most common cancer accounting for 9.8% of all detected cancers. In 2008, the age-standardised incidence rate was 17.3 cases per 100,000, 30.1 in more developed regions, and 10.7 in less developed regions (Ferlay 2013).

Colorectal adenocarcinoma arises in the colonic mucosa and progressively invades through the layers of bowel wall into surrounding structures leading to peritoneal, neural, lymphatic and haema-

tological metastasis (Gore 1997). This process provides the basis of the internationally recognised TNM staging system (Sobin 2009) and earlier Dukes classification (Dukes 1932). The first site of haematological metastasis is the liver via the portal vein, after which distant metastasis occurs most commonly to the lungs but also the bones and brain (Guthrie 2002). Prognosis is closely related to stage, with higher grade more invasive metastatic tumours having poorer prognosis (Maringe 2013). Approximately two-thirds of patients will present with a primary CRC amenable to radical surgery (+/- adjuvant therapy) (Jeffery 2007).

Following surgery (+/- adjuvant therapy) however, 30% to 50% of patients will develop recurrence (Labianca 2010). The most common site for recurrence is the liver followed by the lungs but can also occur in the abdomen and pelvis (Cunningham 2010; Jeffery 2007). Patients undergoing secondary surgery with curative intent have substantially improved five- and 10-year survival rates with a median survival time of 35.8 to 84.8 months. Surgery for isolated hepatic metastasis improves five-year survival by 36% to 58%, for isolated lung metastasis by 27% to 41%; chemotherapy can prolong life by one to two years, and improve quality of life (Arriola 2006; Cunningham 2010; Tsikitis 2009).

Index test(s)

CEA is a relatively simple and low-cost biomarker that can be detected by a blood test. The analysis of CEA in clinical studies utilises the technique of immunoassay in a variety of forms and from a number of different manufacturers. Earlier methods were manual immunoassays such as radio-immunoassay but most laboratories now utilise fully automated non-isotopic methods. The reproducibility of these fully automated methods are, in general, superior to the older manual methods. Unfortunately, the details of the methods used in clinical studies and their analytical performance is often lacking in publications (Wild 2013).

Data from external quality assessment schemes have repeatedly shown good precision for most methods at low CEA concentrations. In 2010 the mean within laboratory precision over a 12-month period at a concentration of 3 µg/L (equivalent to 54 U/L) was < 9% for all major methods. A greater analytical challenge is the difference in method bias (Wild 2013). Despite the availability of an international reference preparation (IRP 73/601) since 1975 and its widespread use in commercial assays since the early 1990's, method bias may be +/-20% and the degree of this bias may be sample-dependent (Bormer 1991; Laurence 1975). CEA has a complex molecular structure and the antibodies used in the immunoassays recognise different epitopes of the molecule and this is considered to be a major source of the method bias (Bormer 1991). Consequently, the interpretation of data from clinical studies, in particular the use of any particular threshold, whether that is 3, 5 or 7 µg/L, needs to consider the actual method utilised. Due to the good reproducibility but significant method-dependen-

dent bias, it is advised that the same assay technique should be used throughout any follow-up period (Duffy 2013a).

CEA is a glycoprotein involved in cell adhesion produced during foetal development. Production usually ceases at birth, but elevated levels can be detected in colorectal, breast, lung and pancreatic cancer, in smokers, and in benign conditions such as cirrhosis of the liver, jaundice, diabetes, pancreatitis, chronic renal failure, colitis, diverticulitis, irritable bowel syndrome, pleurisy and pneumonia (Newton 2011; Sturgeon 2009). It is produced in 90% of CRC and is known to contribute to the malignant characteristics of the tumour, and to have an important role in CRC metastasis (Dallas 2012). CEA levels may rise four and a half to eight months prior to the development of cancer-related symptoms (Goldstein 2005). Depending on the threshold used, the sensitivity of CEA varies depending on the stage of disease; Dukes type A, B, C, and D tumours are reported to be associated with levels of CEA > 5µg/L in 3%, 25% 45% and 65% of cases respectively (Sturgeon 2009). Furthermore, CEA is most sensitive for hepatic and retroperitoneal metastases and least sensitive for local recurrences and peritoneal or pulmonary disease (Scheer 2009).

