Cochrane Library

Cochrane Database of Systematic Reviews

Blood CEA levels for detecting recurrent colorectal cancer (Review)

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

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 2015, Issue 12. Art. No.: CD011134.
DOI: 10.1002/14651858.CD011134.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Blood CEA levels for detecting recurrent colorectal cancer

Brian D Nicholson[1], Bethany Shinkins[2], Indika Pathiraja[1], Nia W Roberts[3], Tim J James[4], Susan Mallett[5], Rafael Perera[1], John N Primrose[6], David Mant[1]

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. [2]Academic Unit of Health Economics, University of Leeds, Leeds, UK. [3]Bodleian Health Care Libraries, University of Oxford, Oxford, UK. [4]Clinical Biochemistry, Oxford University Hospitals NHS Trust, Oxford, UK. [5]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK. [6]Department 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 12, 2015.
Review content assessed as up-to-date: 29 January 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* 2015, Issue 12. Art. No.: CD011134. DOI: 10.1002/14651858.CD011134.pub2.

ABSTRACT

Background

Testing for carcino-embryonic antigen (CEA) in the blood is a recommended part of follow-up to detect recurrence of colorectal cancer following primary curative treatment. There is substantial clinical variation in the cut-off level applied to trigger further investigation.

Objectives

To determine the diagnostic performance of different blood CEA levels in identifying people with colorectal cancer recurrence in order to inform clinical practice.

Search methods

We conducted all searches to January 29 2014. We applied no language limits to the searches, and translated non-English manuscripts. We searched for relevant reviews in the MEDLINE, EMBASE, MEDION and DARE databases. We searched for primary studies (including conference abstracts) in the Cochrane Central Register of Controlled Trials (CENTRAL), in MEDLINE, EMBASE, and the Science Citation Index & Conference Proceedings Citation Index - Science. We identified ongoing studies by searching WHO ICTRP and the ASCO meeting library.

Selection criteria

We included cross-sectional diagnostic test accuracy studies, cohort studies, and randomised controlled trials (RCTs) of post-resection colorectal cancer follow-up that compared CEA to a reference standard. We included studies only if we could extract 2 x 2 accuracy data. We excluded case-control studies, as the ratio of cases to controls is determined by the study design, making the data unsuitable for assessing test accuracy.

Data collection and analysis

Two review authors (BDN, IP) assessed the quality of all articles independently, discussing any disagreements. Where we could not reach consensus, a third author (BS) acted as moderator. We assessed methodological quality against QUADAS-2 criteria. We extracted binary diagnostic accuracy data from all included studies as 2 x 2 tables. We conducted a bivariate meta-analysis. We used the xtmelogit command in Stata to produce the pooled estimates of sensitivity and specificity and we also produced hierarchical summary ROC plots.

Main results

In the 52 included studies, sensitivity ranged from 41% to 97% and specificity from 52% to 100%. In the seven studies reporting the impact of applying a threshold of 2.5 µg/L, pooled sensitivity was 82% (95% confidence interval (CI) 78% to 86%) and pooled specificity 80% (95% CI 59% to 92%). In the 23 studies reporting the impact of applying a threshold of 5 µg/L, pooled sensitivity was 71% (95% CI 64% to 76%) and pooled specificity 88% (95% CI 84% to 92%). In the seven studies reporting the impact of applying a threshold of 10 µg/L, pooled sensitivity was 68% (95% CI 53% to 79%) and pooled specificity 97% (95% CI 90% to 99%).

Authors' conclusions

CEA is insufficiently sensitive to be used alone, even with a low threshold. It is therefore essential to augment CEA monitoring with another diagnostic modality in order to avoid missed cases. Trying to improve sensitivity by adopting a low threshold is a poor strategy because of the high numbers of false alarms generated. We therefore recommend monitoring for colorectal cancer recurrence with more than one diagnostic modality but applying the highest CEA cut-off assessed (10 µg/L).

PLAIN LANGUAGE SUMMARY

Detecting recurrent colorectal cancer by testing for blood carcino-embryonic antigen (CEA).

Background

After surgery for cancer in the colon or rectum (colorectal cancer), most people are intensively followed up for at least five years to monitor for signs of the cancer returning. When this occurs, it usually causes a rise in a blood protein called CEA (carcino-embryonic antigen). An increased level of CEA can be picked up by a blood test, which is normally done every three to six months after colorectal cancer surgery. Those people with raised CEA levels are further investigated by x-ray imaging (usually a scan of the chest, abdomen and pelvis). We conducted this review to help decide what level of blood CEA should lead to further investigation.

Key Results

This review shows that setting a low cut-off point will increase the number of genuine cases of colorectal cancer recurrence that are detected (true positives), but a low cut-off will also cause unnecessary alarm by incorrectly classifying too many cases that are not actually recurrences (false positives). In addition, this review shows that a rise in CEA does not occur in up to 20% of patients with a true recurrence (false negatives). The current evidence supports using the highest cut-off point assessed (10 µg/L), but that adding another diagnostic modality (e.g. a single scan of the chest, abdomen and pelvis at 12 to 18 months) is necessary in order to avoid the missed cases.

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 2013b; Labianca 2010; Locker 2006; NCCN 2013; NICE 2011).

A previous Cochrane review (Jeffery 2007) of eight randomised controlled trials (RCTs) (Kjeldsen 1997; Makela 1995; Ohlsson 1995; Pietra 1998; Rodriguez-Moranta 2006b; Schoemaker 1998; Secco 2002; Wattchow 2006) evaluated the impact of follow-up strategy on overall survival and the number of recurrences detected. The analysis included very scant data on CEA; data on overall survival were only available from one trial (odds ratio (OR) 0.57, 95% confidence interval (CI) 0.26 to 1.29) and data on recurrence rate only from two (OR 0.85, 95% CI 0.58 to 1.25).The follow-up strategies implemented in each study were instead broadly classed as either intensive or minimal and the investigative modalities included in each strategy varied greatly between studies. Compared to minimal follow-up, it was estimated that an intensive regimen could significantly reduce five-year all-cause mortality (OR 0.73, 95% CI 0.59 to 0.91).

The validity of this conclusion has been questioned because the mechanism by which a mortality reduction of this magnitude

could be achieved by treating asymptomatic recurrence is unclear. There is evidence from one trial that starting chemotherapy for recurrence at an asymptomatic rather than symptomatic stage increases length of survival by a median of five months (Glimelius 1992). There is also observational evidence that surgical resection of metastases when feasible is associated with over 40% survival at five years (Colibaseanu 2013; Gonzalez 2013; Kanas 2012), and one commentator has suggested that advances in chemotherapy, hepatic resection, and multidisciplinary CRC follow-up mean that the clinical benefits of intensive follow-up will be even greater today (Labianca 2010). It is certainly true that there are now a number of well-tolerated effective chemotherapy regimens for recurrent CRC in older populations (Cunningham 2010; Locker 2006). However, the authors of the CEASL (CEA second-look) trial argue that identifying and treating asymptomatic recurrence has the potential to increase overall mortality (Treasure 2014), and the FACS (Follow-up After Colorectal Surgery) trial suggests that the effect of follow-up on absolute mortality is much smaller than that suggested by the 2007 review (Primrose 2014).

Nevertheless, the FACS trial has re-awakened interest in CEA follow-up. It showed that measuring blood CEA three- to six-monthly for five years, augmented by a single CT (computed tomography) scan at 12 to 18 months, leads to earlier diagnosis of recurrence and increases by about three-fold the proportion of recurrences that can be treated with curative intent (Primrose 2014). As CEA monitoring does not involve x-rays, it can be done in the community, and is potentially more cost-effective than CT imaging. The FACS trial result has raised substantial interest in CEA as a first-line follow-up modality.

CEA is a glycoprotein involved in cell adhesion produced during foetal development. Production usually ceases at birth, but elevated levels can be detected in people with colorectal, breast, lung and pancreatic cancer, in smokers, and in people with 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). Prior to first diagnosis, CEA levels may rise between four and eight months before the development of cancer-related symptoms (Goldstein 2005). Approximately 90% of colorectal cancers produce CEA (Dallas 2012). Predicting those people who do not secrete CEA is a challenge, with conflicting reports regarding whether well- or poorly-differentiated tumours are associated with increased secretion (Davidson 1989). During follow-up, CEA appears to be most sensitive for detecting hepatic and retroperitoneal metastases, and is least sensitive for local recurrences and peritoneal or pulmonary disease (Scheer 2009; Tsikitis 2009). However, CEA needs to be seen as a triage test (where a rise should lead to further investigation rather than initiation of therapy), as it gives no information about the location and extent of recurrence (Duffy 2013b).

Although serial CEA measurements are taken during follow-up, the decision to investigate further with imaging is usually based on a single elevated CEA measurement (although a repeat blood test is often done to confirm the raised level). An absolute threshold somewhere between 3 and 7 µg/L is typically used to trigger further investigation. In the FACS trial, the threshold used was based on the difference of the CEA level at a single time point from the postoperative baseline (Primrose 2014).

The most recent systematic review exploring the accuracy of CEA for diagnosing recurrent CRC includes a meta-analysis of 20 studies (Tan 2009). These studies implemented a wide range of thresholds (3 to 15 µg/L) and measured CEA using a variety of test kits. The pooled estimates of sensitivity and specificity were 64% (95% CI 61% to 67%) and 90% (95% CI 89% to 91%) respectively. The pooled area under the curve (AUC) was 0.79 (standard error = 0.054). A subgroup analysis of four studies that reported accuracy at a threshold of 3 µg/L gave an improved sensitivity of 73% (95% CI 69% to 77%) but at the expense of a reduced specificity of 68% (95% CI 65% to 72%). Based on a metaregression analysis, the authors suggest that a cut-off of 2.2 µg/L provides the ideal balance between sensitivity and specificity, but this is based on extrapolation beyond the data analysed, as the lowest threshold applied in any included study was 3 µg/L. We were also unable to identify some of the data included in the analysis from the published studies.

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 high-income countries and 10.7 in low- or middle-income countries) (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 haematological metastasis (Gore 1997). This process provides the basis of the internationally recognised TNM (tumour node metastasis) staging system (Sobin 2009) and the 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 in the lungs but also in the bones and brain (Guthrie 2002). Prognosis is closely related to stage, with higher-grade metastatic tumours having a poorer prognosis (Maringe 2013). Approximately two-thirds of patients will present with a primary CRC amenable to radical surgery (Jeffery 2007).

Following surgery, however, 30% to 50% of patients will develop recurrence (Labianca 2010), although the results of the FACS trials suggest that perhaps half these cases result from inadequate preliminary staging and might have been detectable through more rigorous investigation at the time of primary treatment (Primrose 2014). The most common site for recurrence is the liver, followed

by the lungs, but it can also occur in the abdomen and pelvis (Cunningham 2010; Jeffery 2007).

As stated in the Background, the effectiveness of treatment of recurrence is a matter of hot debate (Godlee 2014; Treasure 2014). In the absence of trials of treatment versus no treatment, most estimates of impact are based on observational data. Patients undergoing secondary surgery with curative intent have a median survival time of 35.8 to 84.8 months. Chemotherapy has been estimated to prolong life by one to two years (Arriola 2006; Cunningham 2010; Tsikitis 2009). However, apart from the Nordic trial showing that the initiation of chemotherapy at an asymptomatic stage increases survival (Glimelius 1992), there is no evidence from trials to confirm that treatment of early-diagnosed asymptomatic recurrence improves survival or other outcomes. There is a need therefore to determine the most accurate means of detecting early-stage recurrence before the impact of treatment strategies can be further explored.

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 use fully automated non-isotopic methods. The reproducibility of these fully automated methods are generally superior to the older manual methods. Unfortunately, the details of the methods used in clinical studies and their analytical performance are often lacking (Wild 2013).

Data from external quality assessment schemes have repeatedly shown good precision for most methods at low CEA concentrations. In 2010, within-laboratory precision over a 12-month period at a concentration of 3 µg/L (equivalent to 54 U/L) was less than 9% on average 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 1990s, method bias may still be ± 20%, and the degree of this bias is often 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, which is considered to be a major source of method bias (Bormer 1991). Consequently, the interpretation of data from clinical studies, especially the use of any particular threshold, needs to take account of the actual method used. Due to the good reproducibility but significant method-dependent bias, it is advised that the same assay technique should be used throughout any follow-up period (Duffy 2013b).

Clinical pathway

Following radical surgery (with or without adjuvant therapy), there is wide variation in the recommended intensive follow-up regimen (Duffy 2013b; 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 the first three years, and every six to 12 months in years four and five. A colonoscopy is recommended at one year, then 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 is recommended every six to 12 months for the first three years in patients considered to be at higher risk. Other laboratory and radiological examinations are not recommended unless patients have suspicious symptoms (Labianca 2010).

The American Society of Clincal Oncology (ASCO) recommends that CEA is performed every three months for the first 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 (> 5 µg/L, confirmed by a repeat test) warrant further evaluation for metastatic disease (Locker 2006). Unlike ASCO, ESMO does not specify a threshold nor limit testing to specific tumour stages. 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 to 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 2013b).

The National Institute for Health and Clinical Excellence (NICE) recommended 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 on the basis of a raised CEA level, patients then undergo further diagnostic testing to confirm recurrence (Duffy 2013a). The modality used to provide a definitive diagnosis is usually either CT or USS, but could also 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 most frequently undertaken modality within an intensive follow-up regimen. Prior testing in this context is irrelevant, because CEA is measured routinely within intensive follow-up programmes.

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 2013b; Newton 2011). CT imaging is the only other test that meta-analysis suggests has potential to detect metastatic recurrence amenable to resection, but it is more expensive than measuring blood CEA. CT-PET is used in some centres, but will only be preferred to standard CT for routine follow-up if future evidence suggests much superior performance. Endoscopic imaging (colonoscopy) is routinely used as an adjunct to CEA or CT imaging or both in follow-up care to detect metachronous polyps or cancer (and rarely intraluminal recurrence). Clinical and ultrasound examination lack sensitivity. MRI can realistically be applied only to the liver and lacks strong evidence of effectiveness in detecting recurrence.

Rationale

This diagnostic test accuracy (DTA) review aims to clarify the accuracy of blood CEA as a triage test for CRC recurrence. If found to be sufficiently accurate, CEA could be a cost-effective means of reducing unnecessary, more expensive investigations.

OBJECTIVES

To determine the diagnostic performance of different blood CEA levels in identifying people with colorectal cancer recurrence in order to inform clinical practice.

Secondary objectives

To identify sources of between- and within-study heterogeneity to inform future study designs.

METHODS

Criteria for considering studies for this review

Types of studies

We include cross-sectional diagnostic test accuracy studies, cohort studies, and RCTs that directly compared follow-up after CRC resection using CEA to a reference standard. We included studies only if we could extract 2 x 2 accuracy data. We excluded case-control studies, as the ratio of cases to controls is determined by the study design, making the data unsuitable for assessing test accuracy.

Participants

Participants were adults with no detectable residual disease after primary treatment with surgical resection (with or without adjuvant therapy) being followed-up for recurrence.

Index tests

Blood carcino-embryonic antigen (CEA).

Target conditions

Recurrence of colorectal cancer 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).

We had hoped to compare the results of using these different reference standards in a sensitivity analysis. However, the majority of studies (73%) reported a composite reference standard, including more than one of the three reference standards listed above, as part of a prespecified clinical pathway and so the specific reference standard applied varied between participants within the same study. Without individual patient data, identifying the exact investigative modality applied as the reference standard was not possible and so we did not conduct the planned sensitivity analysis.

We classified the chosen reference standard (or composite reference standard) used in each study as 'appropriate' (1 to 3 above), 'inappropriate' (a reference standard not included in 1 to 3 above), or 'not stated' for further subgroup analysis.

There were insufficient data available to classify deaths during follow-up as 'death from CRC', 'death with CRC', 'death from other causes', or 'death unspecified', as detailed in the original protocol.

Search methods for identification of studies

Electronic searches

Our information specialist (NR, trained in Cochrane DTA methodology) designed our search strategy, and conducted all searches to January 29 2014. We applied no language limits to the searches, and translated non-English manuscripts to assess suitability for inclusion.

We searched for relevant reviews in the MEDION database (www.mediondatabase.nl), using the search terms 'cea' OR 'carcinoembryonic' or 'carcino-embryonic' and restricting to Malignancy OR Digestive. Using the same terms, we also searched MEDLINE (OvidsSP) [1946 to current, In-process], and EMBASE (OvidSP) [1974 to current] using the Reviews Clinical Query, and the DARE database (the Cochrane Library, Wiley).

We searched for primary studies (including conference abstracts) in the Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Library, Wiley (Appendix 1), MEDLINE (OvidSP) [1946 to current, In-process] (Appendix 2), EMBASE (OvidSP) [1974 to current] (Appendix 3), , and the Science Citation Index & Conference Proceedings Citation Index - Science (Web of Science, Thomson) [1945 to current] (Appendix 4).

We identified ongoing studies by searching WHO ICTRP (apps.who.int/trialsearch/) using the following search terms: (Condition = (colorectal cancer OR colon cancer OR colorectal neoplas* OR colon neoplas* OR rectal cancer OR rectal neoplas*) AND Intervention = (cea OR Carcinoembryonic Antigen OR carcinoembryonic antibod*)), and by searching ClinicalTrials (clinicaltrials.gov) using the following search terms: (Condition = (colorectal cancer OR colon cancer OR colorectal neoplas* OR colon neoplas* OR rectal cancer OR rectal neoplas*) AND Intervention = (cea OR Carcinoembryonic Antigen OR carcinoembryonic antibod*)).

We conducted an additional search of the ASCO meeting library (meetinglibrary.asco.org/) for conference abstracts using the following search terms: (Title word search: "cea OR "carcinoembryonic antigen" OR "carcinoembryonic antigen").

Searching other resources

Following the search of bibliographic databases, we checked reference lists of retrieved reviews and all included studies. In addition, we performed a 'Related articles' search on PubMed on all included studies.

In the protocol, we stated we would contact the principal investigators of all included studies to identify further relevant literature, clarify methodological queries if they exist and to ask for any unpublished data relevant to this review. Unfortunately, due to time constraints and the large number of studies included in our review, we were not able to do this.

Data collection and analysis

Selection of studies

To identify relevant studies, two review authors (BDN and IP) scanned all titles and excluded those studies clearly not relevant to the topic of CEA for the detection of CRC recurrence. Following this, the same two review authors (BDN, IP) independently assessed both the titles and abstracts of the selected studies and retrieved the full-text articles for those deemed to be relevant and for those where a decision could not be made on the basis of the title and abstract alone.

We assessed the remaining full-text articles to see whether 2 x 2 accuracy data were available and, if so, we included the study in the review and implemented a full data extraction. Reasons for exclusions are detailed in Figure 1. A third review author (BS) resolved any disputes over which references should be included.

Data extraction and management

Full data extraction was guided by a background information sheet describing how each item should be interpreted. Two review authors piloted and refined this form, using three initial studies. A third review author resolved any disagreements over extracted data.

We extracted data into an Excel spreadsheet under the following headings: author, year, title, country, study design, setting, dates of data collection, population (n), inclusion criteria, exclusion criteria, included participants (n), age, smoking status, site of primary tumour, stage/grade of primary tumour, investigations done to ensure no residual disease, chemotherapy/radiotherapy, follow-up schedule, cases of recurrence (n), CEA timing, CEA technique, CEA threshold, reference standard, timing of CEA versus reference standard, true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), AUC, QUADAS-2 items (including CEA laboratory technique, Appendix 5).

In the protocol we stated we would contact authors if data were not available, but due to time constraints we were not able to do this.

Assessment of methodological quality

Assessment of methodological quality

QUADAS-2 is a generic set of criteria for assessing the quality of diagnostic accuracy studies. It consists 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 provided to guide judgement of the risk of bias across these four domains (Whiting 2011).

We modified QUADAS-2 to exclude items not applicable to this review. A guide to the operational definitions for the modified QUADAS-2 items can be found in Appendix 5.

We included additional questions regarding index test repetition (4.A.1) and CEA laboratory technique (2.A.2 to 2.A.4). We 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 also 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?". We removed "Was a case-control design avoided?" from the original QUADAS-2 template as we excluded all case-control studies. We also removed "Were the index test results interpreted without knowledge of the results of the reference standard?" as knowledge of the reference test result would not bias the interpretation of a positive or negative CEA result, as CEA is an objective test using a predetermined dichotomous threshold.

For the index test domain, items were weighted so that the use of a prespecified threshold and a consistent method for CEA measurement had more influence on the overall judgement than the items regarding estimation of method reproducibility and indication of method accuracy. We made this decision as the latter two items were very rarely reported.

For the reference standard domain, items were weighted so that correctly classifying recurrent CRC had more influence on the overall judgement than whether the reference standard was interpreted without the knowledge of the index test. We made this decision as there were no blinded studies included in the review.

For the flow and timing domain, the five items were weighted so that the inclusion of all patients in the final analysis had the most influence and everyone receiving a reference standard was second most influential. Repetition of the index test prior to the reference standard, ascertainable timing between the index test and reference standard, and to all patients receiving the same reference standard were weighted equally lower.

Signalling questions weighted as high priority determined the overall rating within each domain.

Two review authors (BDN, IP) assessed the quality of all articles independently, discussing any disagreements. Where they could not reach consensus, a third author (BS) acted as moderator. We used the results of the quality assessment 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

We used descriptive statistics to present summary data for each included study. The Characteristics of included studies tables detail patient sample, study design, CEA technique, follow-up characteristics and the CEA threshold(s) at which accuracy was reported.

We extracted binary diagnostic accuracy data from all included studies as 2 x 2 tables. We present the risk of bias results for each of the four domains of the QUADAS-2 assessment graphically as described by Whiting 2011.

Inferential statistics were guided by Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010).

We used Review Manager 5 to produce forest plots showing the variability of sensitivity and specificity across primary studies, with corresponding 95% confidence intervals for visual comparison. For studies reporting more than one threshold, we extracted 2 x 2 data for all thresholds. We plotted sensitivity and specificity estimates from each study in ROC space, using the inverse standard error of each estimate to adjust the size of each box to represent precision. For both of these graphs, we included sensitivity and specificity at the threshold closest to 5 µg/L (the most commonly reported threshold). We did not conduct a meta-analysis across all of the included studies, as we had a sufficient number of studies to carry out meta-analyses at specific thresholds (see next section), which is clinically more informative.

We used the bivariate model to perform meta-analysis of sensitivity and specificity (Reitsma 2005). We conducted analyses using the xtmelogit command in Stata (Takwoingi 2013).

We estimated the absolute numbers of false alarms (false positives) and missed cases (false negatives) per 1000 patients tested for each three-monthly testing interval by applying the pooled sensitivity and specificity derived from this review to: 1) the observed median reported prevalence of recurrence divided by 15 (national guidance is to conduct 14 to 15 CEA tests during follow-up); 2) the incidence of recurrence data per follow-up period reported by Sargent 2007 (as in reality the proportion developing recurrence between tests is not constant but falls over time).

Investigations of heterogeneity

Based on the results of the quality assessment, we determined the following most likely sources of heterogeneity: effect of CEA threshold, whether a single CEA measurement or serial measurements were evaluated, and the laboratory techniques employed.

For each subgroup analysis, we conducted bivariate meta-analyses (Reitsma 2005), using the xtmelogit command in Stata to produce pooled estimates of sensitivity and specificity. Summary ROC plots and forest plots are reported to provide a basic picture of between-study variability in these accuracy estimates.

CEA Threshold

For tests producing a continuous outcome, the threshold at which a positive result is defined directly impacts on the accuracy of the test. The use of different thresholds between studies is therefore a key source of heterogeneity.

We investigated the effect of threshold by carrying out subgroup meta-analyses for thresholds where sufficient data were available. As some studies reported 2 x 2 data for more than one threshold, this analysis allowed us to include all of the available data. We used

Review Manager 5 to produce a forest plot showing the variability of sensitivity and specificity across primary studies at specific thresholds.

Although the original plan was to apply a meta-analysis method incorporating more than one 2 x 2 table from a single study (Hamza 2009), this method requires data to be reported at consistent thresholds across all included studies, and this was not the case in our review.

Timing of CEA Measurement

Despite sequential CEA measurements being taken in the majority of studies, 2 x 2 data were not reported for each scheduled measurement in any of these studies.

Some studies provided 2 x 2 data for the CEA measurement taken closest to the time point at which recurrence was detected or, for patients who did not experience recurrence, their final follow-up measurement. Others looked across all of the measurements available for each individual to assess whether any of the sequential measurements had crossed the threshold during the entire follow-up period. This approach meant the time interval between a rise in CEA and confirmed recurrence was variable across individuals within the same study, but this interval was not reported in any study. Consequently, we classified a patient without confirmed recurrence during the follow-up period and at least one measurement above the threshold as a false positive in the 2 x 2 table, and a patient with confirmed recurrence but without any CEA rise above the threshold as a false negative.

As this information was not consistently reported in all studies, we could not include this variable in the metaregression analysis. Instead, we explored whether this had a significant impact on accuracy by carrying out a subgroup analysis on those studies that did provide this information. This analysis was also limited to studies reporting accuracy at 5 µg/L (the most commonly reported threshold) to avoid any threshold effects.

Laboratory Technique

The intention was to carry out subgroup analyses on studies using the same laboratory technique in order to assess the effect of technique on accuracy. However, given that so few studies provided sufficient detail regarding the laboratory technique employed, this was not possible. We were interested in exploring whether the implementation of IRP 73/601 reduced between-study variability in sensitivity and specificity. We therefore used the information provided in each study to assess whether laboratory methods predated the introduction of IRP (e.g. manual Radioimmunoassay (RIA) and Immunoradiometric assay (IRMA) methods) and whether the samples were analysed pre-1992. We then carried out a subgroup analysis and compared the widths of the 95% confidence intervals for the pooled estimates of sensitivity and specificity. We again limited this analysis to those studies reporting accuracy at 5 µg/L to avoid threshold effects.

Sensitivity analyses

To explore whether study quality biased the sensitivity and specificity of CEA, we planned a subgroup analysis to include those studies which had a low risk of bias across all four domains. We also carried out a metaregression analysis using the 'Metadas' macro in SAS, including all of the four domains as ordinal covariates (low risk, unclear, high risk).

Assessment of reporting bias

As described in the protocol and by Van Roon 2011, investigation of publication bias in DTA studies is known to be problematic, and so we have not included assessment of reporting bias in this review (Deeks 2005; Leeflang 2008; Song 2002).

RESULTS

Results of the search

Figure 1 summarises the studies that we identified, screened and selected for this review. Our search resulted in 6782 hits, including 6571 primary studies, 128 reviews, 46 conference abstracts, and 37 registered trials. We identified 45 additional articles by checking the reference lists of retrieved reviews and by performing a 'Related articles' search in PubMed. We removed duplicates (n = 3016), leaving 3811 records for title and abstract screening. Of these, we requested 268 full-text articles for review, of which we excluded 216 (see Figure 1 for reasons for exclusion). Fifty-two studies met our inclusion criteria and are included in the final review.

Figure 1. PRISMA flow diagram: results of the search for studies evaluating the diagnostic accuracy of blood CEA to detect recurrent colorectal cancer in patients following curative resection.

Included studies

Prevalence

Included studies were published between 1974 and 2014 and were conducted across 22 countries. All studies were conducted in secondary care, except one Norweigian prospective study (Johnson 1985) in which follow-up was conducted in both primary and secondary care. In total, 9717 patients were included, and 2951 recurrences detected. The median number of participants in the studies was 139 (interquartile range (IQR): 72 to 247) and the proportion of recurrences detected ranged from 13.5% (Fezoulidis 1987) to 72.3% (Ochoa-Figueroa 2012) (median: 29.5%, IQR: 24.3 to 36.3%).

Study Design

In 24 studies (46%) a prospective design was used, three of which were randomised controlled trials (RCTs) (McCall 1994; Ohlsson 1995; Steele 1982). One study prospectively followed up a cohort of patients of whom some were identified retrospectively (Tate 1982), while another sampled retrospectively from a prospective cohort (Korner 2007). The remaining 26 studies (50%) used a retrospective design.

Clincal features of included patients

Location of recurrence

The location of recurrence was reported in 25 studies (48%) including local, locoregional, and distant recurrence. However, the description of CRC recurrence was heterogeneous and all studies lacked 2 x 2 tables for the diagnostic accuracy of CEA to detect recurrence at each location (Characteristics of included studies).

Staging of primary colorectal cancer

Apart from the two studies (4%) which included only patients with rectal cancer (Barillari 1992; Fezoulidis 1987), the majority of studies (n = 50, 96%) included patients with both colon and rectal cancer.

Thirty-three studies (63%) used the Dukes staging to describe the primary CRC. A further 11 studies (21%) used the TNM grading system and one study (2%) used the Astler-Coller staging. The staging was unclear or not reported in the remaining seven studies (13%) (Carlsson 1983; Kohler 1980; Koizumi 1992; Li Destri

1998; Mittal 2011; Ochoa-Figueroa 2012; Wood 1980).

Of those using Dukes staging, seven included Dukes A - D (Banaszkiewicz 2011; Carpelan-Holmström 2004; Jubert 1978; Mach 1978; Mariani 1980; Seregni 1992; Yu 1992); 15 included Dukes A - C (Barillari 1992; Deveney 1984; Farinon 1980; Fezoulidis 1987; Fucini 1987; Graffner 1985; Hine 1984; Irvine 2007; Kato 1980; Korner 2007; Luporini 1979; Mackay 1974; McCall 1994; Ohlsson 1995; Triboulet 1983); three used Dukes B - C (Beart 1981; Steele 1982; Wang 1994); two used Dukes C (Hara 2008; Tobaruela 1997); one used Dukes A - C plus palliative cases (Johnson 1985); one used Dukes A - C plus unknown cases (Tate 1982); and four used Dukes A - D plus unknown cases (Bjerkeset 1988; Engaråas 2003; Miles 1995; Minton 1985).

Of the 11 studies using the TNM grading system: five included TNM I - III (Kanellos 2006a; Ohtsuka 2008; Park 2009; Tang 2009; Yakabe 2010); four used TNM I - IV (Carriquiry 1999; Nishida 1988; Peng 2013; Staib 2000); and one included TNM II - III (Kim 2013). Only one study reported 2 x 2 tables by stage, reporting on TNM II and TNM III (Hara 2010).

The study that used Astler-Coller staging included A - C2 (Lucha 1997).

Smokers

Three studies explicitly excluded smokers (Kanellos 2006a; Mariani 1980; Staib 2000), four studies explicitly included some smokers (but there was no way of identifying these patients in the 2 x 2 tables), and the remaining studies did not report smoking status. In the two studies which gave precise figures for smoking prevalence, it was low at 2% smokers (Fucini 1987) and 9% heavy smokers (Mach 1978).

Investigations for residual disease

In 43 studies (83%) it was not clear which (if any) perioperative investigations were done to ensure there was no residual disease before entering follow-up. In the nine studies that reported this information, three reported using a persistent postoperative elevation of CEA as evidence of residual disease (Hara 2008; Irvine 2007; Steele 1982); one used "signs" of malignancy at the first follow-up examination (Tate 1982); one used preoperative colonoscopy to resect any lesions outside the section of bowel planned for resection (Banaszkiewicz 2011); one reported using the intraoperative detection of gross residual disease (Lucha 1997); one specified no gross residual disease and clear resection margins (Bjerkeset 1988); one used preoperative abdominal CT and interoperative palpation to exclude liver metastases (Kanellos 2006a); and one reported using preoperative barium enema (BE), chest x-ray (CXR), liver function tests (LFTs) and CEA, and postoperative BE and colonoscopy to ensure there was no residual disease (Ohlsson 1995).

Treatment

In 14 studies (27%) some (but not all) patients received chemotherapy, and in no studies was a subgroup analysis performed comparing the diagnostic accuracy of CEA in those receiving chemotherapy compared to those who did not (Characteristics of included studies).

Reference standard

In 38 studies (73%) a composite reference standard was used, the composition of which varied greatly between studies (see Characteristics of included studies). In 12 of these, a predefined multimodal follow-up schedule was used for each patient (although the composition of these varied across studies) (Banaszkiewicz 2011; Carlsson 1983; Fucini 1987; Hara 2008; Irvine 2007; Jubert 1978; Kanellos 2006a; McCall 1994; Ohlsson 1995; Park 2009; Peng 2013; Steele 1982). In 26 studies (50%) a predefined composite follow-up schedule was used to trigger further investigations for suspected recurrence.

A single investigation was used in three studies (6%) (Mittal 2011; Ochoa-Figueroa 2012; Staib 2000), of which one reported 2 x 2 tables separately for PET and for CT (Ochoa-Figueroa 2012).

In the remaining 11 studies (21%), it was unclear what was used as a reference standard.

CEA measurement

The use of predefined follow-up schedules resulted in multiple CEA measurements being available for analysis.

Eight studies (15%) reported the accuracy of the CEA measurement closest to the time at which recurrence was detected by the reference standard, whilst nine studies (17%) defined CEA as positive if any CEA measurement crossed the threshold at any time within the follow-up period. In a subset of studies, the authors stated clearly that a single 'positive' measurement would be followed up by a repeat test to confirm the result.

For the remaining 35 studies (67%), it was impossible to unpick which CEA value had been used, due to limited reporting.

Reporting units

CEA studies have used both ng/mL and µg/L in their publications. Numerically these are the same value and for consistency we have used µg/L throughout the review.

Laboratory technique

Details regarding laboratory methods for CEA analysis were inconsistently reported across the included studies. Based on the available information relating to laboratory technique, we were able to able to group the studies as follows:

1. Twenty-two studies (42%) analysed samples before the introduction of the international reference preparation (IRP) using manual RIA and IRMA methods;
2. Seven studies (13%) used an identifiable laboratory technique following introduction of IRP;
3. Eight studies (15%) used unfamiliar laboratory techniques after the introduction of IRP;
4. Fifteen studies (29%) did not report laboratory technique.

For the seven studies reporting an identifiable laboratory technique following IRP introduction, six distinct techniques were used: Autodelfia post-year 2000 (Carpelan-Holmström 2004; Engaråas 2003); Abbott automated instrumentation (Korner 2007); Bayer Immuno 1 (Irvine 2007); Siemens ADVIA centaur (Kim 2013); Roche elecsys (Mittal 2011); and Diasorin/byk santec liaison (Staib 2000). Across these, four thresholds were reported: 3 µg/

L (Staib 2000); 5 µg/L (Carpelan-Holmström 2004; Kim 2013; Mittal 2011); 5.6 µg/L (Engarås 2003); and 10 µg/L (Irvine 2007; Korner 2007).