Because of the variable sensitivity and expression in benign conditions CEA (and all other existing serum biomarkers), fails to meet gold standard criteria for biomarker use and so CEA is not recommended for screening purposes in the general population (Newton 2011). However, it is recommended for use as a preoperative prognostic marker, as a marker of response to chemotherapy (especially for liver metastasis), and as a triage test for diagnosing recurrent CRC (where a rise should lead to further investigation rather than initiation of therapy) (Duffy 2013a; Sturgeon 2009). Although serial CEA measurements are taken, centres take action on a single CEA level above an absolute threshold level, but there is lack of agreement on the threshold above which action should be taken, or the extent of concentration change that constitutes a clinically significant rise. A threshold between 3 and 7 µg/L is commonly used, some centres look at the difference a pre-operative or post-operative baseline level, some repeat the test before acting, and no guidelines recommend taking into account trend information based on longitudinal CEA measurements.

The most recent meta-analysis includes 20 studies combining diagnostic accuracy data for a range of threshold values (3 to 15 µg/L) measured by a variety of test-kits to investigate the diagnostic value of the absolute level from a single test (Tan 2009). The pooled estimates of diagnostic accuracy were: sensitivity 64% (95% CI 61% to 67%); specificity 90% (89% to 91%); diagnostic OR 18 (12 to 28); and area under the curve (AUC) 0.79 (standard error 0.054). There was a significant degree of heterogeneity reported between studies (Q-value 80.83, P < 0.001) (Tan 2009). When limited to four studies using the 3 µg/L threshold, sensitivity (73% (69 to 77)) increased at the expense of specificity (68% (65 to 72)). Through meta-regression the authors suggest that a cut-off of 2.2µg/L provides the ideal balance between sensitivity and specificity for use in clinical practice (Tan 2009), but this level

generates a high level of false-alarms and is implemented by few clinicians: the COST trial used a CEA cut-off of 5µg/L (Tsikitis 2009), and the FACS trial used a threshold of 7µg/L (Primrose 2014). We propose an update of the Tan study (the search was performed in July 2007) using a less conservative search strategy and conducting analyses following the latest Cochrane DTA guidance.

Clinical pathway

Following radical surgery (+/- adjuvant therapy), there is variation in the recommended intensive follow-up regimen (Duffy 2013a; Labianca 2010; Locker 2006; NCCN 2013; NICE 2011).

The European Society of Medical Oncology (ESMO) recommend history, physical examination, and CEA determination every three to six months for three years, and every six to 12 months at years four and five, colonoscopy at one year then at every three to five years looking for metachronous adenomas and cancers, a CT scan of the chest and contrast-enhanced ultrasound scan (USS) or CT scan of the abdomen every six to 12 months for the first three years in patients considered higher risk, advising against the use of other laboratory and radiological examinations unless patients have suspicious symptoms (Labianca 2010).

The American Society of Clinical Oncology (ASCO) recommends that CEA is performed every three months for three years in patients with stage II or III disease if the patient is a candidate for surgery or systemic therapy, and that raised CEA levels (> 5µg/L, confirmed by a repeat test) warrants further evaluation for metastatic disease (Locker 2006). Unlike ASCO, ESMO does not specify a threshold nor limit testing to tumour stage but the European Group on Tumour Markers (EGTM) specify CEA measurement at baseline and then every two to three months for three years, then six-monthly for five years in patients with stage II-III disease who would tolerate further surgery or systemic therapy. EGTM recommend that any increase in CEA (confirmed by a repeat test) should trigger further investigations (Duffy 2013a).

NICE recommend follow-up from four to six weeks following curative treatment, for all patients who could tolerate and accept the balance of risk and benefits of further treatment, including CEA measurement at least every six months in the first three years, two CT scans of the chest and abdomen in the first three years, and colonoscopy at one year and five years (NICE 2011).

Once recurrence is suspected patients then undergo further diagnostic testing with usually CT or USS to confirm recurrence (Duffy 2013), although the modality used to confirm recurrence varies and can alternatively be clinical assessment, colonoscopy, flexible sigmoidoscopy and barium enema, CT colonography, positron emission tomography-computed tomography (PET-CT), or magnetic resonance imaging (MRI).

Prior test(s)

As detailed above, CEA is often the first investigative modality to be used within an intensive follow-up regimen.

Role of index test(s)

As a triage test to prompt further investigation for CRC recurrence.

Alternative test(s)

Circulating tumour cells and cytokeratins have been examined as possible biomarkers of CRC recurrence but the studies are few and limited. Ca125 is regarded as an emerging biomarker for use in postoperative follow-up but as yet evidence is limited (Duffy 2013a; Newton 2011). CT imaging is the only other test that meta-analysis suggests has potential to detect metastatic recurrence amenable to resection but CT is less cost-effective than CEA. FACS suggested that concurrent CEA and CT does not improve accuracy (Primrose 2014).