Forty-three studies (83%) did not report an estimation of CEA method reproducibility nor an indication of method accuracy. Of the remaining nine studies, three (6%) reported both an estimation of reproducibility and an indication of method accuracy (Carpelan-Holmström 2004; Engarås 2003; Steele 1982), four (8%) clearly reported only an estimation of reproducibility (Fucini 1987; Hine 1984; Mach 1978; Mackay 1974), and the remaining two (4%) reported only the indication of method accuracy (Irvine 2007; Miles 1995).

Excluded studies

Of the 216 excluded full-text articles (Figure 1; Characteristics of excluded studies):

- 152 studies (70%) did not report complete 2 x 2 data, and 74 (34%) reported no 2 x 2 data at all: 59 (27%) only reported recurrences; 16 (7%) only reported CEA positive cases; and three (1%) only CEA negative);
- 23 studies (11%) did not conduct a single-point diagnostic test accuracy study (14 (6%) used alternative analyses (trend, nomogram, slope, or median CEA); five were case-control studies (2%); three (1%) were review articles; and one was an economic analysis);
- 14 studies (6%) did not report an analysis of serum CEA measurements taken as part of a follow-up schedule (seven (5%) reported preoperative CEA measurements; six (3%) reported the prognostic value of one postoperative CEA measurement; and one used intraoperative portal vein sampling);
- eight studies (4%) included fewer than 30 patients;
- six studies were unavailable or needed translation (five studies (2%) were not retrieved after worldwide search by the British Library, and we were not able to translate the remaining study);
- five studies (2%) did not clearly report colorectal cancer recurrence (three (1%) reported on only liver metastases; and two (1%) reported colorectal cancer recurrence together with other cancer types);
- five studies (2%) reported datasets already included in the review;
- three studies (1%) reported non-curative surgery.

We have not included two large RCTs in the review: the FACS trial (as 2 x 2 data were not reported in the published paper (Primrose 2014)), and the CEASL trial, which was published following our search and did not report on negative CEA cases (Treasure 2014).

Methodological quality of included studies

We assessed all 52 studies using the QUADAS-2 framework. Figure 2 shows the summary of overall risk of bias and applicability concerns, and Figure 3 presents the risk of bias and applicability concerns as overall percentages.

Figure 2. QUADAS-2 risk of bias and applicability concerns summary including review authors' judgements about each domain for each included study

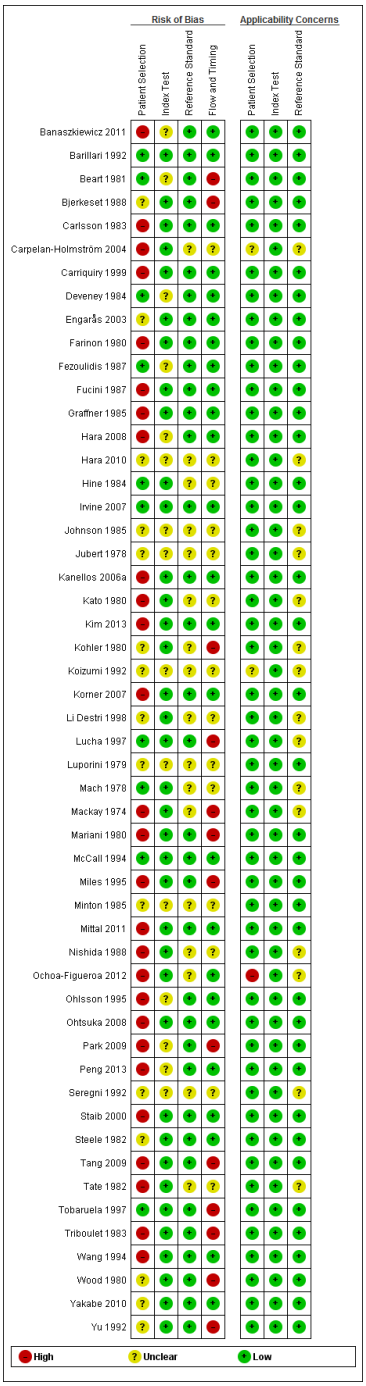
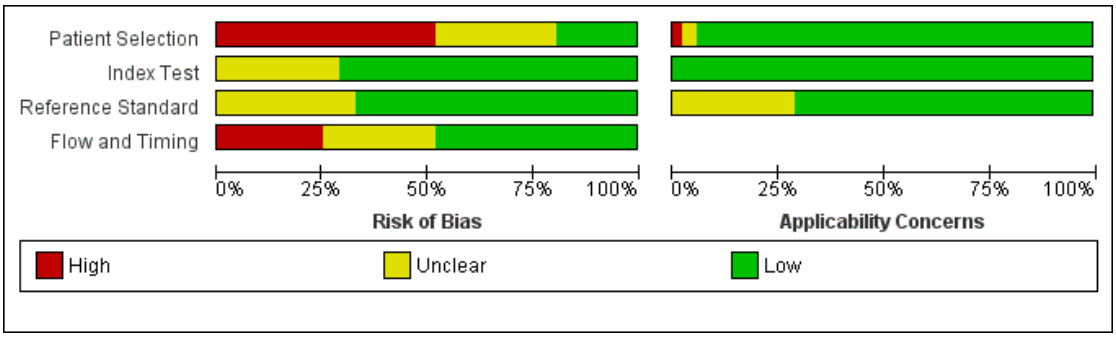


Figure 3. QUADAS-2 risk of bias and applicability concerns graph including review authors' judgements about each domain presented as percentages across included studies



Three studies (6%), including 516 participants of whom 177 experienced recurrence, were assessed as being at low risk of bias and low concern regarding applicability across all domains (Barillari 1992; Irvine 2007; McCall 1994). Across these studies, each reported a different threshold (3, 10, and 5 µg/L respectively) using CEA test kits from three different manufacturers (with poor description of method accuracy). Each study applied a different but "appropriate" follow-up schedule to detect recurrence. Consequently, the planned subgroup analysis of high-quality studies (low risk of bias in all four domains) was not feasible.

Risk of bias

We judged 34 studies (65%) to be at high risk of bias in at least one of the four domains (Figure 3).

For the patient selection domain, items were weighted so that the presence of inappropriate exclusions had more influence on the overall judgement than the presence of a consecutive or random sample. Of the 27 studies judged to be at high risk of bias for patient selection (52%), inappropriate exclusions were based on:

- advanced age (Carlsson 1983; Graffner 1985; Korner 2007; Ohlsson 1995);
- previous malignancy (Ohtsuka 2008);
- poor prognosis for further surgery (Banaszkiewicz 2011; Ohlsson 1995);
- preoperative CEA values (Carpelan-Holmström 2004; Carriquiry 1999; Farinon 1980; Miles 1995; Tang 2009; Wang 1994);
- temporary rises in CEA (Mackay 1974);
- non-rising CEA (Mittal 2011);
- using a minimum follow-up period (Carriquiry 1999; Kim 2013; Korner 2007; Mackay 1974);
- 'incomplete' follow-up measurements (Carpelan-Holmström 2004; Carriquiry 1999; Hara 2008; Kato 1980; Korner 2007; Nishida 1988);
- factors related to other follow-up tests (Fucini 1987; Peng 2013; Tang 2009; Triboulet 1983);
- early signs of malignancy or death (Carlsson 1983; Tate 1982);
- smoking status (Kanellos 2006a; Mariani 1980);
- concomitant benign disease or recent surgery (Kanellos 2006a; Mariani 1980; Park 2009; Peng 2013; Staib 2000);
- patients not presenting a "diagnostic problem" (Staib 2000).

There were no studies deemed to be at high risk of bias based on the judgements made about the index test.

There were no studies at high risk of bias based on the judgements made about the reference standard, and in 17 (33%) the risk was unclear.

Thirteen studies (25%) were deemed to be at high risk of bias based on flow and timing. In four studies, not all patients were included in the final analysis (Beart 1981; Bjerkeset 1988; Kohler 1980; Park 2009). In the remaining nine studies, a raised CEA value triggered the reference standard which could introduce work-up bias and result in false negative CEA results being misclassified as true negative results (Lucha 1997; Mackay 1974; Mariani 1980; Miles 1995; Tang 2009; Tobaruela 1997; Triboulet 1983; Wood 1980; Yu 1992).

Applicability concerns

We judged 37 studies (71%) to be at low risk of applicability concerns in all three domains (Figure 3). We rated only one study (Ochoa-Figueroa 2012) at high risk of applicability concerns in relation to patient selection, as it did not include all patients undergoing postoperative follow-up, but only those referred with sus-

pected recurrence to the Department of Nuclear Medicine for fluoro-deoxy-glucose (FDG) PET-CT. There were no studies deemed to be at high risk for applicability based on the index test or reference standard.

Unclear risk

Of the 364 domains, we deemed 85 (23%) to be at unclear risk of bias or applicability. For the vast majority of these items poor reporting accounted for the unclear rating.

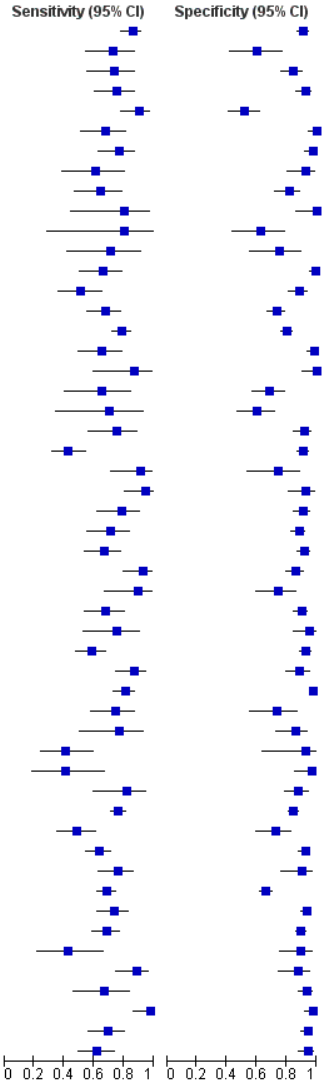
Findings

Diagnostic accuracy

The forest plot in Figure 4 (Analysis 1) shows the range of sensitivity and specificity of CEA for the detection of recurrent colorectal cancer across all 52 included studies.

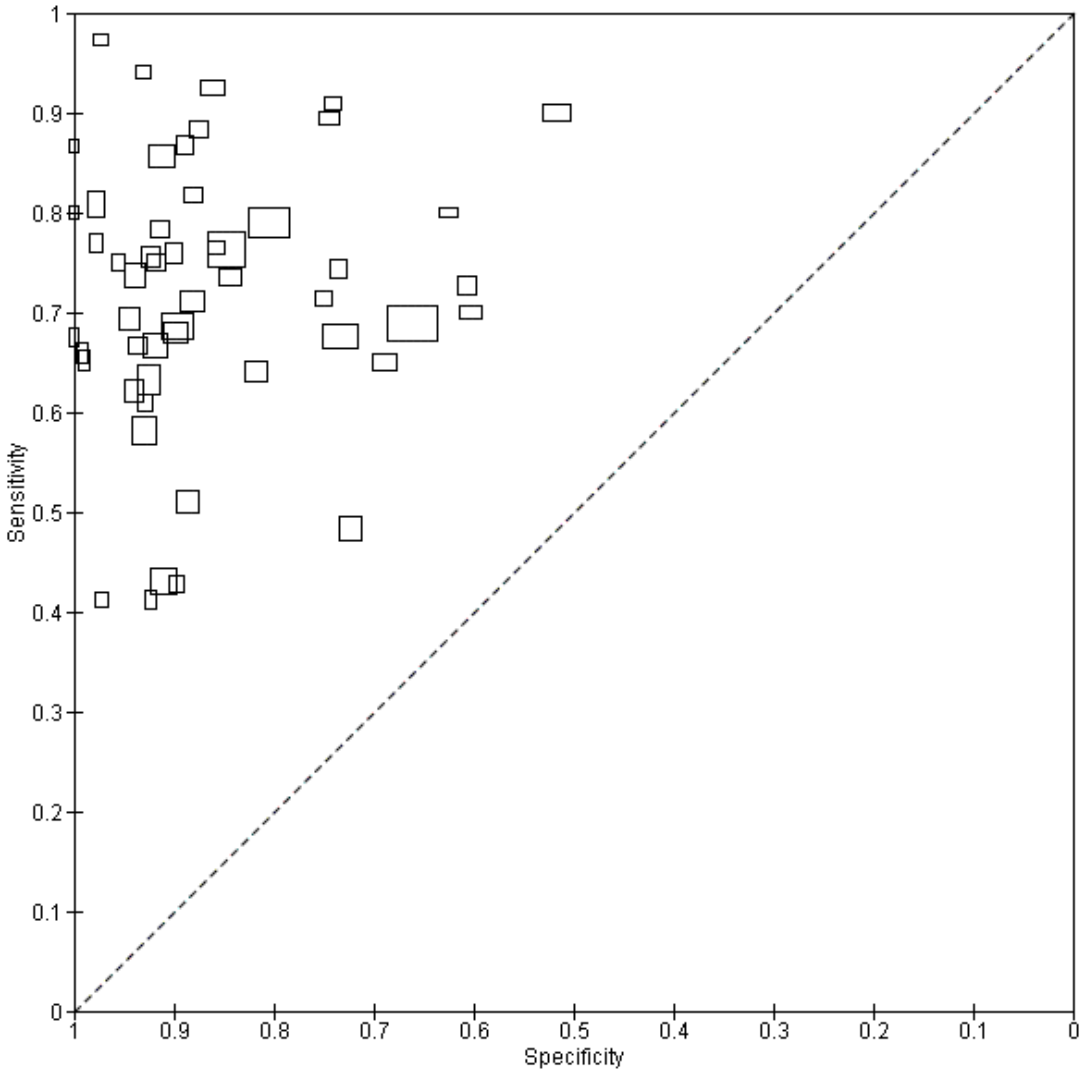
Figure 4. Forest plot for all 52 included studies for the threshold reported closest to 5 µg/LTP = true positive; FP = false positive; FN = false negative; TN = true negativeThe blue square depicts the sensitivity and specificity for each study and the horizontal line represents the corresponding 95% confidence interval for these estimates.

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Banaszkiewicz 2011	96	20	16	208	5.0	0.86 [0.78, 0.92]	0.91 [0.87, 0.95]
Barillari 1992	24	13	9	20	3.0	0.73 [0.54, 0.87]	0.61 [0.42, 0.77]
Beart 1981	25	18	9	97	5.0	0.74 [0.56, 0.87]	0.84 [0.76, 0.90]
Bjerkeset 1988	34	9	11	108	5.0	0.76 [0.60, 0.87]	0.92 [0.86, 0.96]
Carlsson 1983	45	43	5	46	3.0	0.90 [0.78, 0.97]	0.52 [0.41, 0.62]
Carpelan-Holmström 2004	27	0	13	62	5.0	0.68 [0.51, 0.81]	1.00 [0.94, 1.00]
Carriquiry 1999	40	2	12	88	5.0	0.77 [0.63, 0.87]	0.98 [0.92, 1.00]
Deveney 1984	14	3	9	39	5.0	0.61 [0.39, 0.80]	0.93 [0.81, 0.99]
Engarås 2003	25	17	14	76	5.6	0.64 [0.47, 0.79]	0.82 [0.72, 0.89]
Farinon 1980	8	0	2	25	6.0	0.80 [0.44, 0.97]	1.00 [0.86, 1.00]
Fezoulidis 1987	4	12	1	20	2.5	0.80 [0.28, 0.99]	0.63 [0.44, 0.79]
Fucini 1987	10	7	4	21	20.0	0.71 [0.42, 0.92]	0.75 [0.55, 0.89]
Graffner 1985	31	1	16	119		0.66 [0.51, 0.79]	0.99 [0.95, 1.00]
Hara 2008	26	14	25	109	5.0	0.51 [0.37, 0.65]	0.89 [0.82, 0.94]
Hara 2010	50	61	24	168	5.0	0.68 [0.56, 0.78]	0.73 [0.67, 0.79]
Hine 1984	135	89	36	366	20.0	0.79 [0.72, 0.85]	0.80 [0.76, 0.84]
Irvine 2007	30	1	16	92	10.0	0.65 [0.50, 0.79]	0.99 [0.94, 1.00]
Johnson 1985	13	0	2	36		0.87 [0.60, 0.98]	1.00 [0.90, 1.00]
Jubert 1978	13	24	7	53	2.5	0.65 [0.41, 0.85]	0.69 [0.57, 0.79]
Kanellos 2006a	7	25	3	38	5.0	0.70 [0.35, 0.93]	0.60 [0.47, 0.72]
Kato 1980	24	8	8	89	2.5	0.75 [0.57, 0.89]	0.92 [0.84, 0.96]
Kim 2013	34	23	45	234	5.0	0.43 [0.32, 0.55]	0.91 [0.87, 0.94]
Kohler 1980	20	7	2	20	2.5	0.91 [0.71, 0.99]	0.74 [0.54, 0.89]
Koizumi 1992	32	3	2	40	5.0	0.94 [0.80, 0.99]	0.93 [0.81, 0.99]
Korner 2007	29	10	8	106	10.0	0.78 [0.62, 0.90]	0.91 [0.85, 0.96]
Li Destri 1998	32	23	13	171	5.0	0.71 [0.56, 0.84]	0.88 [0.83, 0.92]
Lucha 1997	44	18	22	201	5.0	0.67 [0.54, 0.78]	0.92 [0.87, 0.95]
Luporini 1979	37	22	3	136	5.0	0.93 [0.80, 0.98]	0.86 [0.80, 0.91]
Mach 1978	17	12	2	35	5.0	0.89 [0.67, 0.99]	0.74 [0.60, 0.86]
Mackay 1974	36	17	17	150	40.0	0.68 [0.54, 0.80]	0.90 [0.84, 0.94]
Mariani 1980	18	2	6	43	10.0	0.75 [0.53, 0.90]	0.96 [0.85, 0.99]
McCall 1994	57	15	41	198	5.0	0.58 [0.48, 0.68]	0.93 [0.89, 0.96]
Miles 1995	46	8	7	64	10.0	0.87 [0.75, 0.95]	0.89 [0.79, 0.95]
Minton 1985	105	6	25	264	2.5	0.81 [0.73, 0.87]	0.98 [0.95, 0.99]
Mittal 2011	29	9	10	25	5.0	0.74 [0.58, 0.87]	0.74 [0.56, 0.87]
Nishida 1988	13	7	4	42	2.5	0.76 [0.50, 0.93]	0.86 [0.73, 0.94]
Ochoa-Figueroa 2012	14	1	20	12	10.0	0.41 [0.25, 0.59]	0.92 [0.64, 1.00]
Ohlsson 1995	7	1	10	35		0.41 [0.18, 0.67]	0.97 [0.85, 1.00]
Ohtsuka 2008	18	9	4	66	5.0	0.82 [0.60, 0.95]	0.88 [0.78, 0.94]
Park 2009	222	69	69	382	7.0	0.76 [0.71, 0.81]	0.85 [0.81, 0.88]
Peng 2013	30	18	32	47	5.0	0.48 [0.35, 0.61]	0.72 [0.60, 0.83]
Seregni 1992	86	15	50	185		0.63 [0.55, 0.71]	0.93 [0.88, 0.96]
Staib 2000	44	4	14	36	3.0	0.76 [0.63, 0.86]	0.90 [0.76, 0.97]
Steele 1982	150	175	68	341	5.0	0.69 [0.62, 0.75]	0.66 [0.62, 0.70]
Tang 2009	56	14	20	215	5.0	0.74 [0.62, 0.83]	0.94 [0.90, 0.97]
Tate 1982	70	38	32	328	40.0	0.69 [0.59, 0.77]	0.90 [0.86, 0.93]
Tobaruela 1997	9	4	12	35	5.0	0.43 [0.22, 0.66]	0.90 [0.76, 0.97]
Triboulet 1983	38	6	5	42	20.0	0.88 [0.75, 0.96]	0.88 [0.75, 0.95]
Wang 1994	18	9	9	131	5.0	0.67 [0.46, 0.83]	0.94 [0.88, 0.97]
Wood 1980	36	3	1	108	25.0	0.97 [0.86, 1.00]	0.97 [0.92, 0.99]
Yakabe 2010	43	9	19	154	4.5	0.69 [0.56, 0.80]	0.94 [0.90, 0.97]
Yu 1992	41	7	25	109	15.0	0.62 [0.49, 0.74]	0.94 [0.88, 0.98]



For studies reporting accuracy at more than one threshold, 2 x 2 data at the threshold closest to 5 μg/L are included in the plot (5 μg/L was the most commonly reported threshold).
.Sensitivity ranged from 41% to 97% and specificity from 52% to 100%.
Figure 5 plots each of the 52 studies in ROC space. The size of each box is proportional to the inverse standard error for sensitivity and specificity for each study (a larger box indicates greater precision).

Figure 5. Scatter plot of sensitivity versus specificity for all 52 studies, regardless of threshold. Each box represents the 2 x 2 data extracted from each study, with the width of the boxes being proportional to the inverse standard error of the specificity and the height of the boxes proportional to the inverse standard error of the sensitivity.



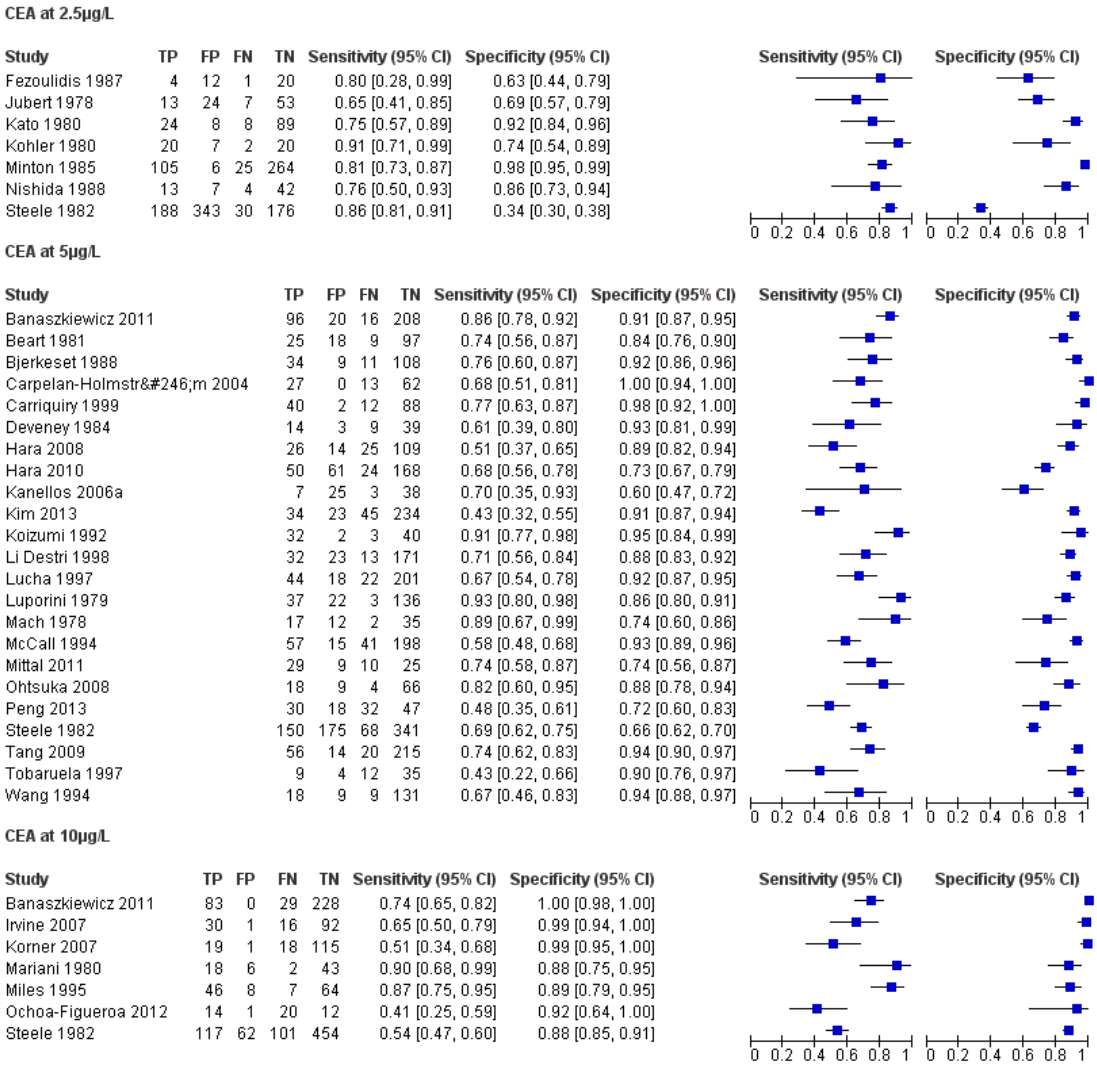
Effect of CEA threshold on diagnostic accuracy

Forty-one studies (79%) reported accuracy at just a single threshold. A wide range of thresholds were reported (2 to 40 µg/L). Four studies (8%) did not report which threshold they used (Graffner 1985; Johnson 1985; Ohlsson 1995; Seregni 1992).Seven studies (13%) reported 2 x 2 data for more than one threshold:

- Banaszkiewicz 2011: 5 and 10 µg/L
- Bjerkeset 1988: 3.5 and 5 µg/L
- Carlsson 1983: 3 and 7.5 µg/L
- Korner 2007: 4 and 10 µg/L
- Mittal 2011: 3, 5, 10, 20, and 50 µg/L
- Steele 1982: 2.5, 5, 10, and 20 µg/L
- Wood 1980: 25 and 40 µg/L

The forest plots in Figure 6 (Analysis 2) show the range of sensitivity and specificity for studies reporting the accuracy of CEA at cut-off values of 2.5, 5 and 10 µg/L.

Figure 6. Forest plot broken down by threshold: CEA at 2.5µg/L, CEA at 5µg/L, CEA at 10µg/L.TP = true positive; FP = false positive; FN = false negative; TN = true negativeThe blue square depicts the sensitivity and specificity for each study and the horizontal line represents the corresponding 95% confidence intervals for these estimates.



The summary ROC curves and the summary estimates including confidence ellipses for the threshold values of 2.5, 5, and 10 µg/L (Analyses 3, 4 and 5) can be found in Figure 7, Figure 8 and Figure 9 respectively.

Figure 7. Summary ROC plot of accuracy at a threshold of 2.5 µg/L.Each box represents the 2 x 2 data extracted from each study. The width of the box is proportional to the number of patients who did not experience recurrence in each study, and the height is proportional to the number of patients that did develop recurrent CRC.The filled circle is the pooled estimate for sensitivity and specificity and the line running through it is the summary ROC curve.The smaller dotted ellipse represents the 95% credible region around the summary estimate; the larger dashed ellipse represents the 95% prediction region.

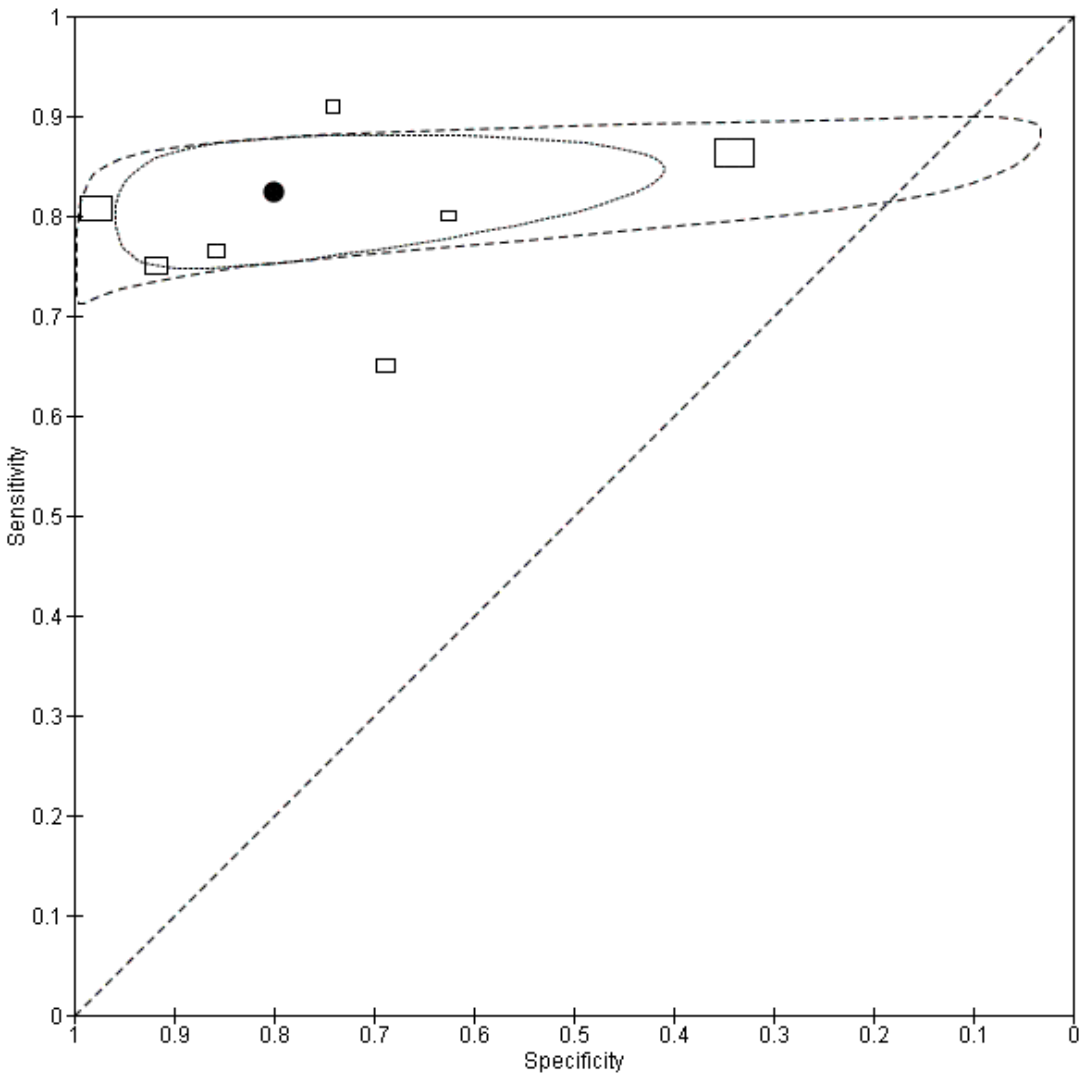


Figure 8. Summary ROC plot of accuracy at a threshold of 5 µg/L. Each box represents the 2 x 2 data extracted from each study. The width of the box is proportional to the number of patients who did not experience recurrence in each study, and the height is proportional to the number of patients that did develop recurrent CRC. The filled circle is the pooled estimate for sensitivity and specificity and the line running through it is the summary ROC curve. The smaller dotted ellipse represents the 95% credible region around the summary estimate; the larger dashed ellipse represents the 95% prediction region.

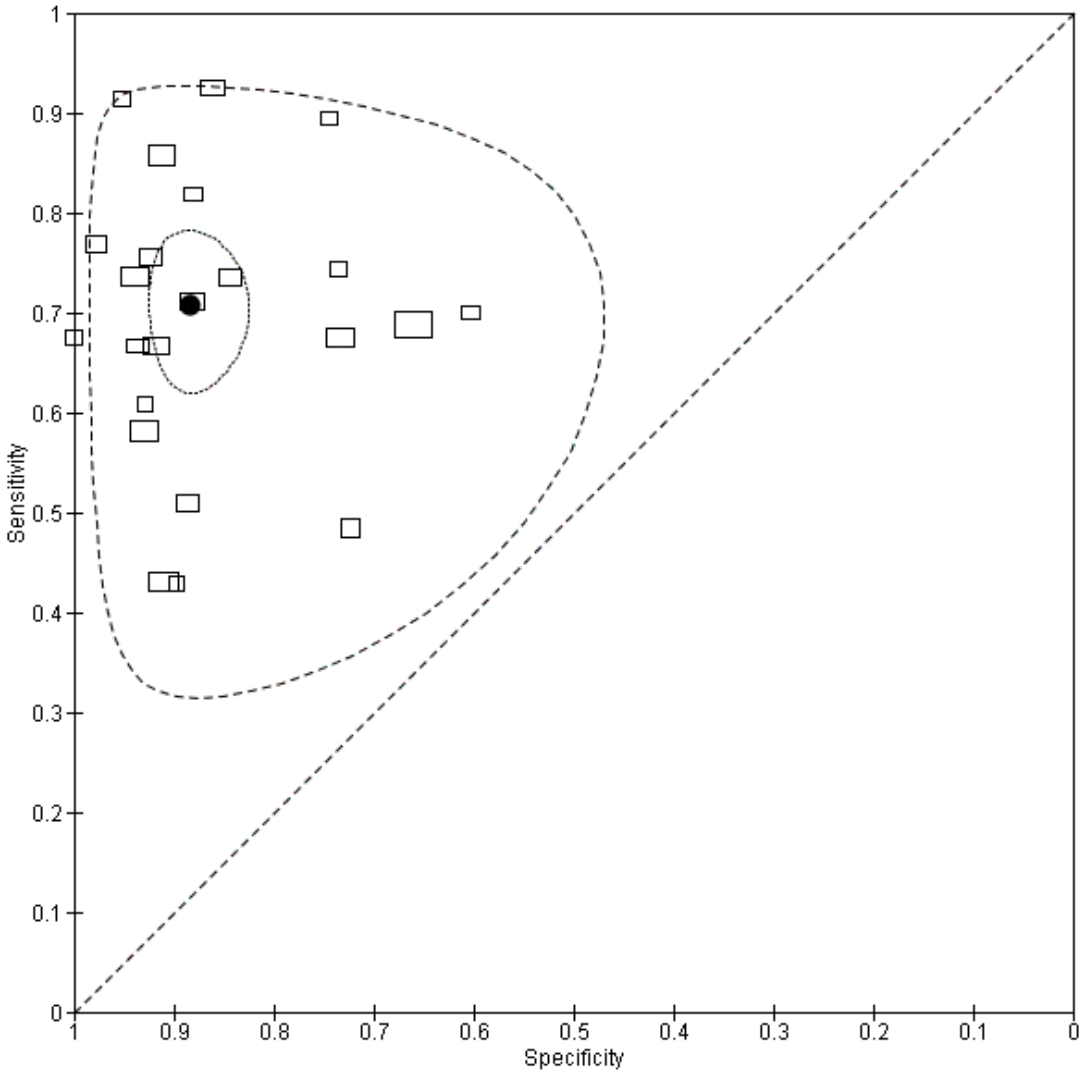
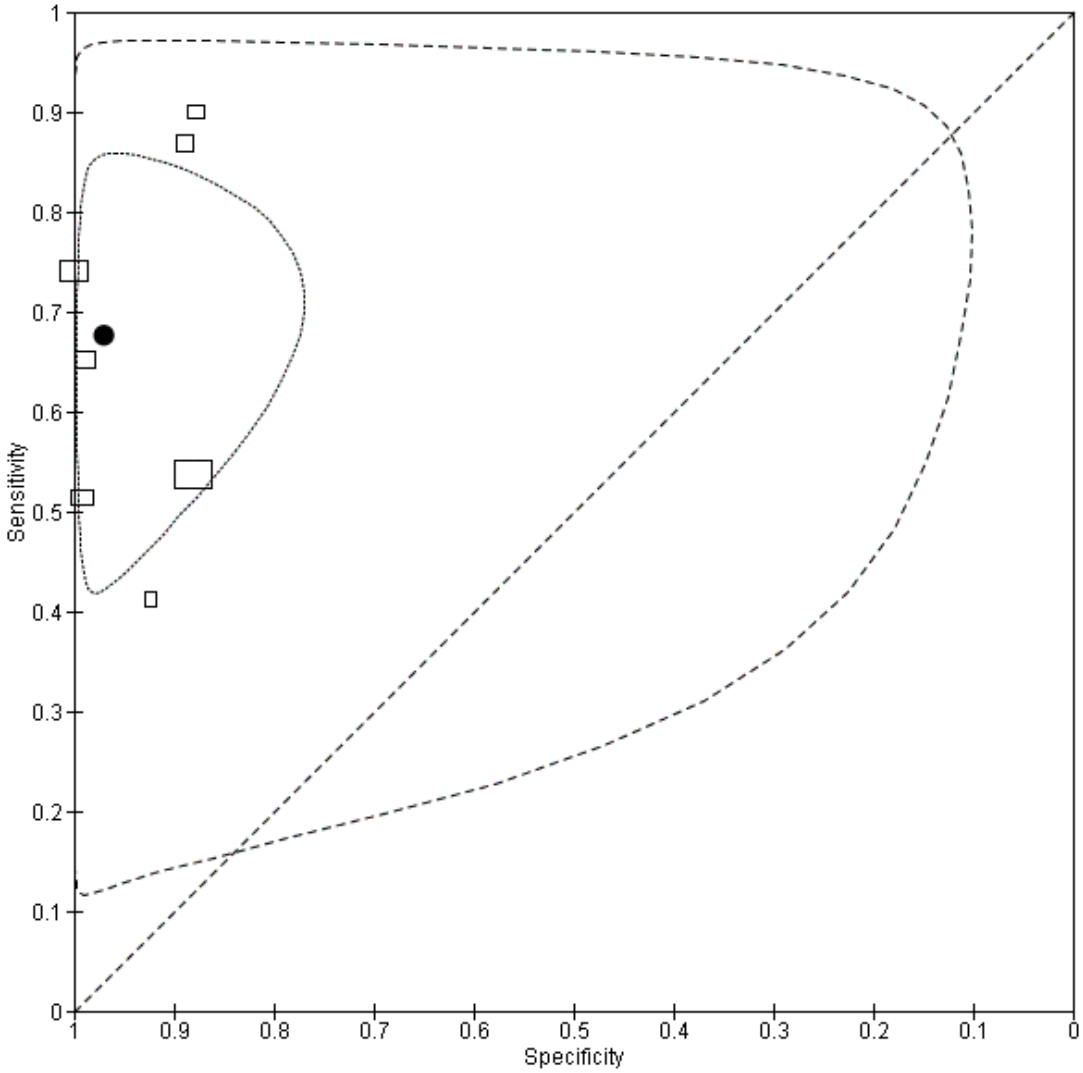


Figure 9. Summary ROC plot of accuracy at a threshold of 10 µg/L. Each box represents the 2 x 2 data extracted from each study. The width of the box is proportional to the number of patients who did not experience recurrence in each study, and the height is proportional to the number of patients that did develop recurrent CRC. The filled circle is the pooled estimate for sensitivity and specificity and the line running through it is the summary ROC curve. The smaller dotted ellipse represents the 95% credible region around the summary estimate; the larger dashed ellipse represents the 95% prediction region.



In the seven studies reporting a threshold of 2.5 µg/L, the sensitivity ranged from 65% to 91% and specificity from 34% to 98%. The pooled sensitivity of these studies was 82% (95% CI 78% to 86%) and pooled specificity 80% (95% CI 59% to 92%). Assuming that the proportion of patients with recurrence in any single testing period is 2% (based on our observed prevalence of recurrence of 30% and national guidance to conduct 14 to 15 CEA tests during follow-up), for every 1000 patients tested at a threshold of 2.5 µg/L, 16 cases of recurrence will be detected, four cases will be missed, and there will be 196 false alarms (people referred unnecessarily for further testing). More precise estimates of test performance using the incidence data reported by Sargent 2007 can be found in Summary of findings 2.

In the 23 studies which reported the impact of applying a threshold of 5 µg/L, sensitivity ranged from 43% to 93% and specificity from 60% to 100%. The pooled sensitivity of these studies was 71% (95% CI 64% to 76%) and pooled specificity 88% (95% CI 84% to 92%). For every 1000 patients tested at a threshold of 5 µg/L, 14 cases of recurrence will be detected, six cases will be missed, and there will be 118 false alarms. More precise estimates of test performance using the incidence data reported by Sargent 2007 can be found in Summary of findings 3

In the seven studies reporting the impact of applying a threshold of 10 µg/L, sensitivity ranged from 41% to 87% and specificity from 88% to 100%. The pooled sensitivity of these studies was 68% (95% CI 53% to 79%) and pooled specificity 97% (95% CI 90% to 99%). For every 1000 patients tested at a threshold of 10 µg/L, 14 cases of recurrence will be detected, seven cases will be missed, and there will be 29 false alarms. More precise estimates of test performance using the incidence data reported by Sargent 2007 can be found in Summary of findings 4.

Effect of the timing of CEA measurement

As previously described, we used two approaches when choosing which CEA measurement to include in the 2 x 2 tables. The first was to evaluate the CEA measurement taken closest to the time point at which recurrence was detected; the second was to look across all measurements to assess whether any had crossed the threshold during the entire follow-up period.

Including only those studies reporting accuracy at a threshold of 5 µg/L, we carried out a subgroup analysis for these two strategies. We adopted the first strategy in eight studies, for which the pooled sensitivity and specificity were 69.0% (95% CI 57.3% to 78.7%) and 90.0% (95% CI 77.8% to 95.9%) respectively. We adopted the second strategy in nine studies, for which the pooled sensitivity and specificity were 64.5% (95% CI 55.2% to 72.9%) and 89.5% (95% CI 83.4% to 93.5%) respectively.

Effect of laboratory technique

We were unable to carry out a subgroup analysis based on specific laboratory techniques, as reporting was so limited that it is was difficult to identify groups of studies where we could be confident that they had all used consistent methods.

For those studies reporting accuracy at a threshold of 5 µg/L, we carried out a subgroup analysis comparing the variability in accuracy before and after the introduction of the international reference preparation (IRP 73/601) calibration. We excluded one study (Li Destri 1998) from this analysis, as there was insufficient information about the timing of the sample analysis and laboratory technique. There were 11 studies predating the introduction of the IRP, providing a pooled sensitivity of 73.6% (95% CI 63.2% to 81.8%) and a pooled specificity of 88.5% (95% CI 83.2% to 92.2%), and 11 studies used methods which incorporated the IRP, resulting in a pooled sensitivity of 67.9% (95% CI 58.6% to 75.9%) and a pooled specificity of 88.6% (95% CI 80.0% to 93.7%). These results indicate no significant reduction in variability, and this was confirmed when we added it as a covariate in the metaregression (P = 0.958).

Effect of patient selection on diagnostic accuracy

When restricting the analyses to the 11 studies deemed to be at low risk of bias in the patient selection domain of the QUADAS-2 assessment, the sensitivity ranged from 43% to 93% and specificity from 61% to 99%.

We added the patient selection risk of bias item as an ordinal covariate (low risk = 6, unclear risk = 6 and high risk = 11) in the metaregression analysis for those studies reporting accuracy at 5 µg/L. The effect of this covariate was not significant (P = 0.771).

Effect of index test on diagnostic accuracy

There were no studies deemed to be at high risk of bias in the index test domain of the QUADAS-2 assessment. When restricting the analyses to the 37 studies (71%) deemed to be at low risk of bias in the index test domain of the QUADAS-2 assessment, the sensitivity ranged from 41% to 97% and specificity from 52% to 100%.

We added the index test risk of bias item as a covariate (low risk = 15, unclear risk = 8) in the metaregression analysis for those studies reporting accuracy at 5 µg/L. The effect of this covariate was not significant (P = 0.901).

Effect of the reference standard on diagnostic accuracy

There were also no studies deemed to be at high risk of bias in the reference standard domain of the QUADAS-2 assessment. When restricting the analyses to the 35 studies (67%) deemed to be at low risk of bias in the reference standard domain of the QUADAS-2 assessment, the sensitivity ranged from 41% to 97% and specificity from 52% to 100%.

We added the reference standard risk of bias item as a covariate (low risk = 17, unclear risk = 6) in the metaregression analysis for those studies reporting accuracy at 5 µg/L. The effect of this covariate was not significant (P = 0.292).

Effect of flow and timing on diagnostic accuracy

When restricting the analyses to the 25 studies (48%) deemed to be at low risk of bias in the flow and timing domain of the QUADAS-2 assessment, the sensitivity ranged from 41% to 95% and specificity from 52% to 100%.

We added the flow and timing risk of bias item as an ordinal covariate (low risk = 12, unclear risk = 6 and high risk = 5) in the metaregression analysis for those studies reporting accuracy at 5

µg/L. The effect of this covariate was not significant (P = 0.664).

Summary of findings

Review question: What is the accuracy of single-measurement blood CEA as a triage test to prompt further investigation for colorectal cancer recurrence after curative resection?				
Population: adults with no detectable residual disease after curative surgery (with or without adjuvant therapy)				
Studies: cross-sectional diagnostic test accuracy studies, cohort studies, and RCTs, reporting 2 x 2 data				
Index test: Blood carcino-embryonic antigen (CEA)				
Reference standard: appropriate[1] imaging, histology, or routine clinical follow-up				
Setting: primary or hospital care.				
Subgroup	**Number (Studies)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Interpretation Assuming a constant incidence of 2%[2] recurrence at each measurement point, testing 1000 people will have the following outcome depending on the CEA threshold applied**
2.5 µg/L	1515 (7)	82% (78 to 86)	80% (59 to 92)	16 cases of recurrence will be detected and 4 cases will be missed. 196 people will be referred unnecessarily for further testing
5 µg/L	4585 (23)	71% (64 to 76)	88% (84 to 92)	14 cases of recurrence will be detected and 6 cases will be missed. 118 people will be referred unnecessarily for further testing
10 µg/L	2341 (7)	68% (53 to 79)	97% (90 to 99)	14 cases of recurrence will be detected and 6 cases will be missed. 29 people will be referred unnecessarily for further testing

[1]as defined in the Reference standards section of the Methods.

[2]three-monthly prevalence is estimated as 2%, as the median prevalence amongst the included studies was 30% and a standard follow-up schedule will include 14 to 15 CEA tests over five years.

Month when CEA measured	per 1000 patients tested at a threshold of 2.5 µg/L					False alarm rate
	Estimated recurrences[1]	Referrals for raised CEA	Cases of recurrence detected	Cases of recurrence missed	False alarms (cases investigated when cancer not present)	
Follow-up years 1 and 2: 3-monthly CEA testing						
3	19	212	16	3	196	92%
6	19	212	16	3	196	92%
9	39	224	32	7	192	86%
12	39	224	32	7	192	86%
15	37	223	30	7	193	87%
18	37	223	30	7	193	87%
21	31	219	25	6	194	89%
24	31	219	25	6	194	89%
Follow-up years 3, 4 and 5: 6-monthly CEA testing						
30	46	229	38	8	191	83%
36	36	223	30	6	193	87%
42	27	217	22	5	195	90%
48	25	216	21	4	195	90%
54	17	211	14	3	197	93%
60	14	208	11	3	197	95%

[1]Estimates are based on data reported by Sargent 2007. Three-monthly data were unavailable, and so constant rates were assumed during each six-month period for the first two years. Estimates are rounded.

Month when CEA measured	per 1000 patients tested at a threshold of 5 µg/L					False alarm rate
	Estimated recurrences[1]	Referrals for raised CEA	Cases of recurrence detected	Cases of recurrence missed	False alarms (cases investigated when cancer not present)	
Follow-up years 1 and 2: 3-monthly CEA testing						
3	19	131	13	6	118	90%
6	19	131	13	6	118	90%
9	39	143	28	11	115	80%
12	39	143	28	11	115	80%
15	37	142	26	11	116	82%
18	37	142	26	11	116	82%
21	31	138	22	9	116	84%
24	31	138	22	9	116	84%
Follow-up years 3, 4 and 5: 6-monthly CEA testing						
30	46	147	33	13	114	78%
36	36	142	26	10	116	82%
42	27	136	19	8	117	86%
48	25	135	18	7	117	87%
54	17	130	12	5	118	91%
60	14	128	10	4	118	92%

[1]Estimates are based on data reported by Sargent 2007. Three-monthly data were unavailable, and so constant rates were assumed during each six-month period for the first two years. Estimates are rounded.

Month when CEA measured	per 1000 patients tested at a threshold of 10 µg/L					False alarm rate
	Estimated recurrences[1]	Referrals for raised CEA	Cases of recurrence detected	Cases of recurrence missed	False alarms (cases investigated when cancer not present)	
Follow-up years 1 and 2: 3-monthly CEA testing						
3	19	42	13	6	30	70%
6	19	42	13	6	29	70%
9	39	55	27	13	29	52%
12	39	55	27	13	29	52%
15	37	54	25	12	29	53%
18	37	54	25	12	29	53%
21	31	50	21	10	29	58%
24	31	50	21	10	29	58%
Follow-up years 3, 4 and 5: 6-monthly CEA testing						
30	46	60	31	15	29	48%
36	36	53	24	12	29	54%
42	27	48	19	9	29	61%
48	25	46	17	8	29	63%
54	17	41	11	6	30	72%
60	14	39	10	5	30	75%

[1]Estimates are based on data reported by Sargent 2007. Three-monthly data were unavailable, and so constant rates were assumed during each six-month period for the first two years. Estimates are rounded.

DISCUSSION

Summary of main results

We include 52 studies in the meta-analysis, covering 9717 patients (median sample size = 139, IQR: 72 - 247). The median proportion of recurrences in each study was 29% (IQR: 24% - 36%), agreeing with previously reported recurrence rates (Labianca 2010).

The diagnostic accuracy of CEA was reported at 15 different thresholds, ranging from 2 to 40 µg/L. Seven studies (13%) reported accuracy at a threshold of 2.5 µg/L, providing a pooled sensitivity of 82% (95% CI 78% to 86%) and a pooled specificity of 80% (95% CI 59% to 92%). The most commonly reported threshold was 5 µg/L (23 studies, 44%), providing a lower sensitivity of 71% (95% CI 64% to 76%) and an increased specificity of 88% (95% CI 84% to 92%). Seven studies (13%) reported accuracy at a threshold of 10 µg/L. Implementing such a high threshold reduced sensitivity to 68% (95% CI 53% to 79%), but provided high specificity of 97% (95% CI 90% to 99%).

Reporting quality was insufficient in important areas such as laboratory techniques. Insufficient detail about laboratory techniques and the frequent use of composite reference standards made it impossible to conduct desirable subgroup analyses. An individual-patient data meta-analysis would be required to fully explore the influence of factors such as preoperative CEA levels, chemotherapy, site of recurrence and smoking status, that are known to impact on CEA levels in follow-up.

Our results compared with other reports

Tan 2009 carried out a meta-analysis of 20 studies that reported the accuracy of CEA for the diagnosis of colorectal cancer recurrence using the Moses-Littenberg Method (Moses 1993). Their pooled estimate for specificity at a threshold of 5 µg/L was the same as ours (88%). Our pooled estimate for sensitivity was higher (71% versus 63%), but this difference is not statistically significant.

The method used by Tan 2009 to identify 2.2 µg/L as the 'optimum' CEA threshold was based on linear extrapolation (the lowest threshold included in their study was 3 µg/L). We instead implement bivariate meta-analyses (Reitsma 2005), as recommended in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). This method is statistically more rigorous than the method implemented in Tan 2009, and directly accounts for the within- and between-study variability in sensitivity and specificity.

We question the Tan 2009 recommendation of 2.2 µg/L (which was based on achieving high sensitivity) not just on the basis of the low specificity (and high false alarm rate), but also because there appears to be a 'ceiling' effect in terms of sensitivity - even at a threshold of 2.5 µg/L, around one in five cases of recurrence would be missed. The failure to exceed a sensitivity of about 80% even with a low threshold or poor specificity reflects the well-documented fact that some recurrent cancers are not associated with a rise in blood CEA levels.

Strengths and weaknesses of the review

Completeness

A key strength of this review is the comprehensiveness of our searches. We avoided the use of search filters and did not restrict our review to English-language publications. Two review authors screened all abstracts independently, with a third independently settling any disagreement over inclusion. We retrieved and analysed all full-text articles that we felt could be potentially relevant based on the title and abstract. We based additional searches on the citation of full-text articles to reduce the risk of missing relevant studies. Foreign-language articles were translated or assessed or both by colleagues of the authors proficient in the language in question.

It is not possible to estimate the impact of unpublished studies on our findings, as little is known about the mechanisms of publication bias for diagnostic accuracy studies (Allen 2013). Despite this, our included studies are likely to represent the vast majority of studies that provide evidence on this topic.

Two review authors then extracted data independently, and three authors independently performed QUADAS-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 DTA guidelines.

Variability

A major weakness of this review is that we considered many included studies to be at high risk of bias. There was also considerable between-study variation in the reporting of: 1) stage of primary disease included; 2) approach to ensuring no residual disease; 3) reporting of smoking; 4) reporting of chemotherapy treatment; and 5) the location of recurrence. All of these factors could plausibly have some influence on CEA levels, but corresponding 2 x 2 tables were not presented for these subgroups, and so it was not possible to adjust for this variation in our analyses.

The QUADAS-2 assessment of methodological quality highlighted the extent of the quality issues in the existing literature. Even the three studies that we assessed as having no risk of bias or applicability concerns were subject to considerable between-study heterogeneity: they each reported accuracy at different CEA thresholds, implemented different CEA laboratory techniques, and used differing composite reference standards to detect recurrence. The varying thresholds made it unfeasible to provide pooled diagnostic accuracy estimates for these high-quality studies.

Over half of the included studies (n = 27, 52%) were at high risk of selection bias, mainly due to inappropriate patient exclusions. We deemed a further 15 studies (29%) to be at unclear risk of bias for patient selection, due to poor reporting. This makes our accuracy estimates susceptible to selection bias, particularly if those excluded were at particularly high or low risk of recurrence. To

investigate this further, we removed those studies at high and unclear risk of bias for patient selection in a sensitivity analysis. The pooled estimates were not significantly different from the overall pooled results (sensitivity = 73%, 95% CI 64% to 80%; specificity = 87%, 95% CI 79% to 92%).

The methods used to measure CEA were also poorly reported: three studies (6%) did not report the CEA threshold used to determine a positive result, 15 studies (29%) did not report which laboratory technique had been used, and 43 studies (83%) failed to report any indicator of method accuracy or an estimate of CEA reproducibility. It is well known that variability exists between laboratory methods and between laboratories, and without this information it is impossible to adjust for any bias that has been introduced by the differences in method. The IRP calibration (73/601, introduced in 1992) attempts to reduce between-laboratory and between-technique variability, so we performed a sensitivity analysis leaving only the studies that were conducted after its introduction. We did not find the pooled accuracy estimates to be significantly different from the overall analysis (sensitivity = 67.9%, 95% CI 58.6% to 75.9%; specificity = 88.6%, 95% CI 80.0% to 93.7%).

A possible source of bias in this review is likely to be the methods used to implement the reference standard. In nine studies, the reference standard was only carried out if a rise in CEA was detected, possibly causing false-negative results to be misclassified as true-negative results. Furthermore, most studies implemented a composite reference standard, but failed to consistently reported which investigation (within the composite) actually diagnosed recurrence. In half of the studies (n = 26, 50%), positive results for certain reference tests triggered the use of other reference tests. These concerns over partial and differential verification were considered in the flow and timing domain of QUADAS-2, explaining why there were no studies deemed to be at high risk of bias in the reference standard domain.

The time between the CEA measurement and the reference test used in the 2 x 2 table was not reported in any of the studies. There is therefore a high chance of misclassification due to disease progression during the time between CEA and the reference test. Understanding this relationship is important in this setting as: a) a high-grade recurrence will progress more quickly than low-grade; b) this information is required to estimate lead time. Furthermore, no study reported 2 x 2 data for each three- to six-month period of follow-up, which would be desirable given that CRC recurrence is known to occur more commonly in the first two years of follow-up, suggesting that a variable threshold may have greater accuracy (Sargent 2007).

Applicability of findings to the review question

All of the studies identified were carried out in hospital outpatient clinics, except one that followed up patients in both primary and secondary care. As the patient population is so well defined in this review (postoperative curative colorectal cancer resection), it is unlikely that the actual clinical setting in which follow-up takes place would have any influence on the severity of disease seen or consequently on the accuracy of CEA.

Changing the setting of follow-up could affect the accuracy of the CEA measurement if transporting blood samples taken in a community setting are stored suboptimally and there are long delays in blood reaching the laboratory. But monitoring CEA in primary care is already common practice in many countries and these potential problems have been successfully addressed. Implementation of the reference standard might also vary if patients being followed up in hospital are more likely to be referred for further investigation for reasons other than a rise in CEA. However, the Australian multicentre RCT investigating GP versus surgical follow-up reported similar recurrence rates and times to detection, irrespective of place of follow-up (Wattchow 2006).

For these reasons, we regard the findings of this review as applicable to follow-up in the primary and specialist care setting.

To make sense of the meta-analysis results and calculate false-alarm rates, the pooled estimates of sensitivity and specificity need to be converted into predictive values, taking into account the incidence of disease in the relevant testing interval. In making this conversion, we assumed that sensitivity and specificity are constant during the follow-up period, which seems reasonable, as we are aware of no evidence that recurrences presenting at different time points have a different propensity to release CEA.

CEA is usually measured about 14 to 15 times during the five years following primary treatment (three-monthly for two years and then six-monthly) and so the crudest estimate of the number of recurrences potentially detectable in each testing interval is 2% (the median incidence of recurrence in the included studies of 30% divided by 15). However, in reality incidence is not constant at each testing point, but changes with time and follow-up interval. So, as some readers will wish to apply the findings of our review to a more precise estimate of incidence from actual clinical practice, we have reported estimates of test performance based on external data from Sargent 2007, which is the best data currently available on the incidence of recurrence at each point during follow-up.

AUTHORS' CONCLUSIONS

Implications for practice

The most important conclusion from this review is that CEA has inadequate sensitivity to be used as the sole method of detecting recurrence. Most national guidelines already recommend that it should be used in conjunction with another mode of diagnosis (such as CT imaging of the thorax, abdomen, and pelvis at 12 to 18 months) to pick up the remaining cases. Our review supports this recommendation. If CEA is used as the sole triage test, a significant

number of cases will be missed, whatever threshold is adopted for defining a positive test.

It is important to point out that this review provides no evidence to help choose which diagnostic modality to use for this supplementary testing, nor the frequency with which it should be undertaken. However, current recommendations are consistent with the results of the FACS trial which showed that regular CEA blood testing achieves similar diagnostic performance to regular CT imaging, if supplemented with a single CT scan at 12 to 18 months (Primrose 2014).

Supplementing CEA with another testing modality to improve sensitivity also makes it easier to adopt a threshold for defining a positive test which reduces the number of patients requiring further investigation with CT imaging or other more invasive investigations. This is important for minimising unnecessary anxiety and radiation hazard for patients. It is also important in health economies such as the NHS, because of the expense and limited capacity for investigations such as CT imaging and colonoscopy.

Current standard practice (based on national recommendations) is to apply a threshold 5 µg/L. At this threshold, assuming that the proportion of patients with recurrence in any single testing period is about 2% (based on our observed prevalence of recurrence of 30% and national guidance to conduct 14 to 15 CEA tests during follow-up), then there would be 118 false alarms and six missed cases for every 1000 patients tested. Increasing the threshold to 10 µg/L reduces the number of false alarms to 29 at a cost of six missed cases (Summary of findings 1). It is possible (although beyond the scope of this review to assess) that these missed cases may be avoided by the strategy of supplementary testing with another investigative modality as recommended above. For those interested in reviewing national recommendations on testing frequency, and the optimal threshold to apply at each time point (which need not necessarily be constant), we have included more precise estimates of test performance derived from incidence data reported by Sargent 2007 for the thresholds of 2.5 µg/L (Summary of findings 2), 5 µg/L (Summary of findings 3), and 10 µg/L (Summary of findings 4).

One potential solution to improve the diagnostic performance of CEA that is not addressed by this review is to treat CEA as a monitoring test rather than a one-off diagnostic test. Studies excluded from this review (Characteristics of excluded studies) for not being DTA studies have investigated the utility of: CEA frequency (Carl 1983), CEA slope (Staab 1985a), CEA doubling time (Ito 2002; Koga 1999) and a CEA nomogram (Minton 1978a; Minton 1978b; Minton 1989). The authors of the FACS trial have more recently pointed out that taking account of the change in CEA results over time and setting a threshold on the basis of the trend in CEA level could have substantially improved CEA performance, with an area under the ROC curve increasing from 0.74 to 0.90 (Shinkins 2014).

Implications for research

It is clear that measuring blood CEA has insufficient sensitivity to be used alone. Future research needs to explore the optimal timing and extent of supplementary CT imaging. It is also becoming clear that using one-off CEA measurements is suboptimal. An analysis of the benefits of making decisions to further investigate on the basis of trends over time needs to be done, and to be augmented by cost-benefit analysis of different strategies for the timing of monitoring tests and the optimal combination of CEA blood testing and CT imaging.

The other clear outcome from this review is the overall poor quality of reporting of diagnostic accuracy studies in this field. This poor reporting is compounded by the considerable between-study heterogeneity and limitations of study quality. In response to the methodological limitations highlighted in this review, authors of future research investigating the diagnostic accuracy of CEA for CRC recurrence should take care to clearly report: the CEA threshold and technique used, with an indication of method accuracy and of CEA reproducibility; the reference test used in any 2 x 2 table reported; 2 x 2 tables for each time point that the index test is measured; and the timing of the CEA test in relation to the index test (preferably as individual patient data).

The lack of significant improvement in diagnostic accuracy following sensitivity analysis using studies deemed to be at low risk of bias in the QUADAS-2 assessment also suggests that modifications to QUADAS-2 may be warranted in assessing the quality of diagnostic tests used for follow-up monitoring.

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, Dr Clare Davenport for her input into the development of theTitle Registration Form, Copy Edit Support (CES) and Henning Keinke Andersen (ME of the CCCG) for careful revision of the review.

REFERENCES

References to studies included in this review

Banaszkiewicz 2011 *{published data only}*
Banaszkiewicz Z, Jarmocik P, Frasz J, Tojek K, Mrozowski M, Jawien A. Usefulness of CEA concentration measurement and classic colonoscopy in follow-up after radical treatment of colorectal cancer. *Polski Przeglad Chirurgiczny* 2011 Jun; **83**(6):310–8.

Barillari 1992 *{published data only}*
Barillari P, Bolognese A, Chirletti P, Cardi M, Sammartino P, Stipa V. Role of CEA, TPA, and Ca 19-9 in the early detection of localized and diffuse recurrent rectal cancer. *Diseases of the Colon and Rectum* 1992 May;**35**(5):471–6.

Beart 1981 *{published data only}*
Beart RW Jr, Metzger PP, O'Connell MJ, Schutt AJ. Postoperative screening of patients with carcinoma of the colon. *Diseases of the Colon and Rectum* 1981;**24**(8):585–8.

Bjerkeset 1988 *{published data only}*
Bjerkeset T, Orjasaeter H, Soreide O. The role of carcinoembryonic antigen (CEA) in routine investigation and postsurgical monitoring in patients with colorectal cancer. *Surgical Research Communications* 1988;**2**:205–12.

Carlsson 1983 *{published data only}*
Carlsson U, Stewénius J, Ekelund G, Leandoer L, Nosslin B. Is CEA analysis of value in screening for recurrences after surgery for colorectal carcinoma?. *Diseases of the Colon and Rectum* 1983;**26**(6):369–73.

Carpelan-Holmström 2004 *{published data only}*
Carpelan-Holmström M, Louhimo J, Stenman UH, Alfthan H, Järvinen H, Haglund C. CEA, CA 242, CA 19-9, CA 72-4 and hCGbeta in the diagnosis of recurrent colorectal cancer. *Tumour Biology* 2004;**25**(5-6):228–34.

Carriquiry 1999 *{published data only}*
Carriquiry LA, Piñeyro A. Should carcinoembryonic antigen be used in the management of patients with colorectal cancer?. *Diseases of the Colon and Rectum* 1999; **42**(7):921–9.

Deveney 1984 *{published data only}*
Deveney KE, Way LW. Follow-up of patients with colorectal cancer. *American Journal of Surgery* 1984;**148**(6):717–22.

Engarås 2003 *{published data only}*
Engarås B. Individual cutoff levels of carcinoembryonic antigen and CA 242 indicate recurrence of colorectal cancer with high sensitivity. *Diseases of the Colon and Rectum* 2003; **46**(3):313–21.

Farinon 1980 *{published data only}*
Farinon AM, Sivelli R, Sianesi M, Zanella E. Carcinoembryonic antigen (CEA) test associated with colonoscopy: a monitoring method for the early detection of recurrent colorectal carcinoma. *Surgery in Italy* 1980;**20** (3):190–8.

Fezoulidis 1987 *{published data only}*
Fezoulidis I, Imhof H, Karner-Hanusch J, Teleky B, Wunderlich M, Schiessel R. The value of computed tomography after operation of carcinoma of the rectum. *Digitale Bilddiagnostik* 1987;**7**(4):194–8.

Fucini 1987 *{published data only}*
Fucini C, Tommasi SM, Rosi S, Malatantis G, Cardona G, Panichi S, et al. Follow-up of colorectal cancer resected for cure. An experience with CEA, TPA, Ca 19-9 analysis and second-look surgery. *Diseases of the Colon and Rectum* 1987; **30**(4):273–7.

Graffner 1985 *{published data only}*
Graffner H, Hultberg B, Johansson B, Möller T, Petersson BG. Detection of recurrent cancer of the colon and rectum. *Journal of Surgical Oncology* 1985;**28**(2):156–9.

Hara 2008 *{published data only}*
Hara M, Kanemitsu Y, Hirai T, Komori K, Kato T. Negative serum carcinoembryonic antigen has insufficient accuracy for excluding recurrence from patients withDukes C colorectal cancer: analysis with likelihood ratio and posttest probability in a follow-up study. *Diseases of the Colon and Rectum* 2008;**51**(11):1675–80.

Hara 2010 *{published data only}*
Hara M, Sato M, Takahashi H, Takayama S, Takeyama H. Does serum carcinoembryonic antigen elevation in patients with postoperative stage II colorectal cancer indicate recurrence. Comparison with stage III. *Journal of Surgical Oncology* 2010;**102**(2):154–7.

Hine 1984 *{published data only}*
Hine KR, Dykes PW. Serum CEA testing in the post-operative surveillance of colorectal carcinoma. *British Journal of Cancer* 1984;**49**(6):689–93.

Irvine 2007 *{published data only}*
Irvine T, Scott M, Clark CI. A small rise in CEA is sensitive for recurrence after surgery for colorectal cancer. *Colorectal Diseases* 2007;**9**(6):527–31.

Johnson 1985 *{published data only}*
Johnson JA, Giercksky KE. [Colorectal cancer recurrence. Early diagnosis by determination of a carcinoembryonic antigen in the serum]. *Tidsskrift for Den Norske Laegeforening* 1985;**105**(29):2044–6.

Jubert 1978 *{published data only}*
Jubert AV, Talbott TM, Maycroft TM. Characteristics of adenocarcinomas of the colorectum with low levels of preoperative plasma carcinoembryonic antigen (CEA). *Cancer* 1978;**42**(2):635–9.

Kanellos 2006a *{published data only}*
Kanellos I, Zacharakis E, Demetriades H, Christoforidis E, Kanellos D, Pramateftakis MG, et al. Value of carcinoembryonic antigen assay in predicting hepatic metastases, local recurrence, and survival after curative resection of colorectal cancer. *Surgery Today* 2006;**36**(10): 879–84.

Kato 1980 *{published data only}*
Kato K, Morimoto T, Kato T, Yasue M, Takagi H, Kito T, et al. CEA assays in postoperative detection of recurrent

colorectal carcinoma (author's transl). *Nihin Gan Chyryo Gakkai Shi* 1980;**15**(7):1137–42.

Kim 2013 ***{published data only}***
Kim HS, Lee MR. Diagnostic accuracy of elevated serum carcinoembryonic antigen for recurrence in postoperative stage II colorectal cancer patients: comparison with stage III. *Annals of Coloproctology* 2013;**29**(4):155–9.