Rationale

CEA alone is potentially the most cost-effective option to detect CRC recurrence. This DTA review aims to clarify the accuracy of single-measurement blood CEA as a triage test for CRC recurrence. We propose this in the knowledge that serial CEA measures are commonly taken as part of a postoperative monitoring schedules, but have chosen to evaluate the diagnostic accuracy of a single CEA measurement because the clinical decision to investigate further for recurrent CRC is most often based on a single measurement alone.

OBJECTIVES

To determine the accuracy of single-measurement blood CEA as a triage test to prompt further investigation for CRC recurrence after curative resection.

Secondary objectives

To identify sources of between- and within-study heterogeneity to inform potential subsequent analyses.

METHODS

Criteria for considering studies for this review

Types of studies

We will include cross sectional diagnostic test accuracy studies, cohort studies, and RCTs in the setting of follow-up after CRC resection using direct comparisons between CEA and the reference standard from which data for a 2 x 2 table can be extracted. We will exclude case-control studies as they are inherently biased in this setting.

Participants

We will limit the review to studies of blood CEA measurement to detect recurrence of colorectal cancer in adults with no detectable residual disease after primary treatment with surgical resection (+/- adjuvant therapy).

Index tests

Blood carcino-embryonic antigen (CEA). Clinical guidance and Tan 2009 demonstrate that the CEA thresholds most commonly used in clinical practice are 3, 5, 7, and 10µg/L (Locker 2006; Tan 2009). We will primarily investigate the accuracy of these thresholds, using subgroup analysis to investigate the accuracy of others.

Target conditions

Recurrence of CRC following curative resection, including locoregional recurrence, and metastatic disease.

Reference standards

1. Imaging done per protocol or to investigate for suspected recurrence (usually CT, MRI or PET-CT, but also endoscopy, CT colonography, ultrasound, and barium enema).
2. The histological confirmation of recurrence following surgery or tissue biopsy.
3. Routine clinical follow-up used as a reference standard to confirm negative index test values where imaging is not indicated as part of the follow-up schedule (standard protocols run for three to five years).

For those studies where it is possible to identify the individual reference standard, these data will be used in sensitivity analyses. However, we anticipate that studies will report on a composite reference standard by following a prespecified clinical pathway as described above, and so the reference standard may vary between patients in the same study. Without individual patient data, identifying the exact investigative modality may not be possible as summary diagnostic accuracy data will be presented.

We will therefore determine whether the chosen reference standard (or composite reference standard) is “appropriate” (1 to 3 above), “inappropriate” (a reference standard not included in 1 to 3 above), or “not stated” for further subgroup analysis.

If the data are available, deaths during follow-up will be recorded as “death from CRC”, “death with CRC”, “death from other causes”, or “death unspecified”.

Search methods for identification of studies

Electronic searches

Our information specialist NR (trained in Cochrane DTA methodology) designed our search strategy.

We will search for related reviews in MEDIONdatabase (<http://www.mediondatabase.nl>), the DARE database (*The Cochrane Library*, Wiley), MEDLINE (OvidSP)[1946-current, In-process] and Embase (OvidSP)[1974-current] using the Reviews Clinical Query. Primary studies (including conference abstracts) will be searched for in MEDLINE (OvidSP)[1946-current, In-process], Embase (OvidSP)[1974-current], Cochrane Central Register of Controlled Trials (*The Cochrane Library*, Wiley) and the Science Citation Index & Conference Proceedings Citation Index - Science (Web of Science, Thomson)[1945-current]. Ongoing studies will be identified by searching WHO ICTRP (<http://apps.who.int/trialsearch/>) and ClinicalTrials (<http://clinicaltrials.gov>). An additional search for conference abstracts will be conducted on the ASCO meeting library (<http://meetinglibrary.asco.org/>)

No language limits will be applied to the search, non-English manuscripts will be translated to assess suitability for inclusion.

An example search strategy for use in MEDLINE was piloted in December 2012 and is shown in [Appendix 1](#). This strategy will be translated for the other databases using the appropriate controlled vocabulary and free-text terms.

Searching other resources

Following the search of bibliographic databases, we will browse reference lists of retrieved reviews and all included studies. In addition, we will perform a ‘Related articles’ search on PubMed on all included studies. The principal investigators of included studies will be contacted to identify further relevant literature, clarify methodological queries if they exist and ask for any unpublished data relevant to this review. The details of searches for conference abstracts and ongoing studies are detailed above.

Data collection and analysis

Selection of studies

To identify relevant studies one review author will scan all titles and exclude those studies clearly not relevant to the topic of CEA in detection of CRC recurrence. Secondly, two review authors will

independently assess the remaining titles and abstracts retrieving the full-text of relevant articles and of those for which a decision cannot be made on the basis of title and abstract alone. A third review author will resolve any disputes over which references should be included. All full-text articles will be first scrutinised for data sufficient to allow population of 2 x 2 table. Following confirmation of 2 x 2 data, full data will be extracted. Reasons for any exclusions will be detailed in a flow diagram.

Data extraction and management

Full data extraction will be guided by a background information sheet describing how each item should be interpreted. This form will be piloted by two review authors using three initial studies and refined if necessary. Any disagreement over extracted data will be discussed and if consensus is not achieved moderated by a third author.

Data will be extracted or calculated from data provided and collated in an Excel spreadsheet under the following headings: Author, year, country, population (n), included participants (n), setting of follow-up, age, smoking status, stage/grade of primary tumour, primary treatment received, investigations done to ensure no residual disease, definition of follow-up schedule, CEA threshold, timing of CEA measurement, reference standard, site of recurrence, timing of CEA versus reference standard, cases of recurrence (n), True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Positive Likelihood Ratio (PLR), Negative Likelihood Ratio (NLR), AUC, QUADAS-2 items (including CEA laboratory technique).

If data are not available we will first contact the authors to request the missing data.

Assessment of methodological quality

Assessment of methodological quality

QUADAS-2 is a generic set of criteria for assessing diagnostic accuracy studies consisting of four key domains: patient selection, index test, reference standard, and the flow of patients through the study and timing of the index test in relation to the reference standard. Signalling questions are included to allow judgement of the risk of bias across the four domains ([Whiting 2011](#)).

We have modified QUADAS-2 excluding items not applicable to this review ([Whiting 2011](#)). A guide to the operational definitions for the modified QUADAS-2 items can be found in [Appendix 2](#). We have included additional questions regarding CEA laboratory technique (2.A.2-2.A.4) to detail variability in laboratory technique over time, and index test repetition (4.A.1) as some centres repeat CEA before conducting the reference standard, which conflicts with our objective to assess singular CEA measurements. We have modified “Was there an appropriate interval between index test(s) and reference standard?” (Yes/No/Unclear) to instead

read “4.A.2. Was the timing between index test(s) and reference standard ascertainable?” (Yes / Unclear). We have modified “Did all patients receive a reference standard?” to instead read “Did all included patients who had at least one CEA measurement receive a reference standard?” to make specific to our study. We have removed “was a case-control design avoided?” from the original QUADAS-2 template as we are excluding all case-control studies, and we have removed “Were the index test results interpreted without knowledge of the results of the reference standard?” as CEA is an objective test (Whiting 2011).

Two review authors will assess quality of all articles independently, will discuss any disagreements, and if consensus cannot be agreed a third author will act as a moderator to reach consensus. The results of the quality assessment will be used for descriptive purposes to provide an evaluation of the overall quality of the included studies and to investigate potential sources of heterogeneity.

Statistical analysis and data synthesis

Descriptive statistics will be used to present a summary of each included study. Tables will detail the patient sample, study design, CEA technique, follow-up characteristics and the CEA threshold(s) at which accuracy was reported. Binary diagnostic accuracy data will be extracted from all included studies as 2 x 2 tables. The low, high or unclear risk of bias for each of the four domains of the QUADAS-2 assessment will be presented graphically as described by Whiting 2011.

Inferential statistics will be guided by Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). RevMan 5 will be used to produce forest plots showing the variability of sensitivity and specificity across primary studies, with corresponding 95% confidence intervals, for visual comparison. We will present forest plots per CEA threshold used and per CEA technique used (Macaskill 2010).

To overcome the methodological limitations of the Moses-Littenberg SROC approach, we will conduct a bivariate meta-analysis. We will be using the `xtnlogit` command in Stata, which will give us the flexibility to explore the influence of covariates (Takwoingi 2013).