Kohler 1980 ***{published data only}***
Kohler JP, Simonowitz D, Paloyan D. Preoperative CEA level: a prognostic test in patients with colorectal carcinoma. *The American Surgeon* 1980;**46**(8):449–52.

Koizumi 1992 ***{published data only}***
Koizumi F, Odagiri H, Fujimoto H, Kawamura T, Ishimori A. [Clinical evaluation of four tumor markers (CEA, TPA, CA50 and CA72-4) in colorectal cancer]. *Rinsho Byori - Japanese Journal of Clinical Pathology* 1992;**40**(5):523–8.

Korner 2007 ***{published data only}***
Körner H, Söreide K, Stokkeland PJ, Söreide JA. Diagnostic accuracy of serum-carcinoembryonic antigen in recurrent colorectal cancer: a receiver operating characteristic curve analysis. *Annals of Surgical Oncology* 2007 Feb;**14**(2): 417–23.

Li Destri 1998 ***{published data only}***
Li Destri G, Greco S, Rinzivillo C, Racalbuto A, Curreri R, Di Cataldo A. Monitoring Carcinoembryonic Antigen in Colorectal Cancer: Is it Still Useful?. *Surgery Today* 1998; **28**(12):1233–6.

Lucha 1997 ***{published data only}***
Lucha PA Jr, Rosen L, Olenwine JA, Reed JF 3rd, Riether RD, Stasik JJ Jr, et al. Value of carcinoembryonic antigen monitoring in curative surgery for recurrent colorectal carcinoma. *Diseases of the Colon and Rectum* 1997;**40**(2): 145–9.

Luporini 1979 ***{published data only}***
Luporini G, Mangiarotti F, Fraschini P, Labianca R, Tassi GC, De Barbieri A. [Importance of the analysis of the carcinoembryonic antigen in clinical oncology]. *Minerva Medica* 1979;**70**(2):127–34.

Mach 1978 ***{published data only}***
Mach JP, Vienny H, Jaeger P, Haldemann B, Egely R, Pettavel J. Long-term follow-up of colorectal carcinoma patients by repeated CEA radioimmunoassay. *Cancer* 1978; **42**(3 Suppl):1439–47.

Mackay 1974 ***{published data only}***
Mackay AM, Patel S, Carter S, Stevens U, Laurence DJ, Cooper EH, et al. Role of serial plasma C.E.A. assays in detection of recurrent and metastatic colorectal carcinomas. *British Medical Journal* 1974 Nov;**4**(5941):382–5.

Mariani 1980 ***{published data only}***
Mariani G, Carmellini M, Bonaguidi F, Benelli MA, Toni MG. Serum CEA Monitoring in the follow-up of Colorectal Cancer Patients with negative preoperative Serum CEA. *European Journal of Cancer* 1980;**16**(8):1099–103.

McCall 1994 ***{published data only}***
McCall JL, Black RB, Rich CA, Harvey JR, Baker RA, Watts JM, et al. The Value of serum carcinoembryonic antigen in predicting recurrent disease following curative resection of colorectal cancer. *Diseases of the Colon and Rectum* 1994;**37**(9):875–81.

Miles 1995 ***{published data only}***
Miles WF, Greig JD, Seth J, Sturgeon C, Nixon SJ. Raised carcinoembryonic antigen level as an indicator of recurrent disease in follow up of patients with colorectal cancer. *British Journal of General Practice* 1995;**45**(395):287–8.

Minton 1985 ***{published data only}***
Minton JP, Hoehn JL, Gerber DM, Horsley JS, Connolly DP, Salwan F, et al. Results of a 400-patient carcinoembryonic antigen second-look colorectal cancer study. *Cancer* 1985;**55**(6):1284–90.

Mittal 2011 ***{published data only}***
Mittal BR, Senthil R, Kashyap R, Bhattacharya A, Singh B, Kapoor R, et al. 18F-FDG PET-CT in evaluation of postoperative colorectal cancer patients with rising CEA level. *Nuclear Medicine Communications* 2011;**32**(9): 789–93.

Nishida 1988 ***{published data only}***
Nishida O, Shiroto H, Satoh N, Nakajima Y. Clnical evaluation of a combination assay of CEA, CA19-9 and TPA in patients with colorectal cancer. *Japanese Journal of Cancer Clinics* 1988;**34**(416):1096–1100.

Ochoa-Figueroa 2012 ***{published data only}***
Ochoa-Figueroa MA, Uña-Gorospe J, Allende-Riera A, Cárdenas-Negro JC, Muñoz-Iglesias J, Cabello-García D, et al. Utility of low dose (18)F-FDG PET-CT in patients with suspected colorectal carcinoma recurrence in conventional diagnostic methods. *Revista Espanola de Medicina Nuclear e Imagen Molecular* 2012;**31**(5):249–56.

Ohlsson 1995 ***{published data only}***
Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg KG. Follow-up after curative surgery for colorectal carcinoma. Randomized comparison with no follow-up. *Diseases of the Colon and Rectum* 1995;**38**(6):619–26.

Ohtsuka 2008 ***{published data only}***
Ohtsuka T, Nakafusa Y, Sato S, Kitajima Y, Tanaka M, Miyazaki K. Different roles of tumor marker monitoring after curative resections of gastric and colorectal cancers. *Digestive Diseases and Sciences* 2008;**53**(6):1537–43.

Park 2009 ***{published data only}***
Park IJ, Choi GS, Lim KH, Kang BM, Jun SH. Serum carcinoembryonic antigen monitoring after curative resection for colorectal cancer: clinical significance of the preoperative level. *Annals of Surgical Oncology* 2009;**16**(11): 3087–93.

Peng 2013 ***{published data only}***
Peng NJ, Hu C, King TM, Chiu YL, Wang JH, Liu RS. Detection of resectable recurrences in colorectal cancer patients with 2-[18F]fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography. *Cancer Biotherapy & Radiopharmaceuticals* 2013;**28**(6):479–87.

Seregni 1992 *{published data only}*
Seregni E, Bombardieri E, Bogni A, Crippa F, De Jager E, Buraggi GL. The role of serum carcinoembryonic antigen (CEA) in the management of patients with colorectal carcinoma: the experience of the Istituto Tumori of Milan. *International Journal of Biological Markers* 1992;**7**(3): 167–70.

Staib 2000 *{published data only}*
Staib L, Schirrmeister H, Reske SN, Beger HG. Is (18)F-fluorodeoxyglucose positron emission tomography in recurrent colorectal cancer a contribution to surgical decision making?. *American Journal of Surgery* 2000;**180**(1): 1–5.

Steele 1982 *{published data only}*
Steele G Jr, Ellenberg S, Ramming K, O'Connell M, Moertel C, Lessner H, et al. CEA monitoring among patients in multi-institutional adjuvant G.I. therapy protocols. *Annals of Surgery* 1982;**196**(2):162–9.

Tang 2009 *{published data only}*
Tang R, Yeh CY, Wang JY, Changchien CR, Chen JS, Hsieh LL. Serum p53 antibody as tumor marker for follow-up of colorectal cancer after curative resection. *Annals of Surgical Oncology* 2009;**16**(9):2516–23.

Tate 1982 *{published data only}*
Tate H. Plasma CEA in the post-surgical monitoring of colorectal carcinoma. *British Journal of Cancer* 1982;**46**(3): 323–30.

Tobaruela 1997 *{published data only}*
Tobaruela E, Enriquez JM, Diez M, Camunas J, Muguerza J, Granell J. Evaluation of serum carcinoembryonic antigen monitoring in the follow-up of colorectal cancer patients with metastatic lymph nodes and a normal preoperative serum level. *International Journal of Biological Markers* 1997;**12**(1):18–21.

Triboulet 1983 *{published data only}*
Triboulet JP, Dessaint JP, Lagache G. [Diagnostic value of the assay of beta 2 microglobulin during the monitoring of surgically-treated colorectal cancers. Comparison with carcinoembryonic antigen]. *Gastroenterologie Clinique et Biologique* 1983;7(10):808–11.

Wang 1994 *{published data only}*
Wang JY, Tang R, Chiang JM. Value of carcinoembryonic antigen in the management of colorectal cancer. *Diseases of the Colon and Rectum* 1994;**37**(3):272–7.

Wood 1980 *{published data only}*
Wood CB, Ratcliffe JG, Burt RW, Malcolm AJ, Blumgart LH. The clinical significance of the pattern of elevated serum carcinoembryonic antigen (CEA) levels in recurrent colorectal cancer. *British Journal of Surgery* 1980;**67**(1): 46–8.

Yakabe 2010 *{published data only}*
Yakabe T, Nakafusa Y, Sumi K, Miyoshi A, Kitajima Y, Sato S, et al. Clinical significance of CEA and CA19-9 in postoperative follow-up of colorectal cancer. *Annals of Surgical Oncology* 2010;**17**(9):2349–56.

Yu 1992 *{published data only}*
Yu BM. Evaluation of combined CA-19-9 and CEA assay in monitoring recurrences and metastases of colorectal cancer. *Zhonghua Wai Ke Za Zhi* 1992;**30**(12):707–9.

References to studies excluded from this review

Afsaneh 2012 *{published data only}*
Motamed-Khorasani A, Small-Howard A, Etemadi H, Beart H. A new strategy for the early detection of colorectal cancer recurrence. Gastrointestinal Cancers Symposium (ASCO). San Francisco, CA, USA, January 19 - 21. 2012.

Ahmed 2013 *{published data only}*
Ahmed H, Bashir H, Khalid Nawaz M, Faruqui Z, Saeed Kazmi A, Ali Syed A, et al. Early detection of recurrence of colorectal carcinoma on F18-FDG PET-CT and its correlation with other clinicopathological parameters. *Nuclear Medicine Communications* 2013;**34**(4):394–5.

Aitkin 2012 *{published data only}*
Aitken K, Barbachano Y, Sharma B, Cunningham D, Cook G, Rao S. Elevated CEA level in the asymptomatic patient with normal conventional imaging: How useful is PET-CT for the detection of colorectal cancer recurrence? Gastrointestinal Cancers Symposium, San Francisco, CA, USA. Journal of Clinical Oncology. 2012.

Amin 2012 *{published data only}*
Amin A, Reddy A, Wilson R, Jha M, Miranda S, Amin J. Unnecessary surgery can be avoided by judicious use of PET/CT scanning in colorectal cancer patients. *Journal of Gastrointestinal Cancer* 2012;**43**(4):594–8.

Arnaud 1979 *{published data only}*
Arnaud JP, Simon P, Ollier JC, Koehl C, Adloff M. [Interest and limits of estimation of the carcinoembryonic antigen in colonic and rectal (author's transl)]. *Journal de Chirurgie* 1979;**116**(11):633–6.

Arnaud 1997 *{published data only}*
Arnaud JP, Cervi C, Bergamaschi R, Tuech JJ. [Value of oncologic follow-up of patients operated for colorectal cancer. A prospective study of 1000 patients]. *Journal de Chirurgie* 1997;**134**(2):45–50.

Arriola 2006 *{published data only}*
Arriola E, Navarro M, Parés D, Muñoz 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 the Colon and Rectum* 2006;**49**(4):478–84.

Auer 1977 *{published data only}*
Auer IO, Schmid L, Hoecht B. The clinical value of the plasma CEA level for postoperative detection of recurrence and metastases of carcinoma of the gastrointestinal tract [[German] Klinische Wertigkeit Der Plasmakonzentration Des Karzinoembryonalen Antigens Bei Der Postoperativen Verlaufskontrolle Von Karzinomen Des Verdauungstraktes]. *Medizinische Klinik* 1977;**72**(21):934–41.

Bakalakos 1999 *{published data only}*
Bakalakos EA, Burak WE, Young DC, Martin EW. Is carcino-embryonic antigen useful in the follow-up

management of patients with colorectal liver metastases?. *American Journal of Surgery* 1999;**177**(1):2–6.

Barrillari 1996 ***{published data only}***
Barillari P, Ramacciato G, Manetti G, Bovino A, Sammartino P, Stipa V. Surveillance of colorectal cancer: effectiveness of early detection of intraluminal recurrences on prognosis and survival of patients treated for cure. *Diseases of the Colon and Rectum* 1996;**39**(4):388–93.

Beatty 1979 ***{published data only}***
Beatty J, Romero C, Brown PW, Lawrence W, Jr, Terz JJ. Clinical value of carcinoembryonic antigen: diagnosis, prognosis, and follow-up of patients with cancer. *Archives of Surgery* 1979;**114**(5):563–7.

Beets 1994 ***{published data only}***
Beets G, Penninckx F, Schiepers C, Filez L, Mortelmans L, Kerremans R, et al. Clinical value of whole-body positron emission tomography with [18F] fluorodeoxyglucose in recurrent colorectal cancer. *British Journal of Surgery* 1994; **81**(11):1666–70.

Bhatavedekar 1992 ***{published data only}***
Bhatavdekar JM, Patel DD, Giri DD, Karelia NH, Vora HH, Ghosh N, et al. Comparison of plasma prolactin and CEA in monitoring patients with adenocarcinoma of colon and rectum. *British Journal of Cancer* 1992;**66**(5):977–80.

Bivins 1974 ***{published data only}***
Bivins BA, Meeker WR Jr, Griffen WO Jr. CEA levels and prognosis in colon carcinoma. *Journal of Surgical Oncology* 1974;**6**(5):413–21.

Boey 1984 ***{published data only}***
Boey J, Cheung HC, Lai CK, Wong J. A prospective evaluation of serum carcinoembryonic antigen (CEA) levels in the management of colorectal carcinoma. *World Journal of Surgery* 1984;**8**(3):279–86.

Borie 2004 ***{published data only}***
Borie F, Daurès JP, Millat B, Trétarre B. Cost and effectiveness of follow-up examinations in patients with colorectal cancer resected for cure in a French population-based study. *Journal of Gastrointestinal Surgery* 2004;**8**(5): 552–8.

Brummendorf 1985 ***{published data only}***
Brummendorf T, Anderer FA, Staab HJ, Hornung A, Kieninger G. [Carcinoembryonic antigen: diagnosis and tumor progression in gastrointestinal tumors]. *Deutsche Medizinische Wochenschrift* 1985;**110**(51-52):1963–8.

Brummendorf 1986 ***{published data only}***
Brummendorf T, Anderer FA, Staab HJ, Hornung A, Kieninger G. [The doubling time of circulating CEA as an individual prognostic criterion of recurrence in patients with gastrointestinal cancers]. *Klinische Wochenschrift* 1986; **64**(2):63–9.

Bucci 1994 ***{published data only}***
Bucci L, Benassai G, Santoro GA. Second look in colorectal surgery. *Diseases of the Colon and Rectum* 1994;**37**(2 Suppl): S123–6.

Camunas 1991 ***{published data only}***
Camunas J, Enriquez JM, Devesa JM, Morales V, Millan I. Value of follow-up in the management of recurrent colorectal cancer. *European Journal of Surgical Oncology* 1991;**17**(5):530–5.

Cangemi 1984 ***{published data only}***
Cangemi V, Fiori E, Santeusanio G. The clinical usefulness of post-operative monitoring of plasmatic CEA in colorectal cancer [[Italian] Utilita clinica del monitoraggio post–operatorio del cea ematico nei cancri del colon e del retto]. *Giornale di Chirurgia* 1984;**5**(1):83–8.

Cangemi 1987 ***{published data only}***
Cangemi V, Volpino P, Fiori E. The role of tumour markers (CEA, TPA, CA 19-9) in colon and rectum carcinomas. *Journal of Nuclear Medicine and Allied Sciences* 1987;**31**(2): 189–93.

Carl 1983 ***{published data only}***
Carl J, Bentzen SM, Norgaard-Pedersen B, Kronborg O. Modelling of serial carcinoembryonic antigen changes in colorectal cancer. *Scandinavian Journal of Clinical & Laboratory Investigation* 1993;**53**(7):751–5.

Carpelan-Holmström 1996 ***{published data only}***
Carpelan-Holmström MA, Haglund CH, Järvinen HJ, Roberts PJ. Serum CA 242 and CEA detect different patients with recurrent colorectal cancer. *Anticancer Research* 1996;**16**(2):981–6.

Castells 1998 ***{published data only}***
Castells A, Bessa X, Daniels M, Ascaso C, Lacy AM, García-Valdecasas JC, et al. Value of postoperative surveillance after radical surgery for colorectal cancer: results of a cohort study. *Diseases of the Colon and Rectum* 1998;**41**(6):714–23.

Catania 1981 ***{published data only}***
Catania G, Basile F, Cardi F, Azzarello G, Mazzarino C, Campo M, et al. [Value of carcinoembryonic antigen (CEA) in the diagnosis and postoperative monitoring of patients with colorectal cancer]. *Minerva Chirurgica* 1981; **36**(9):569–80.

Chang 2012 ***{published data only}***
Chang AC, Warrena LR, Barretoa SG, Williams R. Differing serum CEA in primary and recurrent rectal cancer - a reflection of histology?. *World Journal of Oncology* 2012; **3**(2):59–63.

Chapman 1998 ***{published data only}***
Chapman MA, Buckley D, Henson DB, Armitage NC. Preoperative carcinoembryonic antigen is related to tumour stage and long-term survival in colorectal cancer. *British Journal of Cancer* 1998;**78**(10):1346–9.

Chen 2010 ***{published data only}***
Chen CH, Hsieh MC, Lai CC, Yeh CY, Chen JS, Hsieh PS, et al. Lead time of carcinoembryonic antigen elevation in the postoperative follow-up of colorectal cancer did not affect the survival rate after recurrence. *International Journal of Colorectal Disease* 2010;**25**(5):567–71.

Cho 2007 ***{published data only}***
Cho YB, Chun HK, Yun HR, Lee WS, Yun SH, Lee WY. Clinical and pathologic evaluation of patients with

recurrence of colorectal cancer five or more years after curative resection. *Diseases of the Colon and Rectum* 2007;**50** (8):1204–10.

Choi 1997 ***{published data only}***
Choi JS, Min JS. Significance of postoperative serum level of carcinoembryonic antigen (CEA) and actual half life of CEA in colorectal cancer patients.. *Yonsei Medical Journal* 1997 Feb;**38**(1):1–7.

Colombo 1986 ***{published data only}***
Colombo PL, Lovotti D, Franco F, Coronelli M, Bonacasa R. [Value of CEA in the early diagnosis of colorectal cancer recurrence in the light of second-look results]. *Minerva Medica* 1986;**77**(13):495–8.

Cossu 1984 ***{published data only}***
Cossu F, Fodde M, Ledda P, Pisano G, Rombi GP, Scintu F, et al. [Prognostic value of CEA in postoperative monitoring of colorectal cancers]. *Minerva Medica* 1984;**75**(40): 2373–80.

Dalton 2010 ***{published data only}***
Dalton RSJ, Burn PR, Eyrebrook I. CT surveillance to detect colorectal cancer recurrence is the best option for possible curative resection. *Colorectal Disease* 2010;**12** (Suppl I):1–13.

Dash 2012 ***{published data only}***
S. Dash, A. Gupta. Clinical utility of FDG PET-CT in surveillance of patients with colorectal malignancy when serum CEA level is within normal range - A prospective study. *European Journal of Nuclear Medical Molecular Imaging* 2012;**39**(Suppl 2):S581.

De Brauw 1987 ***{published data only}***
De Brauw LM, Van de Velde CJH, Albers GHR, Zwaveling A. The value of follow-up after surgery for carcinoma of the colon. *Nederlands Tijdschrift voor Geneeskunde* 1987;**131**: 496–500.

De Levin 1982 ***{published data only}***
De Levin RW, Levin E. Correlation of single and serial CEA determinations with the clinical evolution of cancer patients. *Archiv für Geschwulstforschung* 1982;**52**(2):105–12.

De Salvo 1997 ***{published data only}***
De Salvo L, Razzetta F, Arezzo A, Tassone U, Bogliolo G, Bruzzone D, et al. Surveillance after colorectal cancer surgery. *European Journal of Surgical Oncology.* 1997;**23**(6): 522–5.

Dhar 1972 ***{published data only}***
Dhar P, Moore T, Zamcheck N, Kupchik HZ. Carcinoembryonic antigen (CEA) in colonic cancer. Use in preoperative and postoperative diagnosis and prognosis. *JAMA* 1972;**221**(1):31–5.

Di Cristofaro 2012 ***{published data only}***
Di Cristofaro L, Scarpa M, Angriman I, Perissinotto E, Ruffolo C, Frego M, et al. Cost-effectiveness analysis of postoperative surveillance protocols following radical surgery for colorectal cancer. *Acta Chirurgica Belgica* 2012; **112**(1):24–32.

Engarås 2001 ***{published data only}***
Engarås B, Kewenter J, Nilsson O, Wedel H, Hafström L. CEA, CA 50 and CA 242 in patients surviving colorectal cancer without recurrent disease. *European Journal of Surgical Oncology* 2001;**27**(1):43–8.

Farquharson 2012 ***{published data only}***
Farquharson AL, Genever AV, Belfield J, Hersey N, Amin SN, De Noronha R. PET-CT scan is a specific test for the detection of recurrent colorectal cancer but has limitations for patients with mucinous tumours. *Colorectal Disease* 2012;**14**(Suppl I):1–11.

Fernandes 2006 ***{published data only}***
Fernandes LC, Kim SB, Saad SS, Matos D. Value of carcinoembryonic antigen and cytokeratins for the detection of recurrent disease following curative resection of colorectal cancer. *World Journal of Gastroenterololgy* 2006;**12**(24): 3891–4.

Filella 1994 ***{published data only}***
Filella X, Molina R. Use of CA 19-9 in the early detection of recurrences in colorectal cancer: comparison with CEA. *Tumour Biology* 1994;**15**(1):1–6.

Filiz 2009 ***{published data only}***
Filiz AI, Sucullu I, Kurt Y, Karakas DO, Gulec B, Akin ML. Persistent high postoperative carcinoembryonic antigen in colorectal cancer patients--is it important?. *Clinics (Sao Paulo)* 2009;**64**(4):287–94.

Finlay 1983 ***{published data only}***
Finlay IG, McArdle CS. Role of carcinoembryonic antigen in detection of asymptomatic disseminated disease in colorectal carcinoma. *British Medical Journal (Clinical Research Edition)* 1983;**286**(6373):1242–4.

Fiocchi 2011 ***{published data only}***
Fiocchi F, Iotti V, Ligabue G, Malavasi N, Luppi G, Bagni B, et al. Role of carcinoembryonic antigen, magnetic resonance imaging, and positron emission tomography-computed tomography in the evaluation of patients with suspected local recurrence of colorectal cancer. *Clinical Imaging* 2011;**35**(4):266–73.

Florio 1988 ***{published data only}***
Florio MG, Artemisia A, Giorgianni G, Cogliandolo A, Giacobbe G, Manganaro T, et al. Recurrences in patients operated on by colorectal cancer. Diagnostic reliability of two tumour markers. *Chirurgia Gastroenterologica* 1988;**22** (4):475–7.

Fora 2012 ***{published data only}***
Fora AA, Patta AM, Attwood K, Wilding GE, Fakih M. Intensive radiographic and CEA screening and salvage resection in patients with stage II and III colorectal cancer. *Journal of Clinical Oncology* 2012;**30**(Supplement 4): abstract 405.

Forones 1997 ***{published data only}***
Forones NM, Tanaka M, Falcão JB. CEA as a prognostic index in colorectal cancer. *Sao Paulo Medical Journal* 1997 Nov–Dec;**115**(6):1589–92.

Forones 1998 *{published data only}*
Forones NM, Tanaka M, Machado D. Increased carcinoembryonic antigen and absence of recurrence in monitoring colorectal cancer. *Arquivos de Gastroenterologia* 1998;**35**(2):100–3.

Fortner 1988 *{published data only}*
Fortner JG. Recurrence of colorectal cancer after hepatic resection. *America Journal of Surgery* 1988;**155**(3):378–82.

Fournier 1999 *{published data only}*
Fournier RS, Kilroy K, Alavi A. Correlation between FDG PET imaging results and plasma CEA levels in patients with suspected recurrent colon carcinoma. *Journal of Nuclear Medicine* 1999;**40**(5):241P.

Fucini 1983 *{published data only}*
Fucini C, Tommasi MS, Cardona G, Malatantis G, Panichi S, Bettini U. Limitations of CEA monitoring as a guide to second-look surgery in colorectal cancer follow-up. *Tumori* 1983;**69**(4):359–64.

Fucini 1984 *{published data only}*
Fucini C, Malatantis G, D'Elia M, Tommasi MS. CEA monitoring and recurrences in the active follow-up of rectal cancer operated for cure [[Italian] Monitoraggio del cea e recidive nel follow–up attivo del cancro del retto operato radicalmente]. *Rivista Italiana di Colon-Proctologia* 1984;**3**(1):37–45.

Fucini 1985 *{published data only}*
Fucini C, Rosi S, Herd-Smith A. Value and limitations of intensive follow-up after radical surgery for rectocolonic cancer [[Italian] Significato e limiti del follow up intensivo dopo intervento radicale per cancro del colon–retto]. *Minerva Chirurgica* 1985;**40**(11):783–6.

Gail 1981 *{published data only}*
Gail MH. Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics* 1981;**37**(1):67–78.

Gajdukevich 2010 *{published data only}*
Gajdukevich IV, Kitaev AB, Sharapov GN, Byhovets IV, Turlaj DM. The value of biomolecular tumour markers in abdominal cancer recurrence. *Techniques in Coloproctology* 2010;**14**(1):75–6.

Gaudagni 1999 *{published data only}*
Gaudagni F. The clinical utility of serum tumor markers in the management of gastrointestinal cancer patients. *Journal of Clinical Ligand Assay* 1999;**22**(4):364–6.

Graham 1998 *{published data only}*
Graham RA, Wang S, Catalano PJ, Haller DG. Postsurgical surveillance of colon cancer: preliminary cost analysis of physician examination, carcinoembryonic antigen testing, chest x-ray, and colonoscopy. *Annals of Surgery* 1998 Jul;**228**(1):59–63.

Gray 1981 *{published data only}*
Gray BN, Walker C, Barnard R. Value of serial carcinoembryonic antigen determinations for early detection of recurrent cancer. *Medical Journal of Australia* 1981;**1**(4):177–8.

Griesenberg 1999 *{published data only}*
Griesenberg D, Nurnberg R, Bahlo M, Klapdor R. CEA, TPS, CA 19-9 and CA 72-4 and the fecal occult blood test in the preoperative diagnosis and follow-up after resective surgery of colorectal cancer. *Anticancer Research* 1999;**19**(4A):2443–50.

Grossetti 1981 *{published data only}*
Grossetti D, De Certaines J, Trebuchet G. The value and limitations of carcinoembryonic antigen levels (CEA) in colorectal cancers [[French] Valeur et limites du dosage de l'antigene carcino–embryonnaire (Ace) dans les cancers colo–rectaux]. *Annales de Chirurgie* 1981;**35**(19):875–7.

Grossmann 2007 *{published data only}*
Grossmann I, De Bock GH, Meershoek-Klein Kranenbarg WM, Van de Velde CJ, Wiggers T. Carcinoembryonic antigen (CEA) measurement during follow-up for rectal carcinoma is useful even if normal levels exist before surgery. A retrospective study of CEA values in the TME trial. *European Journal of Surgical Oncology* 2007;**33**(2):183–7.

Haga 1990 *{published data only}*
Haga S, Takahashi N, Kato H, Mori M, Umeda H, Azuhata H, et al. Study of serum CEA as an index for forecasting recurrence of colorectal carcinoma. *Journal of Tokyo Women's Medical College* 1990;**60**(2):163–6.

Hall 1994 *{published data only}*
Hall NR, Finan PJ, Stephenson BM, Purves DA, Cooper EH. The role of CA-242 and CEA in surveillance following curative resection for colorectal cancer. *British Journal of Cancer* 1994 Sep;**70**(3):549–53.

Hara 2011 *{published data only}*
Hara M, Sato M, Takahashi H, Takayama S, Takeyama H. Accuracy of monitoring serum carcinoembryonic antigen levels in postoperative stage III colorectal cancer patients is limited to only the first postoperative year. *Surgery Today* 2011 Oct;**41**(10):1357–62.

Herrera 1976 *{published data only}*
Herrera MA, Chu TM, Holyoke ED. Carcinoembryonic antigen (CEA) as a prognostic and monitoring test in clinically complete resection of colorectal carcinoma. *Annals of Surgery* 1976;**183**(1):5–9.

Hida 1996 *{published data only}*
Hida J, Yasutomi M, Shindoh K, Kitaoka M, Fujimoto K, Ieda S, et al. Second-look operation for recurrent colorectal cancer based on carcinoembryonic antigen and imaging techniques. *Diseases of the Colon and Rectum* 1996;**39**(1):74–9.

Hohenberger 1994 *{published data only}*
Hohenberger P, Schlag PM, Gerneth T, Herfarth C. Pre- and postoperative carcinoembryonic antigen determinations in hepatic resection for colorectal metastases. Predictive value and implications for adjuvant treatment based on multivariate analysis. *Annals of Surgery* 1994;**219**(2):135–43.

Holt 2010 *{published data only}*
Holt A, Nelson RA, Lai L. Surveillance with serial serum carcinoembryonic levels detect colorectal cancer recurrences

in patients who are initial nonsecretors. *American Surgeon* 2010;**76**(10):1100–3.

Holubec 2000 *{published data only}*
Holubec L Jr, Pikner R, Topolcan O, Finek J, Holubec L Sen, Pecen L. The usefulness of tumor markers in patients with colorectal carcinoma for the detection of local recurrences and distant metastases. *Coloproctology* 2001;**23**: 26–31.

Holyoke 2975 *{published data only}*
Holyoke ED, Chu TM, Murphy GP. CEA as a monitor of gastrointestinal malignancy. *Cancer* 1975;**35**(3):830–6.

Houlbec 2001 *{published data only}*
Holubec Jr L, Pikner R, Topolcan O, Finek J, Holubec Sen L, Pecen L. The usefulness of tumor markers in patients with colorectal carcinoma for the detection of local recurrences and distant metastases. *Coloproctology* 2001;**23**(1):26–31.

Humphreys 2011 *{published data only}*
Humphreys A, Cornish J, Stevenson J, Corr C, Billings P, Chandran P. Is a normal CEA false reassurance following curative resection for non metastatic colorectal cancer?. Colorectal Disease. Copenhagen: 6th Scientific and Annual Meeting of the European Society of Coloproctology, 2011; Vol. 13:36.

Huyghe 1983 *{published data only}*
Huyghe J. CEA radioimmunoassay. Clinical applications in colorectal cancer. *Acta Chirurgica Belgica* 1983;**83**(2): 77–88.

Iarumov 1998 *{published data only}*
Iarumov N, Ignatov A, Viiachki I. [The pre- and postoperative monitoring of the immunological indices and tumor markers in colorectal carcinoma]. *Khirurgiia* 1998; **51**(3):42–8.

Indinnimeo 1999 *{published data only}*
Indinnimeo M, Cicchini C, Stazi A, Ghini C, Alessandrini G, Reale MG. Carcinoembryonic antigen in recurrence of colorectal cancer. 23rd National Congress of the Societa-Italiana-di-Chirurgia-Oncologica (SICO), Perugia, Italy. 1999.

Ito 2002 *{published data only}*
Ito K, Hibi K, Ando H, Hidemura K, Yamazaki T, Akiyama S, et al. Usefulness of analytical CEA doubling time and half-life time for overlooked synchronous metastases in colorectal carcinoma. *Japanese Journal of Clinical Oncology* 2002;**32**(2):54–8.

Jaeger 1975 *{published data only}*
Jaeger P, Pettavel J, Wuilleret B, Bertholet MM, Mach JP. [Value and limits of the determination of carcinoembryonic antigen (CEA) in postoperative evaluation of patients with colonic and rectal carcinomas]. *Schweizerische Medizinische Wochenschrift Journal Suisse de Medecine* 1975;**105**(46): 1533–8.

Jiang 1989 *{published data only}*
Jiang R. [Clinical significance of serum CEA determination in the diagnosis of colorectal cancer]. *Chung-Hua Chung Liu Tsa Chih [Chinese Journal of Oncology]* 1989;**11**(5): 348–51.

Kanellos 2006b *{published data only}*
Kanellos I, Zacharakis E, Demetriades H, Christoforidis E, Kanellos D, Pramateftakis MG, et al. Value of carcinoembryonic antigen assay in predicting hepatic metastases, local recurrence, and survival after curative resection of colorectal cancer. *Surgery Today* 2006;**36**(10): 879–84.

Karesen 1980 *{published data only}*
Karesen R, Hertzberg J, Johannesen J, Thoresen BO, Orjasaeter H. Carcinoembryonic antigen in the diagnosis and follow-up of colorectal carcinoma. *American Journal of Proctology, Gastroenterology & Colon & Rectal Surgery* 1980; **31**(11):18–22.

Kawamura 2010 *{published data only}*
Kawamura YJ, Tokumitsu A, Mizokami K, Sasaki J, Tsujinaka S, Konishi F. First alert for recurrence during follow-up after potentially curative resection for colorectal carcinoma: CA 19-9 should be included in surveillance programs. *Clinical Colorectal Cancer* 2010;**9**(1):48–51.

Kerr 2012 *{published data only}*
Kerr NA, Jha B, Edwards T, Karnati G, Mackey PM. The detection of colorectal cancer recurrence following curative resection. *Colorectal Disease* 2012;**14**(Suppl 1):12–40.

Khan 2009 *{published data only}*
Khan K, Rathore M, Loughlin V, Tham TCK, Bhatti MI, Allen D, et al. Retrospective analysis of resected primary colorectal cancer revealed no correlation between node harvest and node involvement. *European Journal of Cancer* 2009;**7**(2-3):326.

Kimura 1986 *{published data only}*
Kimura O, Kaibara N, Nishidoi H, Okamoto T, Takebayashi M, Kawasumi H, et al. Carcinoembryonic antigen slope analysis as an early indicator for recurrence of colorectal carcinoma. *Japanese Journal of Surgery* 1986;**16** (2):106–11.

Kishimoto 2010 *{published data only}*
Kishimoto G, Murakami K, Con SA, Yamasaki E, Domeki Y, Tsubaki M, et al. [Follow-up after curative surgery for colorectal cancer: impact of positron emission tomography - computed tomography (PET/CT)]. *Revista de Gastroenterologia del Peru* 2010;**30**(4):328–33.