A key objective of this meta-analysis is to explore possible CEA thresholds to identify CRC recurrence, but we anticipate many studies will not present data on multiple thresholds. If sufficient data are available we will use the multivariate random-effects meta-analysis approach described by Hamza 2009.

Investigations of heterogeneity

To investigate heterogeneity, we will use a meta-regression approach to add covariates to the Rutter and Gatsonis HSROC mod-

els (Leefflang 2008). For continuous variables, we will categorise data into clinically meaningful groups if a linear association is not found, or if feasible we will add polynomial terms to the model (Royston 2008). We will use likelihood ratio tests to determine statistical significance.

We will investigate heterogeneity as far as is possible from the primary study data, and as a minimum attempt analysis using study date (data gathering pre-1995, post-1995), laboratory method, and reference standard used.

We will contact study authors to request supplementary data if they are not documented within selected publications.

Sensitivity analyses

For sensitivity analysis, we will remove each study from the pooled estimate and recalculate; we will remove RCT data and recalculate; we will examine the effect of reference standard by removing studies in turn that have used “appropriate”/“inappropriate”/“not stated” reference standards, or have used specific tests (e.g. composite/CT/MRI/colonoscopy); and we will use the QUADAS-2 assessment to restrict our analyses to only high-quality studies (defined as overall low risk of bias, or low concern over applicability) for each QUADAS-2 domain.

If sufficient data are available regarding laboratory technique, we will assess whether laboratory technique influences the results by excluding studies without sufficient data (those that score “no” or “unclear” on any QUADAS-2 items 2.A.2-4), and analysing per laboratory technique used.

Assessment of reporting bias

As described by van Roon 2011, investigation of publication bias in DTA studies is known to be problematic because many studies are undertaken without ethical approval or study registration and so it is not possible to identify all studies until final publication. Funnel plots used to detect publication bias in reviews of RCTs have also been shown to be misleading for DTA reviews (Deeks 2005; Leefflang 2008; Song 2002). Assessment of reporting bias will therefore not be included in this review.

ACKNOWLEDGEMENTS

The review authors would like to thank Professor Paul Glasziou for his input especially into the development of the modified QUADAS-2 assessment tool, and Dr Clare Davenport for her input into the development of the TRF.

REFERENCES

Additional references

Arriola 2006

Arriola E, Navarro M, Pares D, Munoz M, Pareja L, Figueras J, et al. Imaging techniques contribute to increased surgical rescue of relapse in the follow-up of colorectal cancer. *Diseases of Colon and Rectum* 2006;**49**(4):478–84.

Borner 1991

Borner OP. Standardization, specificity, and diagnostic sensitivity of four immunoassays for carcinoembryonic antigen. *Clinical Chemistry* 1991;**37**(2):231–6.

Cunningham 2010

Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal Cancer. *Lancet* 2010;**375**: 1030–47.

Dallas 2012

Dallas MR, Liu G, Chen W, Thomas SN, Wirtz D, Huso DL, et al. Divergent roles of CD44 and carcinoembryonic antigen in colon cancer metastasis. *The FASEB journal* 2012;**26**:2648–56.

Deeks 2005

Deeks JJ, Mackaskil P, Irwig I. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**:882–93.

Duffy 2013

Duffy MJ. Tumour markers in clinical practice: a review focusing on common solid cancers. *Medical Principles and Practice* 2013;**22**:4–11.

Duffy 2013a

Duffy MJ, Lamerz R, Haglund C, Nicolini A, Kalousova M, Holubec L, et al. Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European Group on Tumour Markers (EGTM) 2014 Guidelines Update. *International Journal of Cancer* 2013;**134**(11): 2513–22. [DOI: 10.1002/ijc.28384]

Dukes 1932

Dukes CE. The classification of cancer of the rectum. *Journal of Pathological Bacteriology* 1932;**35**:323–32.

Ferlay 2013

Ferlay I, Soerjomataram, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide. IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr>, accessed on 31 March 2014.

Goldstein 2005

Goldstein MJ, Mitchell EP. Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. *Cancer Investigations* 2005;**23**:339–51.

Gore 1997

Gore RM. Colorectal cancer. Clinical and pathologic features. *Radiological Clinics of North America* 1997;**35**: 403–29.

Guthrie 2002

Guthrie JA. Colorectal cancer: follow-up and detection of recurrence. *Abdominal Imaging* 2002;**27**:570–7.

Hamza 2009

Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology* 2009;**9**(73):1–15.

Jeffery 2007

Jeffery M, Hickey BE, Hilder PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database of Systematic Reviews* 2007, Issue 1. [DOI: 10.1002/14651858.CD002200.pub2]

Kjeldsen 1997

Kjeldsen BJ, Kronborg O, Fenger C, Jørgensen OD. A prospective randomised study of follow-up after radical surgery for colorectal cancer. *British Journal of Surgery* 1997;**84**:666–9.

Labianca 2010

Labianca R, Nordlinger B, Beretta GD, Brouquet A, Cervantes A. Primary colon cancer: ESMO Clinical Practice Guidelines for Diagnosis, adjuvant treatment and follow-up. *Annals of Oncology* 2010;**21**(Suppl 5):v70–v77.

Laurence 1975

Laurence DJ, Turberville C, Anderson SG, Neville AM. First British standard for carcinoembryonic antigen (CEA). *British Journal of Cancer* 1975;**32**(3):295–9.

Leefflang 2008

Leefflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic review of diagnostic test accuracy. *Annals of Internal Medicine* 2008;**149**:889–97.

Locker 2006

Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, MacDonald JS, et al. ASCO 2006 update of Recommendations for the use of tumour markers in gastrointestinal cancer. *Journal of Clinical Oncology* 2006;**24**(33):5313–27.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. *Cochrane Handbook for Systematic Reviews of Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010.

Makela 1995

Makela JT, Seppo OL, Kairaluoma MI. Five-year follow-up after radical surgery for colorectal cancer. *Archives of Surgery* 1995;**130**:1062–7.

Maringe 2013

Maringe C, Walters S, Rachet B, Butler J, Fields T, Finan P, et al. Stage at diagnosis and colorectal cancer survival in six high-income countries: A population-based study of patients diagnosed during 2000–2007. *Acta Oncologica* 2013;**52**:919–32.

- NCCN 2013**
NCCN 2013. NCCN Guidelines Version 3.2013 Colon Cancer. NCCN Clinical Practice Guidelines in Oncology 2013.
- Newton 2011**
Newton KF, Newman W, Hill J. Review of biomarkers in colorectal cancer. *Colorectal Disease* 2011;**14**:3–17.
- NICE 2011**
NICE clinical guideline. *Colorectal Cancer*. Vol. **NICE clinical guideline 131**, National Institute for Health and Clinical Excellence, 2011.
- Ohlsson 1995**
Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg K. Follow-up after curative surgery for colorectal carcinoma. *Diseases of Colon and Rectum* 1995;**38**:619–26.
- Pietra 1998**
Pietra N, Sarli L, Costi R, Ouchemi C, Grattarola M, Peracchia A. Role of follow-up in management of local recurrences of colorectal cancer. *Diseases of Colon and Rectum* 1998;**41**:1127–33.
- Primrose 2014**
Primrose JN, Perera R, Gray A, Rose P, Fuller A, Corkhill A, et al. FACS Trial Investigators. Effect of 3–5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: FACS randomised clinical trial. *JAMA* 2014;**311**(3):263–70.
- Renehan 2004**
Renehan AG, O'Dwyer AT, Whynes DK. Cost effectiveness analysis of intensive versus conventional follow up after curative resection. *BMJ* 2004;**328**:81.
- Rodriguez 2006**
Rodriguez-Moranta F, Salo J, Arcusa A, Boadas J, Pinol V, Bessa X, et al. Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: A prospective, multicenter, randomised, controlled trial. *Journal of Clinical Oncology* 2005;**24**:1–8.
- Royston 2008**
Royston P, Sauerbrei W. *Multivariable Model-building: A Pragmatic Approach Based on Fractional Polynomials for Continuous Variables*. Wiley-Blackwell, 2008.
- Scheer 2009**
Scheer A, Auer RAC. Surveillance after Curative Resection of Colorectal Cancer. *Clinics in Colon and Rectal Surgery* 2009;**22**(4):242–50.
- Schoemaker 1998**
Schoemaker D, Black R, Giles L, Toouli J. Yearly colonoscopy, liver CT, and chest radiography do not influence 5-year survival of colorectal patients. *Gastroenterology* 1998;**114**:7–14.
- Secco 2002**
Secco GB, Fardelli RM, Gianquinto D, Bonfante P, Baldi E, Ravera G, et al. Efficacy and cost of risk adapted follow-up in patients after colorectal surgery: a prospective, randomised and controlled trial. *European Journal of Surgical Oncology* 2002;**28**:418–23.
- Sobin 2009**
Sobin LH, Gospodarowicz, Wittekind C. *TNM Classification of Malignant Tumours*. 7. Wiley-Blackwell, 2009.
- Song 2002**
Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analysis of diagnostic accuracy. *International Journal of Epidemiology* 2002;**31**: 88–95.
- Sturgeon 2009**
Sturgeon CM, Lai LC, Duffy MJ. Serum tumour markers: how to order and interpret them. *BMJ* 2009;**339**:852–8.
- Takwoingi 2013**
Takwoingi Y. Meta-analysis of test accuracy studies in Stata: a bivariate model approach. Available from: <http://srdta.cochrane.org/> November 2013, issue Version 1.0.
- Tan 2009**
Tan E, Gouvas N, Nicholls RJ, Ziprin P, Xynos E, Tekkis PP. Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surgical Oncology* 2009;**18**:15–24.
- Tjandra 2007**
Tjandra JJ, Chan MKY. Follow-up after curative resection of colorectal cancer: a meta-analysis. *Diseases of the Colon and Rectum* 2007;**50**:1783–99.
- Tsikitis 2009**
Tsikitis VL, Malireddy K, Green EA, Christensen B, Whelan R, Hyder J, et al. Postoperative surveillance recommendations for early stage colon cancer based on results from the clinical outcomes of surgical therapy trial. *Journal of Clinical Oncology* 2009;**27**:3671–6.
- van Roon 2011**
van Roon AHC, van Dam L, Zauber AG, van Ballegooijen M, Borsboom GJJM, Steyerberg EW, et al. Guaiac-based faecal occult blood tests versus faecal immunochemical tests for colorectal cancer screening in average-risk individuals. *Cochrane Database of Systematic Reviews* 2011, Issue 8. [DOI: 10.1002/14651858.CD009276]
- Wattchow 2006**
Wattchow DA, Weller DP, Esterman A, Pilotto LS, McGorm K, Hammett Z, et al. General Practice vs surgical-based follow-up for patients with colon cancer: randomised controlled trial. *British Journal of Cancer* 2006;**94**:1116–21.
- Whiting 2011**
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.
- Wild 2013**
Wild D. *The Immunoassay Handbook: Theory and Applications of Ligand Binding, ELISA and Related Techniques*. 4th Edition. Elsevier, 31 Jan 2013.
- * Indicates the major publication for the study