Koch 1977 *{published data only}*
Koch M, McPherson TA, Morrish DW. Carcinoembryonic antigen: 3 years' experience in a cancer clinic.. *Canadian Medical Association Journal* 1977 Apr 9;**116**(7):769–71.

Koch 1979 *{published data only}*
Koch M, McPherson TA. Predictive value of plasma CEA in patients with colorectal carcinoma. *Journal of Surgical Oncology* 1979;**12**(4):319–25.

Koch 1982 *{published data only}*
Koch M, Washer G, Gaedke H, McPherson TA. Carcinoembryonic antigen: usefulness as a postsurgical method in the detection of recurrence in Dukes stages B2 and C colorectal cancers. *Journal of the National Cancer Institute* 1982;**69**(4):813–5.

Koga 1999 *{published data only}*
Koga H, Moriya Y, Akasu T, Fujita S. The relationship between prognosis and CEA-dt after hepatic resection in patients with colorectal carcinomas. *European Journal of Surgical Oncology* 1999;**25**(3):292–6.

Korner 2005 *{published data only}*
Körner H, Söreide K, Stokkeland PJ, Söreide JA. Systematic follow-up after curative surgery for colorectal cancer in Norway: a population-based audit of effectiveness, costs, and compliance. *Journal of Gastrointestinal Surgery* 2005;**9**(3):320–8.

Kumar 2011 *{published data only}*
Kumara K, Aggarwalb D, Ardakanib A, Syedb M, Ingham Clark C. Efficacy of an intensive colorectal cancer follow-up programme. *European Journal of Surgical Oncology* 2011;**37**(11):997.

Lagache 1980 *{published data only}*
Lagache G, Dessaint JP, Triboulet JP. Indications for repeat operations for recurrence of rectocolic cancer: Contribution of serum carcino-embryonic antigen (CEA) levels [[French] Valeur du taux serique d'antigene carcino–embryonnaire (A.C.E.) dans l'indication de reintervention pour recidive de cancer colo–rectal]. *Chirurgie - Memoires de l'Academie de Chirurgie* 1980;**106**(5):322–34.

Lauterbach 1987 *{published data only}*
Lauterbach M, Dehne A, Hesse V, Hofig G. [CEA determination in the follow-up of colorectal cancers]. *Zentralblatt fur Chirurgie* 1987;**112**(15):968–74.

Lavin 1981 *{published data only}*
Lavin PT, Day J, Holyoke ED, Mittelman A, Chu TM. A statistical evaluation of baseline and follow-up carcinoembryonic antigen in patients with resectable colorectal carcinoma. *Cancer.* 1981;**47**(4):823–6.

Lechner 2000 *{published data only}*
Lechner P, Lind P, Goldenberg DM. Can postoperative surveillance with serial CEA immunoscintigraphy detect resectable rectal cancer recurrence and potentially improve tumor-free survival?. *Journal of American College of Surgeons* 2000;**191**(5):511–8.

Leventakos 2013 *{published data only}*
Leventakos K, Lu SS, Perry DJ. Intensive CT scan surveillance for patients who have undergone curative intent treatment for colorectalcancer: The Medstar Washington Hospital Center experience. *Journal of Clinical Oncology* 2013;**31**(Suppl):abstract e14675.

Levy 2012 *{published data only}*
Levy M, Lipska L, Visokai V, Veskrna K, Simsa J. Tumor markers in colorectal cancer relapse. *Tumor Biology* 2012;**33**(Suppl I):S15–S80.

Lipska 2007 *{published data only}*
Lipska L, Visokai V, Levy M, Svobodova S, Kormunda S, Finek J. Tumor markers in patients with relapse of colorectal carcinoma. *Anticancer Research* 2007;**27**(4A):1901–5.

Lipska 2010 *{published data only}*
Lipska L, Visokai V, Levy M. Resectability of colorectal cancer relapse. *European Journal of Surgical Oncology* 2010;**36**(9):876–7.

Lorenz 1986 *{published data only}*
Lorenz M, Happ J, Hottenrott C, Maul FD, Baum RP, Hor G, et al. [Clinical evaluation of the tumor marker CA 19-9 in comparison with carcinoembryonic antigen (CEA) in surgical pre- and postoperative diagnosis]. *Nuclear-Medizin* 1986;**25**(1):9–14.

Lunde 1982 *{published data only}*
Lunde OC, Havig O. Clinical significance of carcinoembryonic antigen (CEA) in patients with adenocarcinoma in colon and rectum. *Acta Chirurgica Scandinavica* 1982;**148**(2):189–93.

Ma 2006 *{published data only}*
Ma CJ, Hsieh JS, Wang WM, Su YC, Huang CJ, Huang TJ, et al. Multivariate analysis of prognostic determinants for colorectal cancer patients with high preoperative serum CEA levels: prognostic value of postoperative serum CEA levels. *Kaohsiung Journal of Medical Sciences* 2006;**22**(12):604–9.

Mach 1974 *{published data only}*
Mach JP, Jaeger P, Bertholet MM, Ruegsegger CH, Loosli RM, Pettavel J. Detection of recurrence of large-bowel carcinoma by radioimmunoassay of circulating carcinoembryonic antigen (C.E.A.). *Lancet* 1974;**2**(7880):535–40.

Makela 1995 *{published data only}*
Makela JT, Laitinen SO, Kairaluoma MI. Five-year follow-up after radical surgery for colorectal cancer. Results of a prospective randomized trial. *Archives of Surgery* 1995;**130**(10):1062–7.

Makis 2013 *{published data only}*
Makis W, Kurzencwyg D, Hickeson M. 18F-FDG PET/CT superior to serum CEA in detection of colorectal cancer and its recurrence. *Clinical Imaging* 2013;**37**(6):1094–7.

Mant 2013 *{published data only}*
Mant D, Perera R, Gray A, Rose P, Fuller A, Corkhill A, et al. Effect of 3-5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: FACS randomized controlled trial. Journal of Clinical Oncology. Chicago: Annual Meeting of the American Society of Clinical Oncology, ASCO, 2013; Vol. 20.

Martin 1976 *{published data only}*
Martin EW, Jr, Kibbey WE, DiVecchia L, Anderson G, Catalano P, Minton JP. Carcinoembryonic antigen: clinical and historical aspects. *Cancer* 1976;**37**(1):62–81.

Martin 1979 *{published data only}*
Martin EW Jr, Cooperman M, King G, Rinker L, Carey LC, Minton JP. A retrospective and prospective study of serial CEA determinations in the early detection of recurrent colon cancer. *American Journal of Surgery* 1979;**137**(2):167–9.

Martin 1980 *{published data only}*
Martin EW Jr, Cooperman M, Carey LC, Minton JP. Sixty second-look procedures indicated primarily by rise in serial carcinoembryonic antigen. *Journal of Surgical Research* 1980;**28**(5):389–94.

Marucci 1983 *{published data only}*
Marucci MM, Capussotti L, Molinaro G, Duglio A, Torossian K, Marini C. [Carcinoembryonic antigen in colorectal tumors. Prognostic evaluation and postoperative monitoring]. *Minerva Chirurgica* 1983;**38**(11):751–5.

May 2012 *{published data only}*
May DJ, Richardson JRC, Saunders BW, Miles AJG. Intensive long term follow-up after T1 and T2 node negative colorectal cancer is not necessary. *Colorectal Disease* 2012;**14**:22–3.

Mazilu 2012 *{published data only}*
Mazilu L, Ciufu N, Galan M, Suceveanu AI, Suceveanu AP, Parepa IR, et al. Postherapeutic follow-up of colorectal cancer patients treated with curative intent. *Chirurgia* 2012; **107**(1):55–8.

McCarthy 1985 *{published data only}*
McCarthy SM, Barnes D, Deveney K, Moss AA, Goldberg HI. Detection of recurrent rectosigmoid carcinoma: prospective evaluation of CT and clinical factors. *American Journal of Roentgenology* 1985;**144**(3):577–9.

Meling 1992 *{published data only}*
Meling GI, Rognum TO, Clausen OP, Børmer O, Lunde OC, Schlichting E, et al. Serum carcinoembryonic antigen in relation to survival, DNA ploidy pattern, and recurrent disease in 406 colorectal carcinoma patients. *Scandinavian Journal of Gastroenterology* 1992 Dec;**27**(12):1061–8.

Mentges 1986 *{published data only}*
Mentges B, Grussner R, Klotter HJ, Batz W. 10 years experience with CEA in the postoperative monitoring of patients undergoing surgery to cure colorectal carcinomas. *Acta Medica Austriaca* 1986;**13**:93–4.

Mentges 1988 *{published data only}*
Mentges B. [Effect of serial CEA determination on diagnosis, therapy and prognosis of recurrent colorectal cancer]. *Langenbecks Archiv für Chirurgie* 1988;**373**(4): 227–34.

Metzger 1983 *{published data only}*
Metzger U, Decurtins M, Joller H. Carcinoembryonic antigen (CEA) in large bowel cancer treated by surgery [[German] Das karzinoembryonale antigen (Cea) beim operierten dickdarmkarzinom]. *schweizerische medizinische wochenschrift* 1983;**113**(15):548–9.

Metzger 1985 *{published data only}*
Metzger U, Bronz K, Buhler H, Dolder A, Seefeld U, Hollinger A, et al. [Prospective follow-up study of radically resected colorectal carcinoma. Status after 5 years]. *Schweizerische Medizinische Wochenschrift Journal Suisse de Medecine* 1985;**115**(29):1001–4.

Minton 1978a *{published data only}*
Minton JP, James KK, Hurtubise PE, Rinker L, Joyce S, Martin EW, Jr. The use of serial carcinoembryonic antigen determinations to predict recurrence of carcinoma of the colon and the time for a second-look operation. *Surgery, Gynecology & Obstetrics* 1978;**147**(2):208–10.

Minton 1978b *{published data only}*
Minton JP, Martin EW Jr. The use of serial CEA determinations to predict recurrence of colon cancer and when to do a second-look operation. *Cancer* 1978;**42**(3 Suppl):1422–7.

Minton 1989 *{published data only}*
Minton J, Chevinsky AH. CEA directed second-look surgery for colon and rectal cancer. *Chirurgiae et Gynaecologiae* 1989;**78**(1):32–7.

Miwa 1980 *{published data only}*
Miwa Y. Serum CEA levels in various gastrointestinal diseases with special reference to colorectal cancer. *Japanese Journal of Gastroenterology* 1980;**77**(7):1069–75.

Moertel 1978 *{published data only}*
Moertel CG, Schutt AJ, Go VL. Carcinoembryonic antigen test for recurrent colorectal carcinoma. *JAMA* 1978;**239** (11):1065–6.

Morelli 1985 *{published data only}*
Morelli M, Nardi M, Valle M, Balducci D, Serafini D, Bernardini P. [Experience in the use of the tumor markers CEA, GICA and TPA in the postoperative monitoring of colorectal neoplasms]. *Minerva Medica* 1985;**76**:34–35.

Moreno Carretero 1998 *{published data only}*
Moreno Carretero G, Cerdan Miguel FJ, Maestro de las Casas ML, Martinez Cortijo S, Ortega MD, Pardo Martinez M, et al. Serum and tissue CEA in colorectal cancer: clinical relevance. *Revista Espanola de Enfermedades Digestivas* 1998; **90**(6):391–401.

Moschl 1980 *{published data only}*
Moschl P, Staritz C, Keiler A, Kreuzer W, Fasching W. [Carcino-embryonic antigen as screening protein in the follow-up of patients with surgically-treated gastrointestinal cancer (author's transl)]. *Wiener Klinische Wochenschrift* 1980;**92**(4):128–30.

Nicolini 1995 *{published data only}*
Nicolini A, Caciagli M, Zampieri F, Ciampalini G, Carpi A, Spisni R, et al. Usefulness of CEA, TPA, GICA, CA 72.4, and CA 195 in the Diagnosis of primary colorectal cancer and at its relapse. *Cancer Detection & Prevention* 1995;**19** (2):183–95.

Nicolini 2005 *{published data only}*
Nicolini A, Ferrari P, Anselmi L, Metelli MR, Carpi A, Spisni R, et al. Recurrences of colorectal cancer: time distribution and diagnostic sensitivity of serum CEA, TPA, CA19.9, CA72.4 tumour markers. *Journal of Clinical Oncology* 2005;**23**(16):301S–S.

Nicolini 2010 *{published data only}*
Nicolini A, Ferrari P, Duffy MJ, Antonelli A, Rossi G, Metelli MR, et al. Intensive risk-adjusted follow-up with the CEA, TPA, CA19.9, and CA72.4 tumor marker panel and abdominal ultrasonography to diagnose operable

colorectal cancer recurrences: effect on survival. *Archives of Surgery* 2010;**145**(12):1177–83.

Northover 1985 ***{published data only}***
Northover JM. Carcinoembryonic antigen and recurrent colorectal cancer. *British Journal of Surgery* 1985;**Suppl**: S44–6.

Northover 1986 ***{published data only}***
Northover J. Carcinoembryonic antigen and recurrent colorectal cancer. *Gut* 1986;**27**(2):117–22.

Northover 2003 ***{published data only}***
Northover J. Follow-up after curative surgery for colorectal cancer. *Scandinavian Journal of Surgery: SJS* 2003;**92**(1): 84–9.

Novis 1986 ***{published data only}***
Novis BH, Gluck E, Thomas P, Steele GD, Zurawski VR Jr, Stewart R, et al. Serial levels of CA 19-9 and CEA in colonic cancer. *Journal of Clinical Oncology* 1986;**4**(6):987–93.

Nowacki 1983 ***{published data only}***
Nowacki M. [Clinical usefulness of the study of serum carcinoembryonic antigen (CEA) for the determination of the degree of tumor progression, prognosis and monitoring of surgical treatment of neoplasms of the large intestine]. *Nowotwory* 1983;**33**(1):13–26.

Ntinas 2004 ***{published data only}***
Ntinas A, Zambas N, Al Mogrambi S, Petras P, Chalvatzoulis E, Frangandreas G, et al. Postoperative follow-up of patients with colorectal cancer: a combined evaluation of CT scan, colonoscopy and tumour markers. *Techniques in Coloproctology* 2004;**8**(Suppl 1):s190–2.

O'Dwyer 1987 ***{published data only}***
O'Dwyer PJ, Mojzisik C, McCabe DP, Sickle-Santanello BJ, Farrar WB, Martin EW Jr. Variation in recognition of recurrent colonic cancer by different CEA assays. *Diseases of the Colon and Rectum* 1987;**30**(2):133–6.

O'Dwyer 1988 ***{published data only}***
O'Dwyer PJ, Mojzisik C, McCabe DP, Farrar WB, Carey LC, Martin EW Jr. Reoperation directed by carcinoembryonic antigen level: the importance of a thorough preoperative evaluation. *American Journal of Surgery* 1988;**155**(2):227–31.

Obradovic 2011 ***{published data only}***
Obradovic VB, Artiko V, Sobic D, Petrovic N, Todorovic-Tirnanic M, Brajkovic L, et al. Diagnosis of recurrences and metastases of colorectal carcinomas using 18FDG PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**:S274.

Odariuk 1989 ***{published data only}***
Odariuk TS, Sevost'ianov SI, Mit'kov VV, Orlova LP. [A comparative evaluation of methods for diagnosing recurrence of rectal cancer after radical surgery]. *Voprosy Onkologii* 1989;**35**(9):1097–103.

Ovaska 1989 ***{published data only}***
Ovaska JT, Jarvinen HJ, Mecklin JP. The value of a follow-up programme after radical surgery for colorectal carcinoma. *Scandinavian Journal of Gastroenterology* 1989; **24**(4):416–22.

Ozhiganov 1986 ***{published data only}***
Ozhiganov EL, Kuznetsova LF. [Determination of carcinoembryonic and carbohydrate antigens in the diagnosis of recurrences and metastases of rectal cancer]. *Meditsinskaia Radiologiia* 1986;**31**(8):40–4.

Ozkan 2012a ***{published data only}***
Ozkan E, Soydal C, Araz M, Aras G. Serum carcinoembryonic antigen measurement, abdominal contrast-enhanced computed tomography, and fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography in the detection of colorectal cancer recurrence: a correlative study. *Nuclear Medicine Communications* 2012;**33**(9):990–4.

Ozkan 2012b ***{published data only}***
Ozkan E, Soydal C, Araz M, Kir KM, Ibis E. The role of 18F-FDG PET/CT in detecting colorectal cancer recurrence in patients with elevated CEA levels. *Nuclear Medicine Communications* 21012;**33**(4):395–402.

Park 2012 ***{published data only}***
Park I, You Y, Skibber JM, Rodriguez-Bigas MA, Feig B, Nguyen S, et al. Recurrence patterns after multidisciplinary therapy for rectal cancer in 725 patients. *Annals of Surgical Oncology* 2012;**19**:S15–6.

Park 2013 ***{published data only}***
Park JW, Chang HJ, Kim BC, Yeo HY, Kim DY. Clinical validity of tissue carcinoembryonic antigen expression as ancillary to serum carcinoembryonic antigen concentration in patients curatively resected for colorectal cancer. *Colorectal Disease* 2013;**15**(9):e503–11.

Pecorella 1996 ***{published data only}***
Pecorella G, Bracchitta S, Petrolito E, Cacciaguerra B, Blanco F, Cirino E. [Follow up in carcinoma of the large intestine]. *Annali Italiani di Chirurgia* 1996;**67**(1):41–7.

Peethambaram 1997 ***{published data only}***
Peethambaram P, Weiss M, Loprinzi CL, Novotny P, O'Fallon JR, Erlichman C, et al. An evaluation of postoperative follow-up tests in colon cancer patients treated for cure. *Oncology* 1997;**54**(4):287–92.

Pereira 2004 ***{published data only}***
Pereira JMC, Martinez C, Pimenta T, Oliveira MC. Carcinoembryonic antigen dosage and colorectal cancer recurrence. *Diseases of the Colon and Rectum* 2004;**47**(6): 986.

Persijin 1981 ***{published data only}***
Persijn JP, Hart AA. Prognostic significance of CEA in colorectal cancer: a statistical study. *Journal of Clinical Chemistry & Clinical Biochemistry* 1981;**19**(11):1117–23.

Pfeiffer 1979 ***{published data only}***
Pfeiffer R, Reis HE, Wittig HD. CEA determination with the Hansen method: Postsurgical follow-up of patients with colorectal carcinoma [[German] Erfahrungen mit der cea–bestimmung bei der postoperativen verlaufskontrolle

von patienten mit colorectalem carcinom]. *Journal of Cancer Research and Clinical Oncology* 1979;**93**(1):85–92.

Philips 1984 ***{published data only}***
Phillips RK, Hittinger R, Blesovsky L, Fry JS, Fielding LP. Local recurrence following 'curative' surgery for large bowel cancer: I. The overall picture. *British Journal of Surgery* 1984;**71**(1):12–6.

Pietra 1998 ***{published data only}***
Pietra N, Sarli L, Costi R, Ouchemi C, Grattarola M, Peracchia A. Role of follow-up in management of local recurrences of colorectal cancer - A prospective, randomized study. *Diseases of the Colon and Rectum* 1998;**41**(9): 1127–33.

Plebani 1996 ***{published data only}***
Plebani M, De Paoli M, Basso D, Roveroni G, Giacomini A, Galeotti F, et al. Serum tumor markers in colorectal cancer staging, grading, and follow-up. *Journal of Surgical Oncology* 1996;**62**(4):239–44.

Pompecki 1980 ***{published data only}***
Pompecki R, Winkler R. Clinical value of serial serum CEA determinations in postoperative follow-up of colorectal cancer [[German] Klinische bedeutung der routinemassigen serum–cea–bestimmung fur die postoperative kontrolle des kolorektalen karzinoms]. *Medizinische Welt* 1980;**31**(49): 1780–3.

Pribelsky 2002 ***{published data only}***
Pribelsky M, Pechan J, Krizan M, Pindak D. CEA and relapse after the operation of colorectal carcinoma. *Bratislavske Lekarske Listy* 2002;**103**(11):428–31.

Primrose 2011 ***{published data only}***
Primrose JN, Fuller A, Rose P, Perera-Salazar R, Mellor J, Corkhill A, et al. Follow-up after colorectal cancer surgery: Preliminary observational findings from the UK FACS trial. Journal of Clinical Oncology. Chicago: ASCO Annual Meeting 2011, 2011.

Primrose 2014 ***{published data only}***
Primrose JN, Perera R, Gray A, Rose P, Fuller A, Corkhill A, et al. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial.. *JAMA* 2014;**311**(3): 263–70.

Quentmeier 1990 ***{published data only}***
Quentmeier A, Schlag P, Smok M, Herfarth C. Re-operation for recurrent colorectal cancer: the importance of early diagnosis for resectability and survival. *European Journal of Surgical Oncology* 1990;**16**(4):319–25.

Reddy 2013 ***{published data only}***
Reddy GA, Abrar ML, Rajender K, Anish B, Mittal BR. Role of FDG PET/CT in evaluation of suspected recurrence of disease in patients of colon carcinoma in relation to carcinoembryonic antigen (CEA) levels. *European Journal of Nuclear Medicine and Molecular Imaging* 2013;**40**:S474–S.

Revetria 1989 ***{published data only}***
Revetria P, Repetto L, Perino M, Ciabattoni N, Gambetta G, Ferro A. Postoperative monitoring of carcinoma of the large bowel. Diagnostic and prognostic value of tumor markers CEA, CA 19-9 and TPA. [Italian]. *Chirurgia* 1989; **2**(3):121–4.

Rezamansourian 2011 ***{published data only}***
Rezamansourian A, Ghaemi E. The prevalence of elevated carcinoembroynic antigen at Gorgan south-east Caspian sea of Northern Iran. *Journal of Clinical and Diagnostic Research* 2011;**5**(1):74–7.

Rieger 1975 ***{published data only}***
Rieger A, Wahren B. CEA levels at recurrence and metastases; importance for detecting secondary disease. *Scandinavian Journal of Gastroenterology* 1975;**10**(8): 869–74.

Rockall 1999 ***{published data only}***
Rockall TA, McDonald PJ. Carcinoembryonic antigen: its value in the follow-up of patients with colorectal cancer. *International Journal of Colorectal Disease* 1999;**14**(1):73–7.

Rocklin 1990 ***{published data only}***
Rocklin MS, Slomski CA, Watne AL. Postoperative surveillance of patients with carcinoma of the colon and rectum. *American Surgeon* 1990;**56**(1):22–7.

Rocklin 1991 ***{published data only}***
Rocklin MS, Senagore AJ, Talbott TM. Role of carcinoembryonic antigen and liver function tests in the detection of recurrent colorectal carcinoma. *Diseases of the Colon and Rectum* 1991;**34**(9):794–7.

Rodriguez-Moranta 2006a ***{published data only}***
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, randomized, controlled trial. *Journal of Clinical Oncology* 2006;**24**(3):386–93.

Rognum 1986 ***{published data only}***
Rognum TO, Heier HE, Orjasaeter H, Thorud E, Brandtzaeg P. Comparison of two CEA assays in primary and recurrent large bowel carcinoma with different DNA ploidy pattern. *European Journal of Cancer & Clinical Oncology* 1986;**22**(10):1165–9.

Sagar 1989 ***{published data only}***
Sagar PM, Cooper EH, Finan PJ. A prospective assessment of the tumor-markers Ca50, Ca195 and CEA in prediction of tumor stage and recurrence in colorectal-cancer. *British Journal of Cancer* 1989;**60**(3):488.

Sandelewski 2005 ***{published data only}***
Sandelewski A, Kokocinska D, Partyka R, Kocot J, Starzewski J, Chanek I, et al. [Usefulness of evaluation of carcinoembryonic antigen (CEA) and soluble fragments of cytokeratin 18-th (TPS) in postoperative monitoring of patients with colorectal cancer]. *Polski Merkuriusz Lekarski* 2005;**18**(108):647–50.

Sanli 2012 ***{published data only}***
Sanli Y, Kuyumcu S, Ozkan ZG, Kilic L, Balik E, Turkmen C, et al. The utility of FDG-PET/CT as an effective tool for detecting recurrent colorectal cancer regardless of serum CEA levels. *Annals of Nuclear Medicine* 2012;**26**(7):551–8.

Sardi 1989 *{published data only}*
Sardi A, Agnone CM, Nieroda CA, Mojzisik C, Hinkle G, Ferrara P, et al. Radioimmunoguided surgery in recurrent colorectal cancer: the role of carcinoembryonic antigen, computerized tomography, and physical examination. *Southern Medical Journal* 1989;**82**(10):1235–44.

Sarikaya 2007 *{published data only}*
Sarikaya I, Bloomston M, Povoski SP, Zhang J, Hall NC, Knopp MV, et al. FDG-PET scan in patients with clinically and/or radiologically suspicious colorectal cancer recurrence but normal CEA. *World Journal of Surgical Oncology* 2007; **5**:64.

Secco 1989 *{published data only}*
Secco GB, Fardelli R, Campora E, Rovida S, Onetto M, Marroni P, et al. CEA as a prognostic factor and early indicator of recurrence in colorectal cancer. *Journal of Experimental and Clinical Cancer Research* 1989;**8**(3):173–7.

Secco 2000 *{published data only}*
Secco GB, Fardelli R, Rovida S, Gianquinto D, Baldi E, Bonfante P, et al. Is intensive follow-up really able to improve prognosis of patients with local recurrence after curative surgery for rectal cancer?. *Annals of Surgical Oncology* 2000;**7**(1):32–7.

Segol 1977 *{published data only}*
Segol PH, Travert G, Davy A. Serum carcinoembryonic antigen determinations during post operative follow up of colorectal carcinomas [[French] Le dosage de l'antigene carcino embryonnaire dans la surveillance post operatoire des cancers colo rectaux]. *Lyon Chirurgical* 1977;**73**(4): 289–92.

Shirley 2012 *{published data only}*
Shirley LA, McNally M, Huntington J, Jones N, Malhotra L, Bloomston M, et al. Correlation of postoperative CEA trends with survival and patterns of recurrence after hepatectomy for colorectal cancer metastases. Journal of Clinical Oncology. San Francisco, CA United States: 2012 Gastrointestinal Cancers Symposium, 2012.

Simo 2002 *{published data only}*
Simo M, Lomena F, Setoain J, Perez G, Castellucci P, Costansa JM, et al. FDG-PET improves the management of patients with suspected recurrence of colorectal cancer. *Nuclear Medicine Communications* 2002;**23**(10):975–82.

Sirisriro 1996 *{published data only}*
Sirisriro R, Podoloff DA, Patt YZ, Curley SA, Kasi LP, Bhadkamkar VA, et al. 99Tcm-IMMU4 imaging in recurrent colorectal cancer: efficacy and impact on surgical management. *Nuclear Medicine Communications* 1996;**17** (7):568–76.

Song 2010 *{published data only}*
Song S, Minsky B, Polite B, Liauw S. Post-treatment CEA values, and CEA trends at the time of uncertain post-treatment imaging, have prognostic value after combined modality therapy for rectal cancer. International Journal of Radiation Oncology Biology Physics. 52nd Annual Meeting of the American Society for Radiation Oncology, San Diego, CA, USA. 2010:S313.

Sorensen 2010 *{published data only}*
Sorensen NF, Jensen AB, Wille-Jorgensen P, Friberg L, Rordam L, Ingeman L, et al. Strict follow-up programme including CT and 18F-FDG-PET after curative surgery for colorectal cancer. *Colorectal Disease* 2010;**12**:e224–8.

Staab 1985a *{published data only}*
Staab HJ, Anderer FA, Stumpf E, Hornung A, Fischer R, Kieninger G. Eighty-four potential second-look operations based on sequential carcinoembryonic antigen determinations and clinical investigations in patients with recurrent gastrointestinal cancer. *American Journal of Surgery* 1985;**149**(2):198–104.

Staab 1985b *{published data only}*
Staab HJ, Brummendorf T, Hornung A, Anderer FA, Kieninger G. The clinical validity of circulating tumor-associated antigens CEA and CA 19-9 in primary diagnosis and follow-up of patients with gastrointestinal malignancies. *Klinische Wochenschrift* 1985;**63**(3):106–15.

Stautner-Brückmann 1990 *{published data only}*
Stautner-Brückmann C, Schneider W, Gresser U, Richter-Turtur M, Eibl-Eibesfeld B, Zoller WG, et al. Is ultrasound superior to carcinoembryonic antigen (CEA) measurement in recurrent tumor screening of patients with colorectal carcinoma?. *Bildgebung* 1990;**57**(1-2):17–20.

Steele 1980 *{published data only}*
Steele G Jr, Zamcheck N, Wilson R. Results of CEA-initiated second-look surgery for recurrent colorectal cancer. *American Journal of Surgery* 1980;**139**(4):544–8.

Stuckle 2000 *{published data only}*
Stuckle CA, Ibing HP, Adamietz IA. Value of computed tomography and tumor markers in detection of recurrent rectal carcinoma after surgery and radiotherapy [[German] Wertigkeit der computertomographie und der tumormarker zur beurteilung von lokalrezidiven beim operierten und nachbestrahlten rektumkarzinom]. *Tumor Diagnostik und Therapie* 2000;**21**(3):61–7.

Su 2012 *{published data only}*
Su BB, Shi H, Wan J. Role of serum carcinoembryonic antigen in the detection of colorectal cancer before and after surgical resection. *World Journal of Gastroenterology* 2012; **18**(17):2121–6.

Sugarbaker 1976 *{published data only}*
Sugarbaker PH, Zamcheck N, Moore FD. Assessment of serial carcinoembryonic antigen (CEA) assays in postoperative detection of recurrent colorectal cancer. *Cancer* 1976;**38**(6):2310–5.

Szymendera 1982 a *{published data only}*
Szymendera JJ, Nowacki MP, Szawlowski AW, Kaminska JA. Predictive value of plasma CEA levels: preoperative prognosis and postoperative monitoring of patients with colorectal carcinoma. *Diseases of the Colon and Rectum* 1982;**25**(1):46–52.

Szymendera 1982 b *{published data only}*
Szymendera JJ, Wilczynska JE, Nowacki MP, Kaminska JA, Szawowski AW. Serial CEA assays and liver scintigraphy

for the detection of hepatic metastases from colorectal carcinoma. *Diseases of the Colon and Rectum* 1982;**25**(3): 191–7.

Szymendera 1985 ***{published data only}***
Szymendera JJ, Nowacki MP, Kozlowicz-Gudzinska I, Kowalska M. Value of serum levels of carcinoembryonic antigen, CEA, and gastrointestinal cancer antigen, GICA or CA 19-9, for preoperative staging and postoperative monitoring of patients with colorectal carcinoma. *Diseases of the Colon and Rectum* 1985;**28**(12):895–9.

Takashima 1982 ***{published data only}***
Takashima S, Kosaka T, Uemura T. Clinical meanings of carcinoembryonic antigen (CEA) in patients with colorectal cancer. *Journal of the Japan Society of Colo-Proctology* 1981; **35**(2):137–46.

Tomoda 1981 ***{published data only}***
Tomoda H, Furusawa M. The usefulness and limitations of CEA assay in the management of colorectal cancer. *Japanese Journal of Surgery* 1981;**11**(1):33–8.

Tsai 2009 ***{published data only}***
Tsai HL, Chu KS, Huang YH, Su YC, Wu JY, Kuo CH, et al. Predictive factors of early relapse in UICC stage I-III colorectal cancer patients after curative resection. *Journal of Surgical Oncology* 2009;**100**(8):736–43.

Tsikitis 2009 ***{published data only}***
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**(22):3671–6.

Verberne 2013 a ***{published data only}***
Verberne CJ, Wiggers T, Vermeulen KM, De Jong KP. Detection of recurrences during follow-up after liver surgery for colorectal metastases: both carcinoembryonic antigen (CEA) and imaging are important. *Annals of Surgical Oncology* 2013;**20**(2):457–63.

Verberne 2013 b ***{published data only}***
Verberne C, Doornbos PM, Grossmann I, De Bock GH, Wiggers T. Intensified follow-up in colorectal cancer patients using frequent carcino-embryonic antigen (CEA) measurements and CEA-triggered imaging. European Cancer Congress 2013, ECC 2013, Amsterdam. European Journal of Cancer. 2013; Vol. 49:S480.

Wan 1994 ***{published data only}***
Wang JY, Tang R, Chiang JM. Value of carcinoembryonic antigen in the management of colorectal cancer. *Diseases of the Colon and Rectum* 1994 Mar;**37**(3):272–7.

Wanebo 1978a ***{published data only}***
Wanebo HJ, Rao B, Pinsky CM, Hoffman RG, Stearns M, Schwartz MK, et al. Preoperative carcinoembryonic antigen level as a prognostic indicator in colorectal cancer. *New England Journal of Medicine* 1978;**299**(9):448–51.

Wanebo 1978b ***{published data only}***
Wanebo JH, Stearns M, Schwartz MK. Use of CEA as an indicator of early recurrence and as a guide to a selected second-look procedure in patients with colorectal cancer. *Annals of Surgery* 1978;**188**(4):481–93.

Wang 2007 ***{published data only}***
Wang JY, Lu CY, Chu KS, Ma CJ, Wu DC, Tsai HL, et al. Prognostic significance of pre- and postoperative serum carcinoembryonic antigen levels in patients with colorectal cancer. *European Surgical Research* 2007;**39**(4):245–50.

Wang 2010 ***{published data only}***
Wang ZY. Serum p53 antibody as a tumor marker in the follow-up of colorectal cancer after curative resection. *Techniques in Coloproctology* 2010;**14**(1):57–8.

Wedell 1981 ***{published data only}***
Wedell J, Meier zu Esssen P, Luu TH, Fiedler R, Van Calker H, Koldowski P, et al. A retrospective study of serial CEA determinations in the early detection of recurrent colorectal cancer. *Diseases of the Colon and Rectum* 1981;**24**(8): 618–21.

Weiss 1998 ***{published data only}***
Weiss NS, Cook LS. Evaluating the efficacy of screening for recurrence of cancer. *Journal of the National Cancer Institute* 1998;**90**(24):1870–2.

Wichmann 2000a ***{published data only}***
Wichmann MW, Lau-Werner U, Müller C, Hornung HM, Stieber P, Schildberg FW, et al. Carcinoembryonic antigen for the detection of recurrent disease following curative resection of colorectal cancer. *Anticancer Research* 2000;**20** (6D):4953–5.

Wichmann 2000b ***{published data only}***
Wichmann MW, Müller C, Lau-Werner U, Strauss T, Lang RA, Hornung HM, et al. The role of carcinoembryonic antigen for the detection of recurrent disease following curative resection of large-bowel cancer. *Langenbecks Archives of Surgery* 2000;**385**(4):271–5.

Wichmann 2002 ***{published data only}***
Wichmann MW, Müller C, Hornung HM, Lau-Werner U, Schildberg FW, Colorectal Cancer Study G. Results of long-term follow-up after curative resection of Dukes A colorectal cancer. *World Journal of Surgery* 2002;**26**(6):732–6.

Wolf 1997 ***{published data only}***
Wolf RF, Cohen AM. The miniscule benefit of serial carcinoembryonic antigen monitoring after effective curative treatment for primary colorectal cancer. *Journal of the American College of Surgeons* 1997;**185**(1):60–4.

Wood 1975 ***{published data only}***
Wood CB, Malcolm AJH, Burt R. Assessment of patients with elevated carcinoembryonic antigen (CEA) level after surgery for primary colorectal cancer. *Bulletin de la Societe Internationale de Chirurgie* 11975;**34**(6):413–6.

Yu 2013 ***{published data only}***
Yu H, Zhao M, Xing J, Jin H, Li Y. Relationship of the applied value of 18F-FDG PET/CT in postoperative relapse with metastasis of colorectal cancer and CEA levels during PET/CT scanning. *Chinese Journal of Clinical Oncology* 2013;**40**(12):717–20.

Zeng 1993 *{published data only}*

Zeng Z, Cohen AM, Urmacher C. Usefulness of carcinoembryonic antigen monitoring despite normal preoperative values in node-positive colon cancer patients. *Diseases of the Colon and Rectum* 1993;**36**(11):1063–8.

Zervos 2001 *{published data only}*

Zervos EE, Badgwell BD, Burak WE Jr, Arnold MW, Martin EW. Fluorodeoxyglucose positron emission tomography as an adjunct to carcinoembryonic antigen in the management of patients with presumed recurrent colorectal cancer and nondiagnostic radiologic workup. *Surgery* 2001;**130**(4): 636–43.

Ziegenbein 1980 *{published data only}*

Ziegenbein R, Jacobasch KH, Pilgrim G, Seifart W. [The CEA-concentration in plasma of patients with colorectal carcinoma and polyps has been determined by radioimmunoassay (RIA, Z-gel-method)]. *Archiv für Geschwulstforschung* 1980;**50**(2):165–8.

Zuniga 1989 *{published data only}*

Zuniga A, Rahmer A, Guzman S, Llanos O, Lopez F, Herreros R. [Colorectal cancer: follow-up after curative resection]. *Revista Medica de Chile* 1989;**117**(3):273–8.

Additional references

Allen 2013

Allen VB, Gurusamy KS, Takwoingi Y, Kalia A, Davidson BR. Diagnostic accuracy of laparoscopy following computed tomography (CT) scanning for assessing the resectability with curative intent in pancreatic and periampullary cancer. *Cochrane Database of Systematic Reviews* 2013, Issue 11. [DOI: 10.1002/14651858.CD009323.pub2]

Bormer 1991

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

Colibaseanu 2013

Colibaseanu DT, Mathis KL, Abdelsattar ZM, Larson DW, Haddock MG, Dozois EJ. Is curative resection and long-term survival possible for locally re-recurrent colorectal cancer in the pelvis?. *Diseases of the Colon and Rectum* 2013; **56**(1):14–9.

Cunningham 2010

Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal cancer. *Lancet* 2010;**375** (9719):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. *FASEB Journal* 2012;**26** (6):2648–56.

Davidson 1989

Davidson BR, Sams VR, Styles J, Dean C, Boulos PB. Comparative study of carcinoembryonic antigen and epithelial membrane antigen expression in normal colon, adenomas and adenocarcinomas of the colon and rectum. *Gut* 1989;**30**(9):1260–5.

Deeks 2005

Deeks JJ, Mackaskil P, Irwig L. 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**(9):882–93.

Duffy 2013a

Duffy MJ. Tumor markers in clinical practice: a review focusing on common solid cancers. *Medical Principles and Practice* 2013;**22**(1):4–11.

Duffy 2013b

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.

Glimelius 1992

Glimelius B, Påhlman L, Graf W, Tveit K, Adami HO, Iasarett V, et al. Expectancy or primary chemotherapy in patients with advanced asymptomatic colorectal cancer: a randomized trial. *Journal of Clinical Oncology* 1992;**10**(6): 904–11.

Godlee 2014

Godlee F. Colorectal cancer: a cautionary tale. *BMJ* 2014; **348**:g3311.

Goldstein 2005

Goldstein MJ, Mitchell EP. Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. *Cancer Investigations* 2005;**23**(4):339–51.

Gonzalez 2013

Gonzalez M, Poncet A, Combescure C, Robert J, Ris HB, Gervaz P. Risk factors for survival after lung metastasectomy in colorectal cancer patients: a systematic review and meta-analysis. *Annals of Surgical Oncology* 2013;**20**(2):572–9.

Gore 1997

Gore RM. Colorectal cancer. Clinical and pathologic features. *Radiological Clinics of North America* 1997;**35**(2): 403–29.

Guthrie 2002

Guthrie JA. Colorectal cancer: follow-up and detection of recurrence. *Abdominal Imaging* 2002;**27**(5):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]

Kanas 2012
Kanas GP, Taylor A, Primrose JN, Langeberg WJ, Kelsh MA, Mowat FS, et al. Survival after liver resection in metastatic colorectal cancer: review and meta-analysis of prognostic factors. *Clinical Epidemiology* 201;**4**:283–301.

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.

Leeflang 2008
Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic review of diagnostic test accuracy. *Annals of Internal Medicine* 2008;**149**(12):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
Macakill P, Gatonis C, Deeks JJ, Harbord RM, Takwoingi Y (editors). Chapter 10: Analysing and presenting results. In: Cochrane Handbook for Systematic Reviews of Test Accuracy Version 1.0, 2010. Available from dta.cochrane.org/handbook-dta-reviews. The Cochrane Collaboration.

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**(5):919–32.

Moses 1993
Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine* 1993;**12**(14): 1293–316.

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**(1):3–17.

NICE 2011
NICE clinical guideline [CG131]. Colorectal Cancer; diagnosis and management. www.nice.org.uk/guidance/cg131/chapter/1-recommendations 2011 (accessed 5th December 2015).

Reitsma 2005
Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.

Rodriguez-Moranta 2006b
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):1–8.

Sargent 2007
Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J, et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data From 20,898 patients enrolled onto 18 randomized trials from the ACCENT group. *Journal of Clinical Oncology* 2007;**25**(29):4569–74.

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.

Shinkins 2014
Shinkins B, Nicholson BD, James TJ, Primrose JN, Mant D. Carcinoembryonic antigen monitoring to detect recurrence of colorectal cancer: how should we interpret the results?. *Clinical Chemistry* 2014;**60**(12):1572–4.

Sobin 2009
Sobin LH, Gospodarowicz MK, 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**(1): 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: 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**(1):15–24.

Treasure 2014
Treasure T, Monson K, Fiorentino F, Russell C. The CEA Second-Look Trial:a randomised controlled trial of carcinoembryonic antigen prompted reoperation for recurrent colorectal cancer.. *BMJ Open* 2014;**4**:e004385.

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**(8): 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*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Banaszkiewicz 2011

Study characteristics	
Patient sampling	**Country** Poland **Study design** Retrospective casenote review **Setting** Hospital **Dates of data collection** N/R **Population (n)** 965 **Inclusion criteria** Patients after radical surgery in whom prognosis following a possible second operation was good **Exclusion criteria** Non-radical surgery or concomitant disease making survival of a second operation unlikely **Participants included (n)** 340
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A - D **Perioperative Investigations done to ensure no residual disease** Endoscopic polypectomy **Chemotherapy/radiotherapy?** Radical **Recurrences (n)** 112 **Site of recurrences** Liver 44, Local 32, Lung 7, Disseminated 12, other 6, 2 sites 11
Index tests	**CEA timing** CEA 3, 6, 12 months, then once a year up to 5 years **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?**

	N/R
Target condition and reference standard(s)	**Follow-up schedule** Follow-up visits at 3, 6, 12 months, then once a year up to 5 years. Follow-up schedule included patient's history and physical examination, measurement of CEA serum concentration and classic colonoscopy
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low

Banaszkiewicz 2011 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Yes		

Barillari 1992

Study characteristics	
Patient sampling	**Country** Italy **Study design** Prospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 66 **Inclusion criteria**

	Rectal cancer treated for cure **Exclusion criteria** N/R **Participants included (n)** 66
Patient characteristics and setting	**Age range** 62.3 yrs (mean) **Smoking status** N/R **Site of primary tumour** rectum **Stage of primary tumour** 6 Stage A, 32 Stage B, 28 Stage C **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 33 **Site of recurrences** Local 10, Metastatic 25 (Lungs 3, peritoneum 10,bones 2,liver 21, multiple 8)
Index tests	**CEA timing** 3-monthly CEA **CEA technique** CEA was analysed using a direct radioimmunologic method (CEA-PR; Sorin Biomedica) **CEA threshold** 3 μg/L **Definition of positive** Any elevation of 1 of the antigen levels greater than the limit defined by the between assay coefficient of variation (calculated on the basis of 2 standard deviations) was defined as significant, and the assay was repeated after 10 days **Which CEA value (s) used?** Repeated value.
Target condition and reference standard(s)	**Follow-up schedule** 3-monthly to 60 months: blood CEA, TPA, CA19.9 and clinical exam. 6, 18, 30, 42, 54 months: USS Abdomen, CXR, Barium Enema. 12, 24, 36, 48, 60 months: colonoscopy, CT body. 6, 18, 30, 42 months: Bone scan **Reference standard** Abdominal or total body CT, a chest x-ray examination, a bone scan, an endoscopy, and a clinical examination were performed. An exploratory laparotomy was performed when all three markers were elevated, even if recurrence was not confirmed by total body CT scan and clinical examinations
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	

Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Barillari 1992 (*Continued*)

			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Yes		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Yes		

Beart 1981

Study characteristics	
Patient sampling	**Country** USA **Study Design** Prospective **Setting** Department of Surgery and Oncology, Mayo Clinic and Mayo Foundation **Dates of data collection** 1976 - 1986 **Population (n)** 149 **Inclusion criteria** Resection of Dukes' B2 or C colorectal carcinoma was followed from the time of operation until the time of tumour recurrence or writing the published paper **Exclusion criteria** N/R **Participants included (n)** 149
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R

Beart 1981 (*Continued*)

	Site of primary tumour Colon **Stage of primary tumour** Dukes B or C **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Some got radiotherapy, chemotherapy, and/or immunotherapy. Numbers not specified **Recurrences (n)** 34 **Site of recurrences** Liver metastasis 14, Chest 6, Pelvic disease 12
Index tests	**CEA Timing** At least every 15 week **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R
Target condition and reference standard(s)	**Follow-up schedule** At least every 15 weeks a complete history was taken and physical examination was carried out. A CXR was obtained, and laboratory determinations included complete blood count, alkaline phosphatase, SGOT, SGPT, and CEA. LDH and proctoscopic examinations were done every 6 months. A BE and liver scanning were done annually **Reference standard** Additional tests including CT, laparoscopy, liver biopsy, and abdominal exploration were ordered as indicated by the history, physical examination, or positive laboratory results. All recurrent tumours were documented histologically
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		

Beart 1981 (*Continued*)

Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Bjerkeset 1988

Study characteristics	
Patient sampling	**Country** Norway **Study Design** Prospective **Setting** Hospital **Dates of data collection** 1976 - 1979 **Population (n)** 244 **Inclusion criteria** colorectal cancer resection operated for cure **Exclusion criteria** Did not survive resection, residual disease, resection margins not clear, pre- and post-op CEA determination **Participants included (n)** 164
Patient characteristics and setting	**Age range** N/R **Smoking status** Some, but not quantified **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 58, B 76, C 48, D 50, unknown 12 **Perioperative Investigations done to ensure no residual disease** Clear resection margins, no residual disease **Chemotherapy / Radiotherapy?** 22 Dukes B - C randomised to 5-year follow-up; 21 had preoperative external radiation **Recurrences (n)** 47

Bjerkeset 1988 (*Continued*)

	Site of recurrences Liver 12, Lungs 10, Local 8, Local and distant 5, carcinomatosis 6, multiple 6
Index tests	**CEA timing** 3, 6, 12, 18, 24, then yearly **CEA technique** Roche Ria test- repeat if raised, if repeat raised then test as described in follow-up **CEA threshold** 3.5 µg/L **Definition of positive** Transient **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** 3, 6, 12, 18, 24 months then yearly CEA, clinical, biochemical, immunological (immunoglobulins and complement). CXR. Colonoscopy 6, 18 months. DCBE 1, 3, 5 years. "complimentary radiographic, scintographic, ultrasonographic added if indicated." **Reference standard** If nothing found investigating for increased CEA, then a second-look operation was performed (laparotomy and biopsy) in 23, liver imaging 8, autopsy 5, clinical course 6
Flow and timing	**Timing of CEA vs reference standard (days)** as per follow-up schedule
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
Was the index test repeated prior to the reference standard?	Yes		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Carlsson 1983

Study characteristics	
Patient sampling	**Country** Sweden **Study design** Prospective study **Setting** Hospital **Dates of data collection** N/R **Population (n)** 163 **Inclusion Criteria** Curative operation for colorectal cancer **Exclusion Criteria** Advanced age, moving away, death 3 months postop **Participants Included (n)** 139
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative Investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 50 **Site of recurrences** N/R
Index tests	**CEA timing** Blood tests 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 60 months post-1977- blood tests 3, 6, 12, 18, 24, 30, 36, 42, 48, 60 months **CEA technique** Direct radio immunoassay method developed at the Department of Nuclear Medicine, Malmo General Hospital **CEA threshold** 3 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence

Carlsson 1983 (*Continued*)

Target condition and reference standard(s)	**Follow-up schedule** Until 1977: Follow-up exam and rectoscopy 3, 6, 9, 12, 15, 18 ,21, 24, 26, 42, 48, 60 months. Double contrast enema 3, 12, 24, 36, 48, 60 months. CXR and blood tests 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 60 months. From 1977: Physical exam and rectoscopy 3, 12, 24, 36, 48, 60 months. Double contrast enema 3, 12, 24, 36, 48, 60 months. CXR and blood tests 3, 6, 12, 18, 24, 30, 36, 42, 48, 60 months
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low

Carlsson 1983 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Carpelan-Holmström 2004

Study characteristics	
Patient sampling	**Country** Finland **Study Design** Retrospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 354 **Inclusion criteria**

	Curative surgery, but unclear **Exclusion criteria** Palliative, followed up elsewhere, no preoperative serum samples, no serum at the time of recurrence **Participants included (n)** 102
Patient characteristics and setting	**Age range** 29 - 88 yrs **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A - D (16 Dukes A, 45 Dukes B, 34 Dukes C, and 7 Dukes D) **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 40 **Site of recurrences** Local 17, Liver 10, Various 13
Index tests	**CEA timing** N/R **CEA technique** CEA was measured with a time-resolved immunofluorometric assay (AutoDELFIA®; Wallac, Turku, Finland). The detection limit of the assay is 0.2 µg/L, and the inter-assay coefficient of variation is 3% in the concentration range 3 - 90 µg/L (total CV 4%) **CEA threshold** 5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Reference standard** Clinical follow-up
Flow and timing	**Timing of CEA vs reference standard (days)** Unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Unclear
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		

Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Unclear		

Carriquiry 1999

Study characteristics	
Patient sampling	**Country** Uruguay **Study design** Retrospective casenote review. **Setting** Hospital **Dates of data collection** 1985 - 1998 **Population (n)** 209 **Inclusion criteria** Histologically proven colorectal carcinoma, 3 postoperative CEA measurements, minimum period of follow-up 24 months **Exclusion criteria** Postoperative death and Stage IV (unless radical resection of synchronous liver metastases), no preop CEA **Participants Included (n)** 142
Patient characteristics and setting	**Age range** 30 - 91 yrs **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** TNM staging system: 32 patients had Stage I, 57 had Stage II, 86 had Stage III, and 27 had Stage IV disease **Perioperative Investigations done to ensure no residual disease**

	N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 52 **Site of recurrences** N/R
Index tests	**CEA timing** CEA 3-monthly for 24 months, 4-monthly for yrs 3 - 4, and once per year after this (strict adherence in only 42 patients) **CEA technique** Serum concentrations of CEA were determined by a standard commercially-available immunoenzymatic assay **CEA threshold** 5 µg/L **Definition of positive** 2 consecutive values above 5 regarded abnormal; repeated at 2 - 4 weeks **Which CEA value (s) used?** Repeated value at the time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** CEA 3-monthly for 24 months, 4-monthly for yrs 3 - 4, and once per year after this (strict adherence in only 42 patients). Clinical follow-up, rectoscopy and/or colonoscopy at 1 yr and 3 yrs **Reference standard** USS/MRI/CT indicated on basis of raised CEA or clinical suspicion. CEA second-look surgery never used
Flow and timing	**Timing of CEA vs reference standard (days)** as per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low

DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Yes		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		

Carriquiry 1999 (*Continued*)

Did all patients receive a reference standard?	Yes		

Deveney 1984

Study characteristics	
Patient sampling	**Country** USA **Study design** Prospective **Setting** Hospital **Dates of data collection** starting in 1978 **Population (n)** N/R **Inclusion criteria** Resection for curable adenocarcinoma of the colon or rectum **Exclusion criteria** Dukes D **Participants included (n)** 65
Patient characteristics and setting	**Age range** 67 yrs mean **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** 8 Dukes A tumours, 34 had Dukes B tumours, and 20 had Dukes C tumours **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Some got radio/chemotherapy **Recurrences (n)** 23 **Site of recurrences** N/R
Index tests	**CEA timing** 3-monthly in yr 1, then 6-monthly to year 5 **CEA technique** N/R **CEA threshold**

	5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** 3-monthly in year 1, then 6-monthly to year 5: clinical history, examination, FOBT, LFT, CEA. 6-monthly CXR, CT abdomen, total colonoscopy. BE at 6 and 12 months then annually **Reference standard** A positive finding on any test prompted additional confirmatory tests, including laparotomy, thoracotomy,or percutaneous CT-directed biopsy
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		

Deveney 1984 (*Continued*)

Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Engarås 2003

Study characteristics	
Patient sampling	**Country** Sweden **Study design**

	Prospective **Setting** Hospital **Dates of data collection** 1998 - 1990 **Population (n)** 151 **Inclusion criteria** Surgery with curative intent with 5 years follow-up **Exclusion criteria** N/R **Participants Included (n)** 132
Patient characteristics and setting	**Age range** 27 - 75 **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Duke A 11, B 76, C 43,D 1, Undefined 1 **Perioperative investigations done to ensure no residual disease** Not specified **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 39 **Site of recurrences** N/R
Index tests	**CEA timing** Monthly during year 1 and then at 18 and 24 months **CEA technique** Delfia® test kits (Wallac Oy, Turku, Finland). The accuracy of the assays was assessed by analysis of 2 control samples in each assay and by measurement of the coefficient of variation by duplicate analyses of the samples **CEA threshold** 5.6 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** Monthly outpatient clinic visit during year 1, serum tests monthly during year 1, then 18 and 24 months. Clinical examinations at 1 year and 2 year with CXR, Sigmoidoscopy, BE, and CT Liver **Reference standard** Radiologic and/or endoscopic investigations at surgery or post mortem

Flow and timing	**Timing of CEA vs reference standard (days)** as per follow-up schedule		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Engaråas 2003 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Farinon 1980

Study characteristics	
Patient sampling	**Country** Italy **Study design** Retrospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 87 **Inclusion criteria** Preoperative CEA test > 6, operated in with end-to-end anastomosis **Exclusion criteria** N/R **Participants included (n)** 35

Farinon 1980 (*Continued*)

Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 3, B 26, C 6 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 10 **Site of recurrences** N/R
Index tests	**CEA timing** 3 monthly **CEA technique** CEA radioimmunoassay direct method **CEA threshold** 6 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** CEA and colonoscopy every 3 months **Reference standard** Second look surgery if not clear from CEA + colonoscopy.
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		

Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Farinon 1980 (*Continued*)

Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Fezoulidis 1987

Study characteristics	
Patient sampling	**Country** Germany **Study design** Prospective **Setting** Hospital **Dates of data collection** 1984 - 1986 **Population (n)** 48 **Inclusion criteria** radical surgery **Exclusion criteria** No exclusion criteria were defined; results from all 48 participants are included in the study **Participants Included (n)** 48
Patient characteristics and setting	**Age range** Study does not describe age bands; median age is 64 **Smoking status** N/R **Site of primary tumour** Rectum **Stage of primary tumour** Dukes A 9, Dukes B 16, Dukes C1 19, Dukes C2 4 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 5 **Site of recurrences**

	5 local rectal recurrences
Index tests	**CEA timing** 6 weeks, (then 3-monthly); main text of the study only mentions that CEA was measured postoperatively; ?Only once; it is not clear if there was a sequence of measurements. Table 4 looks more like a one-off **CEA technique** Unknown **CEA threshold** 2.5 µg/L **Definition of positive** Unclear **Which CEA value (s) used?** Probably 4 - 6 weeks postoperatively
Target condition and reference standard(s)	**Follow-up schedule** 4 - 6 weeks postoperatively, then 3-monthly clinical examination, CT, and CEA **Reference standard** 4 patients underwent CT guided biopsy but at unknown stage
Flow and timing	**Timing of CEA vs reference standard (days)** Unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		

Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Fucini 1987

Study characteristics	
Patient sampling	**Country** Italy **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1979 - 1983 **Population (n)** 64 **Inclusion criteria** Potentially curative surgery **Exclusion criteria** Died or demonstrated recurrence before 1982 (introduction of TPA and CA19-9 assays) **Participants Included (n)** 52
Patient characteristics and setting	**Age range** 40 - 77 **Smoking status** 1 smoker **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A: 28, B 17, C19 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 10 **Site of recurrences** N/R
Index tests	**CEA timing** As per protocol, then repeated within 2 weeks considered positive **CEA technique** Double antibody method (CEA-PR, Sorin Biomedica) **CEA threshold** 20 (95% control group) **Definition of positive** 2 consecutive samples **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** CEA + TPA + CA19-9, clinical exam at 3, 7, 14 days then 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 54, 60 months. Blood count at 3, 6, 12, 18, 24, 26, 48, 60 months. Liver USS at 3, 6, 18, 30, 36,

	42, 48, 54, 60 months. CXR 3, 6, 12, 18, 24, 30, 36, 48, 60 months. DCBE at 18, 42, 60 months. Colonoscopy 6, 12, 24, 36, 48, 60 months. APCT 12, 24 months. Random perineal percutaneous needle biopsy (rectal cancer) 6, 12, 18, 24, 36, 48, 60		
Flow and timing	**Timing of CEA vs reference standard (days)** Sensitivity uses CEA at the time of recurrence, specificity uses CEA over threshold at any time during follow-up		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	No		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			

Fucini 1987 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Graffner 1985

Study characteristics	
Patient sampling	**Country** Sweden **Study design** Prospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 190 **Inclusion criteria** Curative resection, age able to attend follow-up **Exclusion criteria**

	Moved from area, died of intercurrent illness, did not follow the schedule **Participants included (n)** 167
Patient characteristics and setting	**Age range** 55 - 74 **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 24, B 89, C 77 **Perioperative Investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 47 **Site of recurrences** Liver 18, anastomotic 4, perineal 7, lungs 4, skin 5, multiple organs 8, skeleton 1
Index tests	**CEA timing** CEA every second month during the first 2 years and every third month thereafter **CEA technique** Radioimmunoassay **CEA threshold** Abnormal blood values (CEA used same method as Colleen et al 1979 "the reference value was calculated from serum sampled from 89 apparently healthy persons aged 25 to 69 years. It was 10+/- 2.5 ug/l (mean+/-S.D)") or a rise of CEA levels within the normal range of more than 50% **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** CEA, ESR, haemoglobin, ALP, glutamyltranspeptidase (GGT), orosomucoid, alpha-antitrypsin, and haptoglobin every second month during the first 2 years and every third month thereafter. Physical exam and rectoscopy 3, 6, 9, 12, 18, 24, 36, 48, 60 months. DCBE and CXR 12, 36, 60 months **Reference standard** CXR, CT liver, CT perineum, endoscopic investigation of anastomosis, DCBE, angiography and bone scintography in selected cases
Flow and timing	**Timing of CEA vs reference standard (days)** if abnormal CEA detected reference standard triggered
Comparative	
Notes	

Graffner 1985 (*Continued*)

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low

Graffner 1985 (*Continued*)

DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Hara 2008

Study characteristics	
Patient sampling	**Country** Japan **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1990 - 2000 **Population (n)** 680 **Inclusion criteria** Curative resection, dukes C **Exclusion criteria** Multiple cancers, insufficient examinations, persistent post-op CEA, and SCC, randomised to pretest probability group **Participants Included (n)** 174
Patient characteristics and setting	**Age range** 60.6 ± 11.1 **Smoking status** N/R **Site of primary tumour** Colorectal

	Stage of primary tumour All Dukes C- Stage 1 18, 2 59, 3 232, 4 39 **Perioperative investigations done to ensure no residual disease** Persistent CEA elevation excluded **Chemotherapy/radiotherapy?** No **Recurrences (n)** 51 **Site of recurrences** N/R
Index tests	**CEA timing** 3-monthly **CEA technique** N/R **CEA threshold** 5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** All patients were followed for more than 5 years or until death with routine serum CEA examination every 3 months. USS and/or CT and CXR examinations were performed every 3 - 6 months **Reference standard** Additional imaging was performed in patients with elevated postoperative CEA levels to determine whether recurrence was present
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low

DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		

Did all patients receive a reference standard?	Yes		

Hara 2010

Study characteristics	
Patient sampling	**Country** Japan **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1990 - 2004 **Population (n)** 488 **Inclusion criteria** Stage II or III curative resection **Exclusion criteria** Patients with squamous cell, carcinoma, more than one cancer, or insufficient follow-up **Participants Included (n)** Stage II: 167 Stage III: 136
Patient characteristics and setting	**Age range** Stage II: 68.3 ± 10.5 (38 - 92) Stage III: 63.4 ± 9.4 (44 - 88) **Smoking status** N/R **Site of primary tumour** Stage II: Colon 112, rectum 55 Stage III: Colon 89, rectum 47 **Stage of primary tumour** Stage II: Depth T1 0, 2 0, 3 142, 4 23 Stage III: Depth T1 3, 2 89, 3 32, 4 12 **Perioperative investigations done to ensure no residual disease** Not specified **Chemotherapy/radiotherapy?** No **Recurrences (n)** Stage II: 23 Stage III: 51 **Site of recurrences** N/R

Hara 2010 (*Continued*)

Index tests	**CEA timing** Unclear **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R
Target condition and reference standard(s)	**Follow-up schedule** All patients underwent routine serum CEA assays and radiological examination
Flow and timing	**Timing of CEA vs reference standard (days)** unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		

Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Unclear		

Hine 1984

Study characteristics	
Patient sampling	**Country** UK **Study design** Prospective

	Setting Hospital **Dates of data collection** N/R **Population (n)** 663 **Inclusion criteria** Radical surgery for colorectal cancer **Exclusion criteria** SCC anus, tumours in the appendix. 6 were lost to clinical follow-up and 5 others were removed from the trial. Removal followed the development of unassociated conditions such as alcoholic cirrhosis which interfered with the interpretation of a significant CEA rise (3 patients) and in 2 patients the onset of psychiatric illness made the use of cancer chemotherapy inadvisable **Participants Included (n)** 626
Patient characteristics and setting	**Age range** 59 **Smoking status** Unknown **Site of primary tumour** 290 rectum, 373 colon **Stage of primary tumour** A in 38, B in 377 and C in 248 **Perioperative investigations done to ensure no residual disease** Not specified **Chemotherapy/radiotherapy?** Patients with at least 2 progressively rising CEA values of > 35 ngml-1 but no other definite evidence of recurrent malignancy were randomised in a prospective trial of cytotoxic therapy **Recurrences (n)** 171 **Site of recurrences** N/R
Index tests	**CEA Timing** At each follow-up visit **CEA Technique** CEA was measured in the unextracted serum by a double antibody radio-immunoassay as developed by Egan et al. (1972) and adapted by Laurence et al. (1972). The inter- and intra-assay variation of the method was found to be < 10%. An upper limit of 15 µg/L will include 99% of a normal population and in the present study a level of > 20 µg/L was regarded as abnormal **CEA threshold** 20 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All

Target condition and reference standard(s)	**Follow-up schedule** 3-monthly for for first 2 postoperative years, then 6 - 12-monthly depending on the surgeon. Full clinical examination including sigmoidoscopy was performed **Reference standard** recurrence was primarily made on the basis of symptoms and signs of disease confirmed by other investigations when indicated (e.g. liver scan, bone scan, biopsy). Thorough clinical examination including sigmoidoscopy. If this indicated recurrent malignancy, confirmatory investigations were ordered and management was initiated appropriate to the results. When clinical examination failed to reveal malignancy, the subsequent course of events depended on the degree of elevation of the CEA. If the level was >20ngml-1 but <35ngml- 1, the test was repeated at monthly intervals until it fell below 20ngml-1 or rose above 35ngml-1. All patients with levels >35ngml-1 and no clinical evidence of recurrence had a further CEA estimation, full blood count, erythrocyte sedimentation rate, liver function tests, barium enema, chest X-ray and isotope and/or ultrasound liver scan, together with bone scan and colonoscopy where indicated. If recurrence was diagnosed from the results of these investigations then appropriate management was instituted.
Flow and timing	**Timing of CEA vs reference standard (days)** Raised CEAs were recalled to clinic within 2 months of the date of the first sample for clinical exam and sigmoidoscopoy. If no recurrence found intensified frequency of testing whilst in the 20 - 35 range. If > 35 but no signs of recurrence, then chemotherapy
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		

Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Unclear		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Unclear		

Irvine 2007

Study characteristics	
Patient sampling	**Country** UK **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1996 - 2000 **Population (n)** 150 **Inclusion criteria** Curative surgery for colorectal cancer **Exclusion criteria** Palliative patients, non-operative patients, 11 who developed metastases or recurrences within 3 months of surgery, persistently elevated CEA postoperatively (deemed non-curative resection) **Participants Included (n)** 139
Patient characteristics and setting	**Age range** 22 - 87 **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 10, B 82, C 47 **Perioperative investigations done to ensure no residual disease** Development of metastases or recurrence within 3 months of surgery, persistently elevated CEA postoperatively **Chemotherapy/radiotherapy?** No **Recurrences (n)** 46 **Site of recurrences** N/R
Index tests	**CEA timing** Postoperatively 3-monthly for 2 yrs, then 6-monthly to 5 yrs. The CEA measurements for each patient were analysed twice, once looking for a small rise in CEA and again looking for a CEA value that rose above the traditional normal limit (10 µg/L) **CEA technique** Bayer immunoassay, which at the levels in this study has an error rate of 2.3% **CEA threshold** 10 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence

Irvine 2007 (*Continued*)

Target condition and reference standard(s)	**Follow-up schedule** 6-monthly CT for 2 years, plus CEA 3-monthly for 2 years, then 6-monthly to 5 years		
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Yes		
			Low
DOMAIN 3: Reference Standard			

Irvine 2007 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Johnson 1985

Study characteristics	
Patient sampling	**Country** Norway **Study design** Propsective **Setting** Hosptial + primary care **Dates of data collection** N/R **Population (n)** 93 **Inclusion criteria** Radical treatment for colorectal cancer **Exclusion criteria**

	Palliative, new cancers, no CEA monitoring **Participants included (n)** 51
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colon 49, rectal 44 **Stage of primary tumour** Dukes A 28, B 27, C 21, palliative 17 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 15 **Site of recurrences** N/R
Index tests	**CEA timing** Postoperatively, then at 3 - 4-monthly intervals **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R **More data available?** N/R
Target condition and reference standard(s)	**Follow-up schedule** Postoperatively, then at 3 - 4 monthly intervals, rising CEA resulted in further investigation, general clinical investigations, angiography of the liver, resection. No fixed schedule
Flow and timing	**Timing of CEA vs reference standard (days)** CEA triggered investigation
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Johnson 1985 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		

Johnson 1985 (*Continued*)

Were all patients included in the analysis?	Unclear		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Unclear		

Jubert 1978

Study characteristics	
Patient sampling	**Country** USA **Study design** Retrospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 97 **Inclusion criteria** Colorectal cancer **Exclusion criteria** N/R **Participants Included (n)** 97
Patient characteristics and setting	**Age range** 65 mean (39 - 89) **Smoking status** Unknown **Site of primary tumour** Colon 56, rectum 41 **Stage of primary tumour** Dukes A 10, B 42, C 34, D 6 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** 5 chemo, 5 immuno

Jubert 1978 (*Continued*)

	Recurrences (n) 20 **Site of recurrences** 7 liver, 13 non-liver
Index tests	**CEA timing** At 6-week intervals postoperatively **CEA technique** N/R **CEA threshold** 2.5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** CEA is done preoperatively and at six week intervals postoperatively. In addition, patients are evaluated postoperatively at 6 to 8 week intervals by physical examination and the usual laboratory and radiological tests, and where indicated, suspicions of recurrence and/or metastasis are documented histologically for the most part **Reference standard** "suspicions of recurrence and/or metastasis are documented histologically for the most part"
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		

Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Unclear		

Kanellos 2006a

Study characteristics	
Patient sampling	**Country** Greece **Study design** Prospective **Setting** Hospital **Dates of data collection** 1991 - 1999 **Population (n)** N/R **Inclusion criteria** Histologically proven colorectal cancer, no detectable liver metastasis, curative surgery for colorectal cancer **Exclusion criteria** Confirmed liver metastasis, peritoneal carcinomatosis, ascites, emergency surgery for obstruction or perforation, smokers, obstructive biliary disease or biliary surgery, or refused consent **Participants Included (n)** 73
Patient characteristics and setting	**Age range** 64.2 (SD: 9.7) **Smoking status** Non-smokers **Site of primary tumour** Colorectal **Stage of primary tumour** Stage I 14, II 37, III 22 **Perioperative investigations done to ensure no residual disease** Pre-op abdominal CT, intraoperative liver palpation to exclude liver metastases **Chemotherapy/radiotherapy?** 22 patients with stage III cancer had adjuvant chemo **Recurrences (n)** 10 **Site of recurrences** N/R
Index tests	**CEA timing** 3-monthly to 3 yrs, the 6-monthly to 5 yrs **CEA technique** Monoclonal antibody technique, using a solid-phase 2-site mouse monoclonal antibody radioimmunoassay kit **CEA threshold** 5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence

Target condition and reference standard(s)	**Follow-up schedule** Every 3 months for the first 3 years and every 6 months thereafter: clinical examination routine biochemical analysis, CXR, and CT		
Flow and timing	**Timing of CEA vs reference standard (days)** Simultaneous, per protocol.		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Kato 1980

Study characteristics	
Patient sampling	**Country** Japan **Study design** Prospective **Setting** Hospital **Dates of data collection** 1977 - 79 **Population (n)** N/R **Inclusion criteria** Surgically treated for adenocarcinoma of the colon or rectum with curative intent **Exclusion criteria**

	Incomplete CEA dataset **Participants Included (n)** 129
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A,B,C **Perioperative investigations done to ensure no residual disease** Not specified **Chemotherapy/radiotherapy?** No **Recurrences (n)** 32 **Site of recurrences** N/R
Index tests	**CEA timing** Unclear **CEA technique** RIA kit by Dynabot **CEA threshold** 2.5 and 5 µg/L **Definition of positive** N/R **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** N/R
Flow and timing	**Timing of CEA vs reference standard (days)** Unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Unclear		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		

Kato 1980 (*Continued*)

Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Unclear		

Kim 2013

Study characteristics	
Patient sampling	**Country** Korea **Study design** Retrospective **Setting** Hospital **Dates of data collection** 2005 - 2009 **Population (n)** N/R **Inclusion criteria** Radical resection **Exclusion criteria** Patients with stage 0, I or IV cancer, insufficient follow-up (less than 3 years), abnormal CEA in the first measurement after surgery (checked within three months after surgery), history of other cancers and/or history of preoperative concurrent chemoradiation therapy were excluded **Participants Included (n)** 336
Patient characteristics and setting	**Age range** Stage 111: 29 - 81, Stage 11: 33 - 83 **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Stage II 189, Stage III 147 **Perioperative investigations done to ensure no residual disease** N/R

	Chemotherapy/radiotherapy? N/R **Recurrences (n)** 79 **Site of recurrences**
Index tests	**CEA timing** CEA levels were assayed with a 3-month interval for the first 2 years and every 6 months thereafter **CEA technique** Immunoassay method (ADIVA Centaur XP immunoassay system, Siemen AG, Erlangen, Germany) **CEA threshold** 5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** CEA levels were assayed with a 3-month interval for the first 2 years and every 6 months thereafter. Chest CT and abdomino-pelvic CT were performed with a 6-month interval for the first 2 years and every year thereafter **Reference standard** The diagnosis of a tumour recurrence was confirmed by biopsy and radiologic evidence
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Kohler 1980

Study characteristics	
Patient sampling	**Country** USA **Study design** Retrospective casenote review. **Setting** Hospital **Dates of data collection** 1971 - 1974 **Population (n)** 144 **Inclusion criteria** Surgically confirmed adenocarcinoma of colon or rectum **Exclusion criteria** N/R **Participants Included (n)** 49
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 22 **Site of recurrences** N/R
Index tests	**CEA timing** Not clear **CEA technique** Hansens radioimmunoassay **CEA threshold** 2.5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** N/R

Flow and timing	**Timing of CEA vs reference standard (days)** N/R		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Kohler 1980 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	No		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Unclear		

Koizumi 1992

Study characteristics	
Patient sampling	**Country** Japan **Study design** Cross-sectional with follow-up of cases **Setting** Hospital **Dates of data collection** 1986 - 1990 **Population (n)** 194 **Inclusion criteria** Unclear **Exclusion criteria** Cases undergoing operation later, benign colorectal disease. **Participants Included (n)** 77

Patient characteristics and setting	**Age range** 32 - 83 **Smoking status** Unknown **Site of primary tumour** Colorectal **Stage of primary tumour** Unknown **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Follow-up schedule** N/R **Recurrences (n)** 34 **Site of recurrences** N/R
Index tests	**CEA timing** N/R **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Reference standard** Unclear
Flow and timing	**Timing of CEA vs reference standard (days)** N/R
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Did the study avoid inappropriate exclusions?	Unclear		
			Unclear
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Koizumi 1992 (*Continued*)

Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Unclear		

Korner 2007

Study characteristics	
Patient sampling	**Country** Norway **Study design** Prospective cohort with retrospective sampling **Setting** Hospital **Dates of data collection** 1996 - 1999 **Population (n)** 314 **Inclusion criteria** Surgically treated for adenocarcinoma of the colon or rectum with curative intent, age < 75 yrs, national guidelines followed **Exclusion criteria** Not systematically followed up for 5 years or until recurrence, incomplete CEA dataset. Dukes D **Participants included (n)** 153
Patient characteristics and setting	**Age range** < 75 **Smoking status** N/R **Site of primary tumour** Colon 102, rectum 50 **Stage of primary tumour** Dukes A 31, B 79, C 42 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 37

	Site of recurrences N/R
Index tests	**CEA timing** CEA 3, 6, 9, 12, 18, 24, 30, 36, 42, 48, 54, 60 months **CEA technique** Immunoassay kit from Abbot diagnostic IL, USA **CEA threshold** 4 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** CEA 3, 6, 9, 12, 18, 24, 30, 36, 42, 48, 54, 60 months. USS Liver & CXR 6, 12, 18, 24, 30, 36, 42, 48, 54, 60 months. Colonoscopy 12, 60 months **Reference standard** Biopsy and/or imaging studies to confirm recurrence, or disease-free interval of 60 months without proof of recurrence
Flow and timing	**Timing of CEA vs reference standard (days)** not specified if different from protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measure-	Yes		

ments?			
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Li Destri 1998

Study characteristics	
Patient sampling	**Country** Italy **Study design** Retrospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 364 **Inclusion criteria** Radical surgery for colorectal cancer CEA measured postoperatively **Exclusion criteria** N/R **Participants included (n)** 239
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 45 **Site of recurrences** hepatic 18, non-hepatic 22, mixed 5
Index tests	**CEA timing** CEA monitoring, conducted every 3 months for years 1, 2, and 3, every 6 months for years 4 and 5, then yearly up to year 10 **CEA technique** The antigen was determined using the radioimmunoassay method **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R

Li Destri 1998 (*Continued*)

Target condition and reference standard(s)	**Follow-up schedule** CEA monitoring, conducted every 3 months for years 1, 2, and 3, every 6 months for years 4 and 5, then yearly up to year 10		
Flow and timing	**Timing of CEA vs reference standard (days)** N/R		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Unclear		

Lucha 1997

Study characteristics	
Patient sampling	**Country** USA **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1981 - 1985 **Population (n)** N/R **Inclusion criteria** Newly diagnosed colorectal cancer undergoing operative resection for cure (Astler Coller A,B,C) **Exclusion criteria**

	Metastatic disease and synchronous cancers **Participants Included (n)** 285
Patient characteristics and setting	**Age range** 66.8 (range, 31 - 96) **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Astler-Coller Stage A 39, B1 57, B2 109, C1 15, C2 60 **Perioperative investigations done to ensure no residual disease** Intraoperative criteria for curative resection included absence of gross residual disease **Chemotherapy/radiotherapy?** No **Recurrences (n)** 66 **Site of recurrences** N/R
Index tests	**CEA timing** 2-monthly for 2 years, 3-monthly for year 3, 6-monthly for years 4 - 5, annually afterwards. A repeat CEA was performed in patients who had an abnormal rise **CEA technique** Abbott **CEA threshold** 5 µg/L **Definition of positive** 2 consecutive samples **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Follow-up schedule** 2 monthly for 2 years, 3 monthly for year 3, 6 monthly for years 4 and 5, annually afterwards. A detailed history and physical examination was performed, and CEA levels were monitored at each encounter **Reference standard** Two successive CEA elevations were investigated with diagnostic imaging and / or endoscopy when indicated
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	
Methodological quality	

Lucha 1997 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			

Lucha 1997 (*Continued*)

Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	No		

Luporini 1979

Study characteristics	
Patient sampling	**Country** Italy **Study design** retrospective **Setting** Hospital **Dates of data collection** 1974 - 1976 **Population (n)** 204 **Inclusion criteria** Large bowel malignancies, radical resection **Exclusion criteria** N/R **Participants Included (n)** 198
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Large intestine **Stage of primary tumour** Dukes A - B 11, C1 39, C2 30, CH (liver involvement) 32 **Perioperative investigations done to ensure no residual disease**

Luporini 1979 (*Continued*)

	N/R **Chemotherapy/radiotherapy?** Yes **Recurrences (n)** 62 **Site of recurrences** N/R
Index tests	**CEA timing** N/R **CEA technique** N/R **CEA threshold** 5 µg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R **More data available?** N/R
Target condition and reference standard(s)	**Follow-up schedule** N/R
Flow and timing	**Timing of CEA vs reference standard (days)** Unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Unclear		

Mach 1978

Study characteristics	
Patient sampling	**Country** Switzerland **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1977 - 1978 **Population (n)** 200 **Inclusion criteria** Histologically confirmed diagnosis of adenocarcinoma of colon or rectum **Exclusion criteria** Incomplete tumour resection **Participants Included (n)** 66
Patient characteristics and setting	**Age range** 65 **Smoking status** 12 patients who had CEA levels fluctuating around the normal limit of 5 ng/ml during the last 2 or 3 years without a definite rise of CEA levels and also without clinical evidence of tumour relapse. Among them were 6 heavy smokers **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes ABCD **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** 2 of the recurrences were reported to have chemo **Recurrences (n)** 19 **Site of recurrences** N/R
Index tests	**CEA timing** 3-monthly **CEA technique** The radioimmunoassay of CEA was performed according to the method of Goldz as modified by Mach el al. The major modification was that duplicates of 1 ml of plasma (10 ml of blood was collected in tubes containing 33 mg of dry E.D.T.A. K3) instead of 5 ml of serum, were extracted in perchloric acid. The sensitivity of the test is 1 µg/L. The normal value determined in 90 nonsmoking blood bank donors, unselected for age and sex, ranged between 0 to 3.5 µg/L. Our CEA assay is similar to the Hansen method,'but our numerical values are slightly higher and should be divided by a factor of 1.5 in order to make a direct comparison **CEA threshold** 5 µg/L

Mach 1978 (*Continued*)

	Definition of positive 1 elevated value **Which CEA value (s) used?** All		
Target condition and reference standard(s)	**Follow up schedule** N/R		
Flow and timing	**Timing of CEA vs reference standard (days)** N/R		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		

Mach 1978 (*Continued*)

			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Unclear		

Mackay 1974

Study characteristics	
Patient sampling	**Country** UK **Study design** Prospective **Setting** Hospital **Dates of data collection** Approx 1970 - 1973 **Population (n)**

	N/R **Inclusion criteria** Surgically resected colorectal carcinoma (a) Their operations were considered to be clinically curative. (b) Pathological staging showed the carcinoma to fall into Dukes (1950) A, B, or C category. (c) The participants had been followed up for at least 12 months and most for 24 months either after the operation or after the first plasma CEA assay **Exclusion criteria** Inadequate follow-up time or because the plasma CEA values had risen temporarily to or remained at levels between 20 and 40 µg/L **Participants included (n)** 220
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Duke ABC **Perioperative investigations done to ensure no residual disease** Unclear **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 53 **Site of recurrences** Liver 31, lung 3, peritoneum and pelvis 17, bones 2, local 6, skin 2
Index tests	**CEA timing** 3 monthly **CEA technique** Double-antibody radioimmunoassay **CEA threshold** 40 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** N/R **Reference standard** Recurrence of tumour was detected clinically or by radioisotope scanning or other radiographic techniques
Flow and timing	**Timing of CEA vs reference standard (days)** Reference standard triggered by a rise in CEA

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge	No		

Mackay 1974 (*Continued*)

of the results of the index tests?			
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

Mariani 1980

Study characteristics	
Patient sampling	**Country** Italy **Study design** Prospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** N/R **Inclusion criteria** Histologically confirmed adenocarcinoma submitted for resection (included ± pre-op measurements) **Exclusion criteria** Heavy smokers (> 15 cigarettes/day) and patients with known, or suspected alcoholic hepatitis **Participants included (n)** 69
Patient characteristics and setting	**Age range** 60.2 ± 11.6 yrs **Smoking status**

	Excluded **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 5, B 18, C 14, D 2. **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 24 **Site of recurrences** N/R
Index tests	**CEA timing** The 4th and 14th day after surgery. Subsequent blood samples were taken at regular intervals (every 2 - 3 months) in the following 12 - 20 months. Moreover, an increased CEA value was always confirmed by repeated assays of the same sample, and by assaying an additional sample obtained from the same patient **CEA technique** Radioimmunoassay (RIA), using commercial EAK kits (purchased through SORIN Biomedica, Saluggia, Italy) **CEA threshold** 10 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** All patients had a blood sample taken for CEA assay preoperatively, then at the 4th and 14th day after surgery. Subsequent blood samples were taken at regular intervals (every 2-3 months) in the following 12-20 months with follow-up examinations; the complete work-up of the patients included physical examination, chest standard X-ray, recto-sigmoidoscopy, liver scan, hemogram and liver function tests; barium enema and bone scan were performed when indicated
Flow and timing	**Timing of CEA vs reference standard (days)** not specified
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		

Mariani 1980 (*Continued*)

Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

McCall 1994

Study characteristics	
Patient sampling	**Country** Australia **Study design** Prospective RCT **Setting** Hospital **Dates of data collection** 1984 - 1990 **Population (n)** 328 **Inclusion criteria** curative resection of colorectal cancers **Exclusion criteria** Patients with metastatic disease at presentation and those who for geographic or medical reasons were not able to be followed were excluded from the trial. Less than two years follow-up completed (16 patients: 10 died of unrelated causes; 6 withdrew consent or were lost to follow-up) and failure to obtain CEA levels (one patient) **Participants Included (n)** 311
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes ABC **Perioperative investigations done to ensure no residual disease**

	N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 98 **Site of recurrences** N/R
Index tests	**CEA timing** Patients entered into both arms of the study had serum CEA levels measured for 5 consecutive years: every 3 months for the first 2 years, then every 6 months for the next 3 years **CEA technique** Enzyme immunoassay method (Abbott Laboratories, North Chicago, IL) **CEA threshold** 5 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** Standard follow up: Clinical review plus CEA, Liver function, and fecal occult blood - 3 monthly til 2 years, 6 monthly til 5 years. CXR, Liver CT, Colonoscopy at 0 and 5 years; Aggressive follow up: As for standard follow-up plus CXR , Liver CT and Colonoscopy annually **Reference standard** Radiology, histology
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			

If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		

Did all patients receive a reference standard?	Yes		

Miles 1995

Study characteristics	
Patient sampling	**Country** Scotland **Study design** Retrospective notes review **Setting** Hospital **Dates of data collection** 1988 - 1992 **Population (n)** 265 **Inclusion criteria** Patients who underwent a resection, with curative intent. **Exclusion criteria** Patients were excluded where, on inspection of the patients' notes, it was found that primary surgery was palliative, follow-up was incomplete or there were fewer than 1 preoperative and 2 postoperative carcinoembryonic antigen level estimations **Participants included (n)** 125
Patient characteristics and setting	**Age range** 69 (41 - 90) **Smoking status** Unknown **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 10, B 27, C 38, D 22, unknown 27 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 53 **Site of recurrences** N/R
Index tests	**CEA timing** Not clear **CEA technique** Using international standard International Reference Preparation 73/601, National Institute for

	Biological Standards and Control **CEA threshold** 10 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** History is recorded and clinical examination (including rectal examination and rigid sigmoidoscopy) , faecal occult blood test and estimation of carcinoembryonic antigen level are undertaken **Reference standard** The presence of recurrent disease is confirmed by clinical examination, colonoscopy, biopsy, chest radiography, ultrasonography, computerized axial tomography scanning and laparotomy
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol, CEA triggers reference standard.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		

Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

Minton 1985

Study characteristics	
Patient sampling	**Country** USA **Study design** Prospective

	Setting Hospital **Dates of data collection** 1978 - 1983 **Population (n)** 400 **Inclusion criteria** post-colorectal cancer resection **Exclusion criteria** N/R **Participants included (n)** 400
Patient characteristics and setting	**Age range** 58 (18 - 84) **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes A 17, B1 91, B2 31, C1 119, C2 122, D 6, unknown 6 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 130 **Site of recurrences** Liver 49, Anastomosis site or mesentery of bowel 26, peritoneum 7, pelvis 6, para-aortic nodes 2, mesentric nodes 2, multiple 7, other 28, no disease found 3
Index tests	**CEA timing** CEA performed every 2 months for the first 2 years, and then every 4 months for the next 3 years. To rule out laboratory variations, a repeat CEA value was required to confirm an abnormal CEA elevation **CEA technique** N/R **CEA threshold** 2.5 µg/L **Definition of positive** Abnormal repeated **Which CEA value (s) used?** Unclear
Target condition and reference standard(s)	**Follow-up schedule** Patients were evaluated postoperatively with each surgeon's customary follow-up procedures and frequency of CEA determinations **Reference standard** Second-look surgery was performed on any potentially resectable recurrent cancer discovered by physical examination or symptoms of bowel or ureteral obstruction, gastrointestinal bleeding, or

	findings from rectal, vaginal, or colostomy examinations. In addition, second-look surgery was done when a persistently rising CEA value was detected. Before the second-look procedure was performed a careful physical examination complemented by chest roentgenogram, bone and brain scans, and appropriate gastrointestinal and genitourinary roentgenograms was done to rule out the possibility of unresectable metastases. A computerized axial tomography (CAT) scan of the abdomen was not required, but was considered appropriate for institutions with that capability
Flow and timing	**Timing of CEA vs reference standard (days)** not specified
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	No		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low

Minton 1985 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Unclear		

Mittal 2011

Study characteristics	
Patient sampling	**Country** India **Study design** Retrospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 73 **Inclusion criteria**

	Histologically proven postoperative CRC resection undergoing PET/CT and conventional imaging to detect suspected recurrence triggered by a rising CEA **Exclusion criteria** N/R **Participants included (n)** 73
Patient characteristics and setting	**Age range** 25 - 80 **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 38 **Site of recurrences** N/R
Index tests	**CEA timing** Within 7 - 10 days of imaging **CEA technique** Electro-chemiluminescent immunoassay **CEA threshold** 3 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At point of recurrence
Target condition and reference standard(s)	**Reference standard** PET/CT
Flow and timing	**Timing of CEA vs reference standard (days)** within 7-10 days of CEA
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		

Mittal 2011 (*Continued*)

Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Nishida 1988

Study characteristics	
Patient sampling	**Country** Japan **Study design** Prospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** N/R **Inclusion criteria** Surgically treated for adenocarcinoma of the colon or rectum with curative intent and CEA measurements **Exclusion criteria** incomplete CEA dataset **Participants included (n)** 66
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Stage I - V **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?**

	No **Follow-up schedule** N/R **Recurrences (n)** 20 **Site of recurrences** N/R
Index tests	**CEA timing** CEA 1 month **CEA technique** RIA kit by Dynabot **CEA threshold** 2.5 µg/L **Definition of positive** N/R **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Reference standard** N/R
Flow and timing	**Timing of CEA vs reference standard (days)** Unclear
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Unclear		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Unclear		

Ochoa-Figueroa 2012

Study characteristics	
Patient sampling	**Country** Spain **Study design** Retrospective **Setting** Hospital **Dates of data collection** 2007 - 2011 **Population (n)** 54 **Inclusion criteria** Referred to the Dept of Nuclear Medicine for FDG PET-CT with suspected CRC recurrence following surgical resection and posterior histological confirmation **Exclusion criteria** Not possible to follow up, mixed malignancy of the salivary gland **Participants Included (n)** 47
Patient characteristics and setting	**Age range** 63 (32 - 87) **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** 38 chemo, 9 chemo and radio **Recurrences (n)** 34 **Site of recurrences** N/R
Index tests	**CEA timing** CEA used as a marker of suspected recurrence or measured when recurrence suspected by CT **CEA technique** Radioimmunoanalysis **CEA threshold** 10 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** Single measurement taken at point of recurrence

Target condition and reference standard(s)	**Reference standard** Histopathology or Clinical evolution, FDG PET-CT		
Flow and timing	**Timing of CEA vs reference standard (days)** CEA prior to Referral; no more clear than this		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
			High
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Ohlsson 1995

Study characteristics	
Patient sampling	**Country** Sweden **Study design** RCT **Setting** Hospital **Dates of data collection** 1983 - 1986 **Population (n)** 107 **Inclusion criteria** Resection with curative intent, recruited to follow-up group **Exclusion criteria**

	Patients operated with local excision or having demonstrable distant metastases were excluded, as were patients in whom age or severe illness was considered to preclude treatment of recurrent disease. Other exclusion criteria were: Inability to cooperate, ulcerative colitis, Crohn's disease, familial polyposis, and incomplete colonoscopy together with uncertain findings at the barium enema examination **Participants Included (n)** 53
Patient characteristics and setting	**Age range** 65.7 (40.6 - 83.3) **Smoking status** N/R **Site of primary tumour** Rectum 19, colon 34 **Stage of primary tumour** Dukes A 10, B 21, C 22 **Perioperative investigations done to ensure no residual disease** Preoperative investigation included barium enema, pulmonary x-ray, and blood tests for liver function test, carcinoembryonic antigen and colonoscopy **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 17 **Site of recurrences** Local 11, liver 3, lung 3, peritoneum 2, ovary 1
Index tests	**CEA timing** 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 60 months **CEA technique** Not specified **CEA threshold** N/R **Definition of positive** N/R **Which CEA value (s) used?** N/R
Target condition and reference standard(s)	**Follow-up schedule** Physical examination, Rigid Proctosigmoidoscopy, Blood tests - CEA, ALP, GGT, Faecal Heamoglobin, CXR: 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 60 months. Endoscopic control of the anastomosis: 9, 21, 42 months. Colonoscopy: 3, 15, 30, 60 months. CT Pelvis: 3, 6, 12, 18, 24 months. **Reference standard** CT/ Endoscopy/colonoscopy
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol, immediate diagnostic work-up did not reveal the site of recurrence in 4 asymptomatic

	patients with raised CEA levels; in these patients the time interval between elevation of CEA and symptoms of tumour recurrence varied between 0.2 and 4.7 (median 0.5) years
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Ohlsson 1995 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Ohtsuka 2008

Study characteristics	
Patient sampling	**Country** Japan **Study design** Retrospective **Setting** Hospital **Dates of data collection** 2002 - 2005 **Population (n)** 138 **Inclusion criteria** Curative resection, stage 0 - III according to the General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum, and Anus, 7th edition, 2006, no residuals **Exclusion criteria** History of another malignancy before or after the operation, lost to follow-up **Participants Included (n)** 97

Patient characteristics and setting	**Age range** 70 (37 - 86) **Smoking status** Chronic benign disease or smoking in 46 cases **Site of primary tumour** 32 right colon, 32 left colon, 30 rectum, 3 multiple **Stage of primary tumour** 0 in 8, I in 12, II in 37, and III in 40 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Yes, but not described **Recurrences (n)** 22 **Site of recurrences**
Index tests	**CEA timing** Every 1 - 3 months during the initial 6 months after the operation, every 3 - 6 months from 6 months to 2 years, and every 6 - 12 months during 2 - 5 years after the operation **CEA technique** CEA, a latex immunoassay, Mitsubishi Chemical Ltd., Japan **CEA threshold** 5 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R
Target condition and reference standard(s)	**Follow-up schedule** the follow-up schedule of the tumour markers and physical examination after the operation were: every 1 - 3 months during the initial 6 months after the operation, every 3 - 6 months from 6 months to 2 years, and every 6 - 12 months during 2 - 5 years after the operation. Radiological examinations including abdominal ultrasonography, computed tomography (CT), chest X-ray, gastrointestinal series, and/or endoscopic evaluation were performed every 6 - 12 months during the follow-up period. Marker evaluations and physical/radiological examinations were performed at shorter-term intervals than those described above in patients with suspected recurrence, those undergoing chemotherapy, or in those demonstrating marker elevations **Reference standard** radiological examinations / histology
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol or reference standard triggered by rise in CEA
Comparative	
Notes	
Methodological quality	

Ohtsuka 2008 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			

Ohtsuka 2008 (*Continued*)

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Yes		

Park 2009

Study characteristics	
Patient sampling	**Country** Korea **Study design** Prospective **Setting** Hospital **Dates of data collection** N/R **Population (n)** 1707 **Inclusion criteria** curative resection for colorectal cancer followed by surveillance programme **Exclusion criteria** Patients with synchronous metastatic disease or patients undergoing palliative resection, and those with carcinoma in situ, inflammatory bowel disease, familial adenomatous polyposis or pathology other than adenocarcinoma were excluded, as were patients with T1 cancer treated by endoscopic mucosal resection or transanal excision. In addition, patients with chronic obstructive lung disease, chronic liver disease, peptic ulcer, and diabetes were excluded **Participants Included (n)** 1263
Patient characteristics and setting	**Age range** 61 (21 - 90) **Smoking status** N/R **Site of primary tumour**

	Colon 631, rectum 632 **Stage of primary tumour** I 212, II 514, III 537 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Yes, but not specified **Recurrences (n)** 291 **Site of recurrences** N/R
Index tests	**CEA timing** per schedule **CEA technique** N/R **CEA threshold** 7 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All, although at point of recurrence for 18.8%
Target condition and reference standard(s)	**Follow-up schedule** 2- or 3- month intervals for the first 2 years and at 6-month intervals thereafter. At each visit, CEA levels are assayed, a full history is obtained, and a physical examination is per- formed. A serum CEA assay is performed with at least a 2- week interval after the administration of chemotherapy. Colonoscopy is performed within 6 months to 1 year following surgery, and every 3years thereafter. Chest radiographs and abdominopelvic computed tomography (CT) are performed 6 months post-operatively and then at yearly intervals. Unscheduled CT or positron emission tomography (PET) scans were performed on patients with increased serum CEA concentrations or patients who were symptomatic **Reference standard** diagnosis of a tumour recurrence was confirmed by biopsy or examination of the resected specimen. Other- wise, tumour recurrence was documented from the first clinical or radiologic sign of disease that showed an unrelenting course leading to tumour progression and/or death. The criteria for establishment of recurrent disease included histologic confirmation, palpable disease, or radiographic evidence of disease with subsequent clinical progression and supportive biochemical data, particularly an increased CEA level
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	
Methodological quality	

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Peng 2013

Study characteristics	
Patient sampling	**Country** China **Study design** Retrospective comparative diagnostic accuracy study **Setting** Hospital **Dates of data collection** 2006 - 2012 **Population (n)** 128 **Inclusion criteria** Colorectal cancer with full response to primary surgery ± chemo, undergoing FDG-PET/CT for either elevated CEA levels or in patients with a suspicion of recurrence without CEA rise **Exclusion criteria** Unstable, severe DM, severe illness, 1 or more additional tumours, unable to remain supine for 30 mins **Participants Included (n)** 96
Patient characteristics and setting	**Age range** 61 (34 - 85) **Smoking status** N/R **Site of primary tumour** Colon 53, rectum 42 **Stage of primary tumour**

	0 in 1, I 15, II 31, III39, IV 9, unknown 1 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Yes, but not specified **Recurrences (n)** 63 **Site of recurrences** N/R
Index tests	**CEA timing** 3-monthly **CEA technique** N/R **CEA threshold** 5 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At time of recurrence
Target condition and reference standard(s)	**Reference standard** FDG-PET/CT +/- histology
Flow and timing	**Timing of CEA vs reference standard (days)** Detection of recurrent lesions within 6 months of the FDG-PET scan/CEA ± histology
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Seregni 1992

Study characteristics	
Patient sampling	**Country** Italy **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1975 - 1990 **Population (n)** 431 **Inclusion criteria** Curative resection **Exclusion criteria** N/R **Participants Included (n)** 336
Patient characteristics and setting	**Age range** 21 - 92 **Smoking status** N/R **Site of primary tumour** Colon 247, rectum 184 **Stage of primary tumour** Dukes A 40, B 186, C 107, D 72 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 136 **Site of recurrences** 50 local recurrences, 136 distant recurrences
Index tests	**CEA timing** N/R **CEA technique** N/R **CEA threshold** N/R **Definition of positive** Unclear **Which CEA value (s) used?** Unclear
Target condition and reference standard(s)	**Reference standard** N/R

Flow and timing	**Timing of CEA vs reference standard (days)** N/R		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Unclear		
Is the same method and instrument used for all CEA measurements?	Unclear		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		

Seregni 1992 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	Unclear		

Staib 2000

Study characteristics	
Patient sampling	**Country** Germany **Study design** Prospective **Setting** Hospital **Dates of data collection** 1994 - 1998 **Population (n)** 100 **Inclusion Criteria** Patients undergoing a whole-body PET scan for suspected relapse after curative resection of histologically confirmed colorectal cancer and who caused a "diagnostic problem". The "diagnostic problems" of the patients that led to a PET scan were (1) staging of rest of the body in patients with known recurrence (n = 30); (2) suspected recurrence (n = 32); (3) increasing CEA level (n = 13); (4) unclear finding on pelvic CT (n = 7); and (5) confirmation of liver metastases (n = 12) and lung metastases (n = 6)

	Exclusion Criteria No CEA evaluation, uncontrolled DM, or acute inflammation **Participants Included (n)** 98
Patient characteristics and setting	**Age range** 62 (32 - 80) **Smoking status** N/R **Site of primary tumour** Rectal 52, sigmoid 12, colon 22, lung or liver metastases 9, peritoneum 1 **Stage of primary tumour** I 8, II 25, III 46, IV 21 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** Chemo/immunotherapy 25 **Recurrences (n)** 58 **Site of recurrences** N/R
Index tests	**CEA timing** N/R **CEA technique** Liaison Kit (Byk-Sangtec, Diet- zenbach, Germany) **CEA threshold** 3 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** At point of recurrence
Target condition and reference standard(s)	**Follow-up schedule** Followed up with the department's established follow-up program. The indication for a whole body PET scan was given for patient s with suspected relapse after curative resection of colorectal cancer and who caused a "diagnostic problem" **Reference standard** FDG-PET/CT
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	
Methodological quality	

Staib 2000 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
			Low
DOMAIN 4: Flow and Timing			

Staib 2000 (*Continued*)

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	Yes		
Did all patients receive a reference standard?	Yes		

Steele 1982

Study characteristics	
Patient sampling	**Country** USA **Study design** RCT **Setting** Hospital **Dates of data collection** 1975 - 1980 **Population (n)** 770 **Inclusion criteria** B2 C colon or rectal cancer, 2 treatment arms: GITSG protocol 7175 was designed to evaluate adjuvant therapy (chemotherapy, radiotherapy, both, and none) following curative resection of Dukes' B2,C1,or C2 rectal carcinoma. Protocol 6175 was the study of the potential benefit of adjuvant therapy (chemotherapy, immunotherapy, both, and none) following clinically curative resection of Dukes' B2, C1, or C2 colon cancers **Exclusion criteria** CEA not recorded post-op **Participants Included (n)** 734
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour**

	Rectal 191, colon 543 **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** CEA < 5 **Chemotherapy/radiotherapy?** Yes, but not described **Recurrences (n)** 149 **Site of recurrences** Colon
Index tests	**CEA timing** On active treatment arms CEA values during and after treatment were to be obtained monthly during the first 3 months,every 3 months for the remainder of the first year, and every six months from then on. For control arms were to have CEA values obtained before operation, 1 week after operation, and at weeks 5, 10, 15, 25 after operation,and every 15 weeks thereafter **CEA technique** Hansen Z-gel technique. Interassay comparisons among the institutions and intra-assay analysis performed in the GITSG CEA reference laboratory at the Mallory Gastrointestinal Institute (Boston, Massachusetts) showed excellent reproducibility and acceptable variation among the various laboratories **CEA threshold** 2.5 µg/L **Definition of positive** Maximum level of CEA **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** Patients in both protocols were scheduled for regular clinic visits every 5 weeks during the first 6 months after surgery and every 15 weeks for the remainder of the first year. Physical examination, complete blood count, and liver function tests were performed at each visit. Liver/ spleen scan,chest posterio-anterior,and lateral roentgenograms were obtained every 6 months. Sigmoidoscopic examination and large-bowel, contrast roentgenograms were performed every year.Histologic evidence of tumor was the fundamental criterion for recurrence. However,roentgenographic evidence was acceptable in cases of lung or bony metastases. In the rectal-cancer adjuvant study, liver metastases were also accepted on the basis of liver scan, and local recurrence was accepted on the basis of perineal pain occurring acutely after a pain-free interval **Reference standard** Histology, XR for bony or lung mets, liver scan for liver mets in rectal study, or perineal pain occurring acutely after a pain-free interval
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Steele 1982 (*Continued*)

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	Yes		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Yes		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Low

Steele 1982 (*Continued*)

DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Tang 2009

Study characteristics	
Patient sampling	**Country** Taiwan **Study design** prospective **Setting** Hospital **Dates of data collection** 1995 - 2007 **Population (n)** N/A **Inclusion criteria** (1) Prior curative resection for histology-proven primary adenocarcinoma of the colorectum between 1995 and 2002, (2) availability of serial serum samples from before the operation and from after the surgery, and (3) follow-up with a definitive clinical outcome **Exclusion criteria** (1) synchronous or metachronous extracolonic cancers, (2) having neoadjuvant therapy for rectal cancer, and (3) fewer than 3 follow-up samples available for s-p53Ab analysis **Participants Included (n)** 305
Patient characteristics and setting	**Age range** 20 - 90 **Smoking status** N/R

	Site of primary tumour Colon 95, rectum 101, both 4 **Stage of primary tumour** I 45, II 130, III 130. **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 76 **Site of recurrences** locoregional 7, intra-abdominal or retro-peritoneal 18, hepatic 29, pulmonary 17, brain or bone 9
Index tests	**CEA timing** The CEA test was defined as positive if 2 consecutive postoperative CEA values were greater than 5 µg/L or the elevated preoperative CEA values had not returned to the normal level (5 µg/L) after surgery **CEA technique** Abbott Architect 2000 (Abbott Laboratories, Abbott Park, IL, USA) **CEA threshold** 5 µg/L **Definition of positive** The CEA test was defined as positive if 2 consecutive postoperative CEA values were greater than 5 µg/L or the elevated preoperative CEA values did not returned to the normal level (5 µg/L) after surgery **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** All cases were followed up at the outpatient department every 3 - 6 months until death or until December 2007. All the patients were followed according to the hospital guidelines of care. Briefly, all patients underwent a follow-up protocol of an outpatient visits every 3 - 6 months. The follow-up included physical examination and carcinoembryonic antigen tests as well as chest X-ray, abdominal sonography or abdominal computer-assisted tomography scan, and colonoscopy every 1 - 3 years after operation **Reference standard** Relapse confirmed by histology or by an imaging study
Flow and timing	**Timing of CEA vs reference standard (days)** Triggered by positive CEA
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	Unclear		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		

Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

Tate 1982

Study characteristics	
Patient sampling	**Country** UK **Study design** Prospective study after some retrospective sampling **Setting** Hospital **Dates of data collection** 1973 - 1978 **Population (n)** 520 **Inclusion criteria** curative resection **Exclusion criteria** Dukes D, no follow-up information available, signs of malignancy on first postoperative examination, malignancy of other sites during follow-up **Participants Included (n)** 468
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** N/R **Stage of primary tumour** A 94, B 226, C 128, unknown 20 **Perioperative investigations done to ensure no residual disease** First postoperative exam **Chemotherapy/radiotherapy?**

	Not stated **Recurrences (n)** 108 **Site of recurrences** N/R
Index tests	**CEA timing** At each follow-up visit **CEA technique** Assayed by a double-antibody radioimmunoassay system **CEA threshold** 40 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** all
Target condition and reference standard(s)	**Follow-up schedule** The follow-up procedure for each patient complied with the normal clinical practice for the hospital concerned and, in addition, at each follow up examination a specimen of plasma was taken for CEA determination. At least 6mly **Reference standard** Variable
Flow and timing	**Timing of CEA vs reference standard (days)** Very variable
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		

Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
			Unclear
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Unclear		

Tobaruela 1997

Study characteristics	
Patient sampling	**Country** Spain **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1988 - 1993 **Population (n)** N/R **Inclusion Criteria** Colorectal cancer, curative surgery for Dukes C disease. **Exclusion Criteria** Dukes A, B, D **Participants Included (n)** 60
Patient characteristics and setting	**Age range** < 5 preop 60.9 (34 - 85) + > 5 preop 64.9 (47 - 83) **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes C = 60 **Perioperative investigations done to ensure no residual disease** No **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 21 **Site of recurrences** Hepatic 9, locoregional 6, combined 3, pulmonary 3
Index tests	**CEA timing** As follow-up schedule **CEA technique** Enzyme-linked immunoassay **CEA threshold** 5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** Physical examination and CEA 3 monthly for 2 years, then 6 monthly up to 5 years. USS abdomen twice a year. CT if CEA increased

	Reference standard CT if CEA increased		
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			

Tobaruela 1997 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

Triboulet 1983

Study characteristics	
Patient sampling	**Country** France **Study design** Prospective cohort study **Setting** Hospital **Dates of data collection** 1976 - 1979 **Population (n)** 91 **Inclusion criteria** Operated on with curative intent for colorectal cancer **Exclusion criteria**

	Conditions which could affect B2 microglobulin level: altered renal function (creatinine > 88.4 umol/l); liver disease: chronic active cirrhosis, primary biliary cirrhosis, acute hepatitis. Metastasis or Dukes D cancers. Patients whose CEA had not returned to normal within 3 months of the operation
Patient characteristics and setting	**Participants included (n)** 91 **Age range** 33 - 80 **Smoking status** N/R **Site of primary tumour** Colon 65, rectum 26 **Stage of primary tumour** Dukes A&B = 50; C = 41 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** No **Recurrences (n)** 43 **Site of recurrences** 12 rectum, 31 colon
Index tests	**CEA timing** Every 3 months **CEA technique** Radioimmunoassay (sorin) **CEA threshold** 20 μg/L **Definition of positive** N/R **Which CEA value (s) used?** N/R
Target condition and reference standard(s)	**Follow-up schedule** CEA & B2m every 3 months for at least 2 years. Clinical and laboratory monitoring was ensured by the same physician during the first two years post-op in a pre-established protocol with a barium enema and / or an endoscopy during the first two years enema. CXR and Liver USS annually. Further investigations if indicated (CT chest, bone scan)
Flow and timing	**Timing of CEA vs reference standard (days)** Yearly CXR and liver USS; enema and/or endoscopy done at least once in the 2 year follow-up
Comparative	
Notes	
Methodological quality	

Triboulet 1983 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	No		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			

Triboulet 1983 (*Continued*)

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	Unclear		
Was the the timing between index test(s) and reference standard ascertainable?	Unclear		
Did all patients receive a reference standard?	No		

Wang 1994

Study characteristics	
Patient sampling	**Country** **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1981 - 1986 **Population (n)** 352 **Inclusion criteria** Operated for histologically proven colorectal cancer **Exclusion criteria** No preoperative CEA or lost to follow-up, Dukes A, or Dukes D **Participants Included (n)** 272
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** Dukes B 160, C 112 **Perioperative investigations done to ensure no residual disease** N/R

	Chemotherapy/radiotherapy? N/R **Recurrences (n)** 27 **Site of recurrences** N/R
Index tests	**CEA timing** Blood samples for CEA measurement were taken a few days before operation and about 1 month after operation and afterward at intervals of 3 - 4 months, combined with physical examination **CEA technique** Radioimmunoassay kit manufactured by Abbott Laboratory (Chicago, IL, USA) **CEA threshold** 5 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** Blood samples for CEA measurement were taken a. few days before operation and about one month after operation and afterward at intervals of three to four months, combined with physical examination. Other procedures such as colonoscopy, liver sonography, and chest x-ray were performed annually, **Reference standard** In the cases where we suspected recurrence the patient underwent additional abdominal computed tomography, bone scanning, or other diagnostic procedures
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	No		
			Low

Wang 1994 (*Continued*)

DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		

Wang 1994 (*Continued*)

Did all patients receive a reference standard?	Yes		

Wood 1980

Study characteristics	
Patient sampling	**Country** UK **Study design** Retrospective **Setting** Hospital **Dates of data collection** 1974 - 1976 **Population (n)** 148 **Inclusion criteria** Apparently curative surgery for adenocarcinoma of the colon and rectum without evidence of metastatic disease **Exclusion criteria** N/R **Participants Included (n)** 148
Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal **Stage of primary tumour** N/R **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 36 **Site of recurrences** Local 17, local + liver 2, local + bone 2, local + metachronous primary 1, liver 8, bone 5, lung 2
Index tests	**CEA timing** Each follow-up visit, 2 consecutive raised CEA triggered investigation for recurrence **CEA technique** CEA levels were assayed by a double antibody radioimmunoassay on unextracted serum

	CEA threshold 25 µg/L **Definition of positive** 2 consecutively elevated values **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** CEA at 3 - 6 months intervals post-operative for up to 56 months or until death **Reference standard** If CEA positive then CXR, Liver scan, and bone scan. If these are negative, additional BE and/or colonoscopy
Flow and timing	**Timing of CEA vs reference standard (days)** per protocol
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		

Wood 1980 (*Continued*)

Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

Yakabe 2010

Study characteristics	
Patient sampling	**Country** Japan **Study design** Prospective

	Setting Hospital **Dates of data collection** 1999 - 2003 **Population (n)** 266 **Inclusion criteria** Curative resection for colorectal cancer, TNM stages I - III, postoperative examinations according to the follow-up schedule **Exclusion criteria** Inappropriate follow-up **Participants Included (n)** 227
Patient characteristics and setting	**Age range** 65.2 (± 10.8) years **Smoking status** N/R **Site of primary tumour** Colon 138, rectum 89 **Stage of primary tumour** I 34, II 94, III 99 **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 62 **Site of recurrences** N/R
Index tests	**CEA timing** 3 months for the first 3 years and every 6 months during years 4 and 5 **CEA technique** Latex immunoassay, Mitsubishi Chemical Ltd, Japan **CEA threshold** 4.5 µg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** History was taken and a physical examination and measurement of tumor markers were performed every 3 months for the first 3 years and every 6 months during years 4 and 5. Chest X- ray and abdominal computed tomography (CT) were done every 6 months for 5 years, and colonoscopy was performed at 1 and 3 years after surgery. Patients were observed until 5 years after surgery or until recurrence was confirmed **Reference standard** Recurrence was confirmed histologically or radiologically

Flow and timing	**Timing of CEA vs reference standard (days)** per protocol		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	Yes		

Yu 1992

Study characteristics	
Patient sampling	**Country** China **Study design** Retrospective observational study **Setting** Teaching hospital in Shanghai **Dates of data collection** May 1988 - March 1990 **Population (n)** 216 **Inclusion criteria** Primary colorectal cancer having curative surgery in the teaching hospital or other hospitals **Exclusion Criteria** N/R **Participants Included (n)** 182

Yu 1992 (*Continued*)

Patient characteristics and setting	**Age range** N/R **Smoking status** N/R **Site of primary tumour** Colorectal cancer 121, colon cancer 95 **Stage of primary tumour** Only reported Dukes stage data for the 28 before- surgery cases (Table 1) **Perioperative investigations done to ensure no residual disease** N/R **Chemotherapy/radiotherapy?** N/R **Recurrences (n)** 66 **Site of recurrences** N/R
Index tests	**CEA timing** N/R **CEA technique** RIA **CEA threshold** 15 μg/L **Definition of positive** 1 elevated value **Which CEA value (s) used?** All
Target condition and reference standard(s)	**Follow-up schedule** CEA first measured at 6 weeks after curative surgery; then every 3 months, plus liver ultrasound test and basic health check **Reference standard** Positive CEA and CA-19-9 triggers ultrasound and CT or colonoscopy
Flow and timing	**Timing of CEA vs reference standard (days)** N/R
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Did the study avoid inappropriate exclusions?	Unclear		
			Low
DOMAIN 2: Index Test All CEA thresholds			
If a threshold was used, was it pre-specified?	Yes		
Is the same method and instrument used for all CEA measurements?	Yes		
Is there an estimation of reproducibility of the method, for example the % coefficient of variation at specific concentrations?	No		
Is there an indication of method accuracy, for example, is there evidence of participation in an external quality assessment and proficiency testing scheme?	No		
			Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
			Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Yu 1992 (*Continued*)

Was the index test repeated prior to the reference standard?	No		
Was the the timing between index test(s) and reference standard ascertainable?	No		
Did all patients receive a reference standard?	No		

ACBE= air contrast barium enema
ALP: alkaline phosphatase
APCT: abdominopelvic computed tomography
BE: barium enema
CT: computed tomography
CXR: chest xray
DCBE: double contrast barium enema
DM: diabetes mellitus
ESR: erythrocyte sedimentation rate
FOBT: faecal occult blood test
LDH: lactate dehydrogenase
LFT: latex fixation test
MRI: magnetic resonance imaging
N/R: not reported
RIA: radioimmunoassay
SCC: squamous cell carcinoma.
SGOT: serum glutamic oxaloacetic transaminase
SGPT: serum glutamate pyruvate transaminase
TNM: primary tumour, regional nodes, metastasis
TPA: tissue plasminogen activator
µg/L = micrograms per litre
USS = ultrasound scan

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Afsaneh 2012	2 x 2 data not ascertainable
Ahmed 2013	Only CEA positive
Aitkin 2012	Only CEA positive

(*Continued*)

Amin 2012	2 x 2 data not ascertainable
Arnaud 1979	2 x 2 data not ascertainable
Arnaud 1997	2 x 2 data not ascertainable
Arriola 2006	2 x 2 data not ascertainable
Auer 1977	Stomach and colorectal cancer combined
Bakalakos 1999	Liver metastases only
Barrillari 1996	Only cases of recurrence
Beatty 1979	Only cases of recurrence
Beets 1994	Only cases of recurrence
Bhatavedekar 1992	Alternative analysis - median CEA
Bivins 1974	n < 30
Boey 1984	Alternative analysis - slope
Borie 2004	Only cases of recurrence
Brummendorf 1985	2 x 2 data not ascertainable
Brummendorf 1986	Alternative analysis - doubling time
Bucci 1994	2 x 2 data not ascertainable
Camunas 1991	Only cases of recurrence
Cangemi 1984	n < 30
Cangemi 1987	Case-control study
Carl 1983	Alternative analysis - slope
Carpelan-Holmström 1996	Only cases of recurrence
Castells 1998	2 x 2 data not ascertainable
Catania 1981	2 x 2 data not ascertainable
Chang 2012	Only cases of recurrence

(*Continued*)

Chapman 1998	Pre-operative CEA
Chen 2010	Only CEA positive cases
Cho 2007	Pre-operative CEA
Choi 1997	Only CEA positive cases
Colombo 1986	2 x 2 data not ascertainable
Cossu 1984	Alternative analysis
Dalton 2010	2 x 2 data not ascertainable
Dash 2012	Only CEA negative cases
De Brauw 1987	2 x 2 data not ascertainable
De Levin 1982	n<30
De Salvo 1997	Only cases of recurrence
Dhar 1972	2 x 2 data not ascertainable
Di Cristofaro 2012	Alter n a tive analysis - economic
Engarås 2001	Only cases of recurrence
Farquharson 2012	Only CEA positive cases
Fernandes 2006	2 x 2 data not ascertainable
Filella 1994	2 x 2 data not ascertainable
Filiz 2009	Not follow-up for recurrence - prognostic value of postoperative CEA
Finlay 1983	Not curative resection
Fiocchi 2011	Not follow-up for recurrence - includes patients with suspicion of recurrence on CT
Florio 1988	2 x 2 data not ascertainable
Fora 2012	2 x 2 data not ascertainable
Forones 1997	Preoperative CEA
Forones 1998	n < 30

(*Continued*)

Fortner 1988	Only cases of recurrence
Fournier 1999	2 x 2 data not ascertainable
Fucini 1983	Duplicated dataset
Fucini 1984	n < 30
Fucini 1985	2 x 2 data not ascertainable
Gail 1981	Alteranative analysis - modelling
Gajdukevich 2010	Not curative surgery
Gaudagni 1999	2 x 2 data not ascertainable
Graham 1998	Only cases of recurrence
Gray 1981	Only cases of recurrence
Griesenberg 1999	Only cases of recurrence
Grossetti 1981	2 x 2 data not ascertainable
Grossmann 2007	2 x 2 data not ascertainable
Haga 1990	Only cases of recurrence
Hall 1994	2 x 2 data not ascertainable
Hara 2011	Duplicate dataset
Herrera 1976	Case-control study
Hida 1996	2 x 2 data not ascertainable
Hohenberger 1994	Only cases of recurrence
Holt 2010	Only cases of recurrence
Holubec 2000	2 x 2 data not ascertainable
Holyoke 2975	n < 30
Houlbec 2001	2 x 2 data not ascertainable
Humphreys 2011	Only CEA negative cases

(*Continued*)

Huyghe 1983	2 x 2 data not ascertainable
Iarumov 1998	Unable to locate full text
Indinnimeo 1999	Unable to locate full text
Ito 2002	Alternative analysis - doubling time
Jaeger 1975	Only cases of recurrence
Jiang 1989	2 x 2 data not ascertainable
Kanellos 2006b	Not follow-up - portal CEA sampling
Karesen 1980	2 x 2 data not ascertainable
Kawamura 2010	Only cases of recurrence
Kerr 2012	2 x 2 data not ascertainable
Khan 2009	2 x 2 data not ascertainable
Kimura 1986	Only cases of recurrence
Kishimoto 2010	2 x 2 data not ascertainable
Koch 1977	2 x 2 data not ascertainable
Koch 1979	Not follow-up for recurrence - prognostic value of postoperative CEA
Koch 1982	Not follow-up for recurrence - prognostic value of postoperative CEA
Koga 1999	Alternative analysis - doubling time
Korner 2005	2 x 2 data not ascertainable
Kumar 2011	Only cases of recurrence
Lagache 1980	Only cases of recurrence
Lauterbach 1987	2 x 2 data not ascertainable
Lavin 1981	Case-control study
Lechner 2000	2 x 2 data not ascertainable
Leventakos 2013	Only cases of recurrence

(*Continued*)

Levy 2012	Duplicate dataset
Lipska 2007	Only cases of recurrence
Lipska 2010	Only cases of recurrence
Lorenz 1986	Not follow-up for recurrence - prognostic value of postoperative CEA
Lunde 1982	Only cases of recurrence
Ma 2006	Not follow-up for recurrence - prognostic value of postoperative CEA
Mach 1974	Case-control study
Makela 1995	2 x 2 data not ascertainable
Makis 2013	2 x 2 data not ascertainable
Mant 2013	Duplicate dataset
Martin 1976	Only CEA positive case
Martin 1979	Only CEA positive case
Martin 1980	Only CEA positive case
Marucci 1983	Not follow-up for recurrence - prognostic value of postoperative CEA
May 2012	2 x 2 data not ascertainable
Mazilu 2012	Unable to locate full text
McCarthy 1985	2 x 2 data not ascertainable
Meling 1992	2 x 2 data not ascertainable
Mentges 1986	2 x 2 data not ascertainable
Mentges 1988	Only cases of recurrence
Metzger 1983	Only cases of recurrence
Metzger 1985	Only cases of recurrence
Minton 1978a	Alternative analysis - nomogram
Minton 1978b	Only cases of recurrence

(*Continued*)

Minton 1989	Alternative analysis - nomogram
Miwa 1980	Only cases of recurrence
Moertel 1978	Only cases of recurrence
Morelli 1985	n<30
Moreno Carretero 1998	2 x 2 data not ascertainable
Moschl 1980	2 x 2 data not ascertainable
Nicolini 1995	2 x 2 data not ascertainable
Nicolini 2005	2 x 2 data not ascertainable
Nicolini 2010	Only cases of recurrence
Northover 1985	2 x 2 data not ascertainable
Northover 1986	Review article
Northover 2003	Review article
Novis 1986	Only cases of recurrence
Nowacki 1983	2 x 2 data not ascertainable
Ntinas 2004	2 x 2 data not ascertainable
O'Dwyer 1987	2 x 2 data not ascertainable
O'Dwyer 1988	Only CEA positive cases
Obradovic 2011	2 x 2 data not ascertainable
Odariuk 1989	Only CEA positive cases
Ovaska 1989	Only cases of recurrence
Ozhiganov 1986	Unable to translate
Ozkan 2012a	2 x 2 data not ascertainable
Ozkan 2012b	2 x 2 data not ascertainable
Park 2012	Only cases of recurrence
Park 2013	2 x 2 data not ascertainable

(*Continued*)

Pecorella 1996	2 x 2 data not ascertainable
Peethambaram 1997	2 x 2 data not ascertainable
Pereira 2004	Unable to locate
Persijin 1981	2 x 2 data not ascertainable
Pfeiffer 1979	2 x 2 data not ascertainable
Philips 1984	2 x 2 data not ascertainable
Pietra 1998	2 x 2 data not ascertainable
Plebani 1996	2 x 2 data not ascertainable
Pompecki 1980	n < 30
Pribelsky 2002	Only cases of recurrence
Primrose 2011	Duplicate dataset
Primrose 2014	2 x 2 data not ascertainable
Quentmeier 1990	Only cases of recurrence
Reddy 2013	Only cases of recurrence
Revetria 1989	Case-control stu dy
Rezamansourian 2011	Review article
Rieger 1975	Only cases of recurrence
Rockall 1999	2 x 2 data not ascertainable
Rocklin 1990	Only cases of recurrence
Rocklin 1991	2 x 2 data not ascertainable
Rodriguez-Moranta 2006a	Only cases of recurrence
Rognum 1986	Only cases of recurrence
Sagar 1989	2 x 2 data not ascertainable
Sandelewski 2005	Only cases of recurrence

(*Continued*)

Sanli 2012	Only CEA positive cases
Sardi 1989	Only cases of recurrence
Sarikaya 2007	Only CEA negative cases
Secco 1989	2 x 2 data not ascertainable
Secco 2000	Only cases of recurrence
Segol 1977	Not follow-up for recurrence - prognostic value of postoperative CEA
Shirley 2012	2 x 2 data not ascertainable
Simo 2002	Only CEA positive cases
Sirisriro 1996	Only CEA positive cases
Song 2010	Alternative analysis - CEA trend
Sorensen 2010	Only CEA positive cases
Staab 1985a	Alternative analysis - slope
Staab 1985b	Alternative analysis - slope
Stautner-Brückmann 1990	Only cases of recurrence
Steele 1980	Only CEA positive cases
Stuckle 2000	2 x 2 data not ascertainable
Su 2012	Only cases of recurrence
Sugarbaker 1976	Only CEA positive cases
Szymendera 1982 a	Only cases of recurrence
Szymendera 1982 b	2 x 2 data not ascertainable
Szymendera 1985	2 x 2 data not ascertainable
Takashima 1982	Only cases of recurrence
Tomoda 1981	Non-curative surgery
Tsai 2009	Only cases of recurrence

(*Continued*)

Tsikitis 2009	Only cases of recurrence
Verberne 2013 a	Liver metastases only
Verberne 2013 b	Liver metastases only
Wan 1994	2 x 2 data not ascertainable
Wanebo 1978a	Only cases of recurrence
Wanebo 1978b	Only cases of recurrence
Wang 2007	Not follow-up for recurrence - prognostic value of postoperative CEA
Wang 2010	2 x 2 data not ascertainable
Wedell 1981	Only cases of recurrence
Weiss 1998	2 x 2 data not ascertainable
Wichmann 2000a	Only cases of recurrence
Wichmann 2000b	Preoperative CEA
Wichmann 2002	Preoperative CEA
Wolf 1997	Only cases of recurrence
Wood 1975	Unable to locate
Yu 2013	Only cases of recurrence
Zeng 1993	Only cases of recurrence
Zervos 2001	2 x 2 data not ascertainable
Ziegenbein 1980	Alternative analysis - trend
Zuniga 1989	2 x 2 data not ascertainable

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

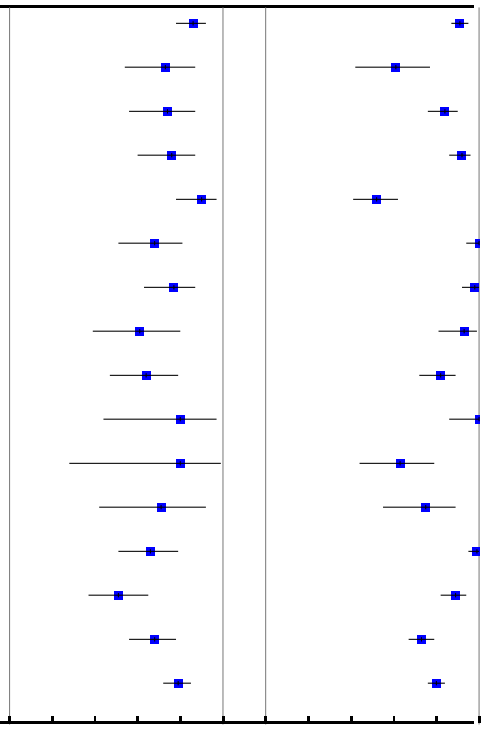
Test	No. of studies	No. of participants
1 CEA - all thresholds	52	9717
2 CEA at 2.5µg/L	7	1515
3 CEA at 5µg/L	23	4585
4 CEA at 10µg/L	7	1607

Test 1. CEA - all thresholds.

Review: Blood CEA levels for detecting recurrent colorectal cancer

Test: 1 CEA - all thresholds

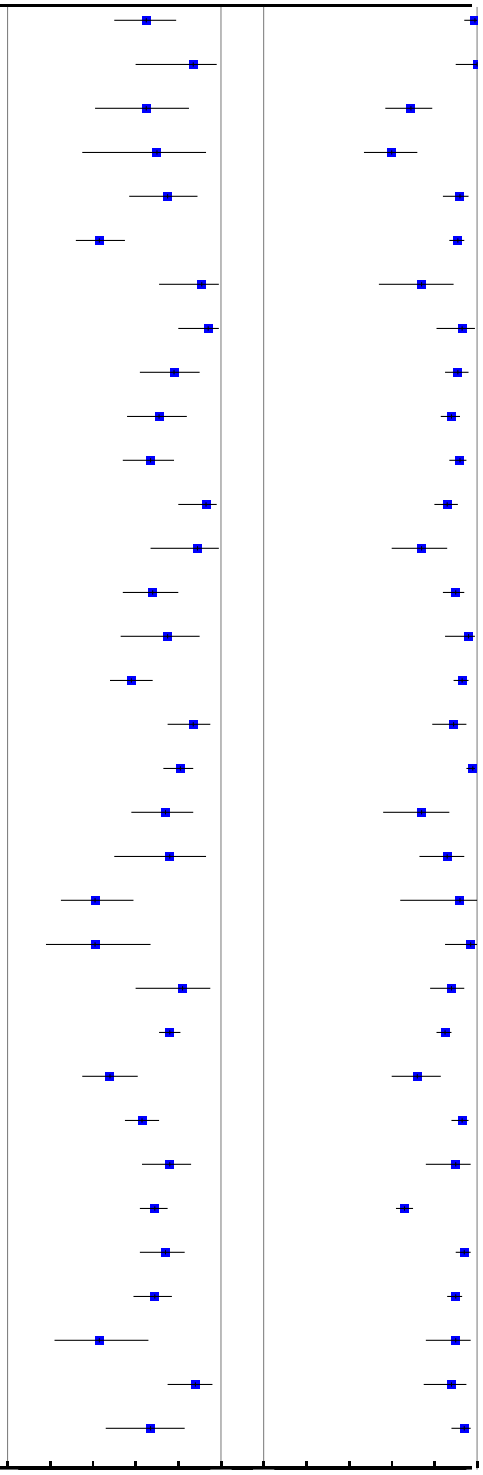
Study	TP	FP	FN	TN	Sensitivity	Specificity
Banaszkiewicz 2011	96	20	16	208	0.86 [0.78, 0.92]	0.91 [0.87, 0.95]
Barillari 1992	24	13	9	20	0.73 [0.54, 0.87]	0.61 [0.42, 0.77]
Beart 1981	25	18	9	97	0.74 [0.56, 0.87]	0.84 [0.76, 0.90]
Bjerkeset 1988	34	9	11	108	0.76 [0.60, 0.87]	0.92 [0.86, 0.96]
Carlsson 1983	45	43	5	46	0.90 [0.78, 0.97]	0.52 [0.41, 0.62]
Carpelan-Holmström 2004	27	0	13	62	0.68 [0.51, 0.81]	1.00 [0.94, 1.00]
Carriquiry 1999	40	2	12	88	0.77 [0.63, 0.87]	0.98 [0.92, 1.00]
Deveney 1984	14	3	9	39	0.61 [0.39, 0.80]	0.93 [0.81, 0.99]
Engar s 2003	25	17	14	76	0.64 [0.47, 0.79]	0.82 [0.72, 0.89]
Farinon 1980	8	0	2	25	0.80 [0.44, 0.97]	1.00 [0.86, 1.00]
Fezoulidis 1987	4	12	1	20	0.80 [0.28, 0.99]	0.63 [0.44, 0.79]
Fucini 1987	10	7	4	21	0.71 [0.42, 0.92]	0.75 [0.55, 0.89]
Graffner 1985	31	1	16	119	0.66 [0.51, 0.79]	0.99 [0.95, 1.00]
Hara 2008	26	14	25	109	0.51 [0.37, 0.65]	0.89 [0.82, 0.94]
Hara 2010	50	61	24	168	0.68 [0.56, 0.78]	0.73 [0.67, 0.79]
Hine 1984	135	89	36	366	0.79 [0.72, 0.85]	0.80 [0.76, 0.84]



(*Continued* . . .)

(. . . *Continued*)

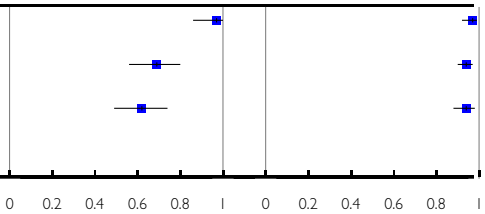
Study	TP	FP	FN	TN	Sensitivity	Specificity
Irvine 2007	30	1	16	92	0.65 [0.50, 0.79]	0.99 [0.94, 1.00]
Johnson 1985	13	0	2	36	0.87 [0.60, 0.98]	1.00 [0.90, 1.00]
Jubert 1978	13	24	7	53	0.65 [0.41, 0.85]	0.69 [0.57, 0.79]
Kanellos 2006a	7	25	3	38	0.70 [0.35, 0.93]	0.60 [0.47, 0.72]
Kato 1980	24	8	8	89	0.75 [0.57, 0.89]	0.92 [0.84, 0.96]
Kim 2013	34	23	45	234	0.43 [0.32, 0.55]	0.91 [0.87, 0.94]
Kohler 1980	20	7	2	20	0.91 [0.71, 0.99]	0.74 [0.54, 0.89]
Koizumi 1992	32	3	2	40	0.94 [0.80, 0.99]	0.93 [0.81, 0.99]
Korner 2007	29	10	8	106	0.78 [0.62, 0.90]	0.91 [0.85, 0.96]
Li Destri 1998	32	23	13	171	0.71 [0.56, 0.84]	0.88 [0.83, 0.92]
Lucha 1997	44	18	22	201	0.67 [0.54, 0.78]	0.92 [0.87, 0.95]
Luporini 1979	37	22	3	136	0.93 [0.80, 0.98]	0.86 [0.80, 0.91]
Mach 1978	17	12	2	35	0.89 [0.67, 0.99]	0.74 [0.60, 0.86]
Mackay 1974	36	17	17	150	0.68 [0.54, 0.80]	0.90 [0.84, 0.94]
Mariani 1980	18	2	6	43	0.75 [0.53, 0.90]	0.96 [0.85, 0.99]
McCall 1994	57	15	41	198	0.58 [0.48, 0.68]	0.93 [0.89, 0.96]
Miles 1995	46	8	7	64	0.87 [0.75, 0.95]	0.89 [0.79, 0.95]
Minton 1985	105	6	25	264	0.81 [0.73, 0.87]	0.98 [0.95, 0.99]
Mittal 2011	29	9	10	25	0.74 [0.58, 0.87]	0.74 [0.56, 0.87]
Nishida 1988	13	7	4	42	0.76 [0.50, 0.93]	0.86 [0.73, 0.94]
Ochoa-Figueroa 2012	14	1	20	12	0.41 [0.25, 0.59]	0.92 [0.64, 1.00]
Ohlsson 1995	7	1	10	35	0.41 [0.18, 0.67]	0.97 [0.85, 1.00]
Ohtsuka 2008	18	9	4	66	0.82 [0.60, 0.95]	0.88 [0.78, 0.94]
Park 2009	222	69	69	382	0.76 [0.71, 0.81]	0.85 [0.81, 0.88]
Peng 2013	30	18	32	47	0.48 [0.35, 0.61]	0.72 [0.60, 0.83]
Seregni 1992	86	15	50	185	0.63 [0.55, 0.71]	0.93 [0.88, 0.96]
Staib 2000	44	4	14	36	0.76 [0.63, 0.86]	0.90 [0.76, 0.97]
Steele 1982	150	175	68	341	0.69 [0.62, 0.75]	0.66 [0.62, 0.70]
Tang 2009	56	14	20	215	0.74 [0.62, 0.83]	0.94 [0.90, 0.97]
Tate 1982	70	38	32	328	0.69 [0.59, 0.77]	0.90 [0.86, 0.93]
Tobaruela 1997	9	4	12	35	0.43 [0.22, 0.66]	0.90 [0.76, 0.97]
Triboulet 1983	38	6	5	42	0.88 [0.75, 0.96]	0.88 [0.75, 0.95]
Wang 1994	18	9	9	131	0.67 [0.46, 0.83]	0.94 [0.88, 0.97]



(*Continued* . . .)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Wood 1980	36	3	1	108	0.97 [0.86, 1.00]	0.97 [0.92, 0.99]
Yakabe 2010	43	9	19	154	0.69 [0.56, 0.80]	0.94 [0.90, 0.97]
Yu 1992	41	7	25	109	0.62 [0.49, 0.74]	0.94 [0.88, 0.98]

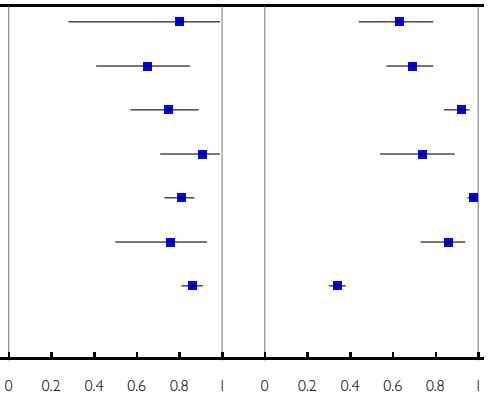


Test 2. CEA at 2.5µg/L.

Review: Blood CEA levels for detecting recurrent colorectal cancer

Test: 2 CEA at 2.5 g/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fezoulidis 1987	4	12	1	20	0.80 [0.28, 0.99]	0.63 [0.44, 0.79]
Jubert 1978	13	24	7	53	0.65 [0.41, 0.85]	0.69 [0.57, 0.79]
Kato 1980	24	8	8	89	0.75 [0.57, 0.89]	0.92 [0.84, 0.96]
Kohler 1980	20	7	2	20	0.91 [0.71, 0.99]	0.74 [0.54, 0.89]
Minton 1985	105	6	25	264	0.81 [0.73, 0.87]	0.98 [0.95, 0.99]
Nishida 1988	13	7	4	42	0.76 [0.50, 0.93]	0.86 [0.73, 0.94]
Steele 1982	188	343	30	176	0.86 [0.81, 0.91]	0.34 [0.30, 0.38]