APPENDICES

Appendix I. MEDLINE search strategy

1	colorectal neoplasms/ or exp adenomatous polyposis coli/ or exp colonic neoplasms/ or colorectal neoplasms, hereditary non-polyposis/ or exp rectal neoplasms/
2	(colorectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab
3	(colon* adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab
4	(bowel adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab
5	(rectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab
6	(rectum adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab
7	1 or 2 or 3 or 4 or 5 or 6
8	Carcinoembryonic Antigen/
9	cea.ti,ab.
10	(carcinoembryonic adj3 antigen?).ti,ab.
11	(carcinoembryonic adj3 antibod*).ti,ab.
12	(carcino-embryonic adj3 antigen?).ti,ab.
13	(carcino-embryonic adj3 antibod*).ti,ab.
14	8 or 9 or 10 or 11 or 12 or 13
15	Neoplasm Recurrence, Local/
16	Recurrence/
17	recur*.ti,ab.
18	relaps*.ti,ab.
19	treatment failure/
20	Reoperation/
21	Follow-Up Studies/ and Postoperative Care/

(Continued)

22	reoperat*.ti,ab.
23	((local or distant) adj2 failure).ti,ab.
24	((therap* or treatment or surg*) adj3 fail*).ti,ab.
25	((therap* or treatment or surg*) adj3 (respond* or response*)).ti,ab
26	((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) adj5 follow up).ti,ab
27	((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) adj5 surveil- lance).ti,ab
28	((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) adj5 monitor*) .ti,ab
29	15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
30	7 and 14 and 29

Appendix 2. Operational guidance for modified QUADAS-2 tool

Unless otherwise specified, each item must be explicitly reported to achieve a “yes” answer.

DOMAIN 1: Patient Selection		
<i>A: Risk of Bias</i>		
1. Was a consecutive or random sample of patients enrolled? Yes/No/Unclear		
2. Did the study avoid inappropriate exclusions?		
	Yes	Patients are included in follow-up post radical CRC resection, OR Exclusions was justified in the text and reviewers reached consensus on the appropriateness of any exclusions. Exclusions based on patient characteristics allowing subgroup analysis (e.g. tumour grade) should be deemed appropriate
	No	Criteria for “yes” not achieved.
	Unclear	Exclusions not reported clearly.
OVERALL RISK OF BIAS: LOW/HIGH/UNCLEAR		
<i>B: Applicability</i>		

(Continued)

1. Is there concern that the included patients do not match the review question?		
	Yes	Patients are not undergoing follow-up post radical CRC resection including CEA measurement
	No	Patients are undergoing follow-up post radical CRC resection including CEA measurement
	Unclear	The included population is not defined.
OVERALL CONCERN REGARDING APPLICABILITY: LOW/HIGH/UNCLEAR		
DOMAIN 2: Index Tests		
<i>A: Risk of Bias</i>		
1. If a threshold was used, was it pre-specified? Yes/No/Unclear		
2. Is the same method and instrument used for all CEA measurements? Yes/No/Unclear		
3. Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations? Yes/No/Unclear		
4. Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme? Yes/No/Unclear		
OVERALL RISK OF BIAS: LOW/HIGH/UNCLEAR		
<i>B: Applicability</i>		
1. Is there concern that the index test, its conduct, or interpretation differ from the review question?		
	Yes	Blood CEA is not interpreted as a stand-alone test to trigger investigation for CRC recurrence
	No	Blood CEA is interpreted as a stand-alone test to trigger investigation for CRC recurrence
	Unclear	It is unclear whether the index test differs from the review question
OVERALL CONCERN REGARDING APPLICABILITY: LOW/HIGH/UNCLEAR		
DOMAIN 3: Reference Standard		
<i>A: Risk of Bias</i>		

(Continued)

1. Is the reference standard likely to correctly classify the target condition? - can we confidently exclude recurrence on the basis of no clinical detection of recurrence when we are assessing the utility of CEA at detecting asymptomatic recurrence amenable to resection?		
	Yes	An appropriate reference standard (as defined in the protocol) is used
	No	An inappropriate reference standard is used
	Unclear	The reference standard used is not clearly specified.
2. Were the reference standard results interpreted without knowledge of the results of the index test? - If tests are done as part of a follow-up regime it must not be assumed that the interpretation of each test is independent of another. It must be clearly stated when reference test interpretation occurred		
	Yes	The reference standard results were interpreted without knowledge of the index test(s)
	No	The reference standard results were interpreted with knowledge of the index test(s)
	Unclear	It is not clear whether interpretation was blinded or not.
OVERALL RISK OF BIAS: LOW/HIGH/UNCLEAR		
<i>B: Applicability</i>		
1. Is there concern that the target condition as defined by the reference standard does not match the review question? Yes/No/Unclear		
OVERALL CONCERN REGARDING APPLICABILITY: LOW/HIGH/UNCLEAR		
DOMAIN 4: Flow and Timing		
<i>A: Risk of Bias</i>		
1. Was the index test repeated prior to the reference standard? Yes/No/Unclear		
2. Was the the timing between index test(s) and reference standard ascertainable?		
	Yes	The timing was ascertainable.
	Unclear	Not reported, variable or could not be clearly determined
3. Did all included patients who had at least one CEA measurement receive a reference standard? Yes/No/Unclear		
4. Did patients receive the same reference standard?		

(Continued)

	Yes	>95% of patients received the same reference standard regardless of index test results or place within a follow-up schedule
	No	>95% of patients did not receive the same reference standard regardless of index test results, or place within the follow-up schedule
	Unclear	It is unclear whether all the included patients received same reference standard regardless of index test results
5. Were all patients included in the analysis? Yes/No/Unclear		
OVERALL RISK OF BIAS: LOW/HIGH/UNCLEAR		

CONTRIBUTIONS OF AUTHORS

BDN wrote the initial draft of the protocol. NWR and BDN devised the search strategy. DM, TJJ, SM, BS, IP, JP, and RP provided comments during protocol development.

DECLARATIONS OF INTEREST

There are no potential conflicts of interest relating to this review.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- HTA - 11/136/81, UK.

This work is partly funded by the National Institute of Health Research (NIHR) Health Technology Appraisal Programme project grant “What CEA level should trigger further investigation during follow up after curative treatment for colorectal cancer?” (HTA - 11/136/81).

- National Institute for Health Research (NIHR) School for Primary Care Research (SPCR), UK.

The Nuffield Department of Primary Care Health Sciences receives funding from the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR).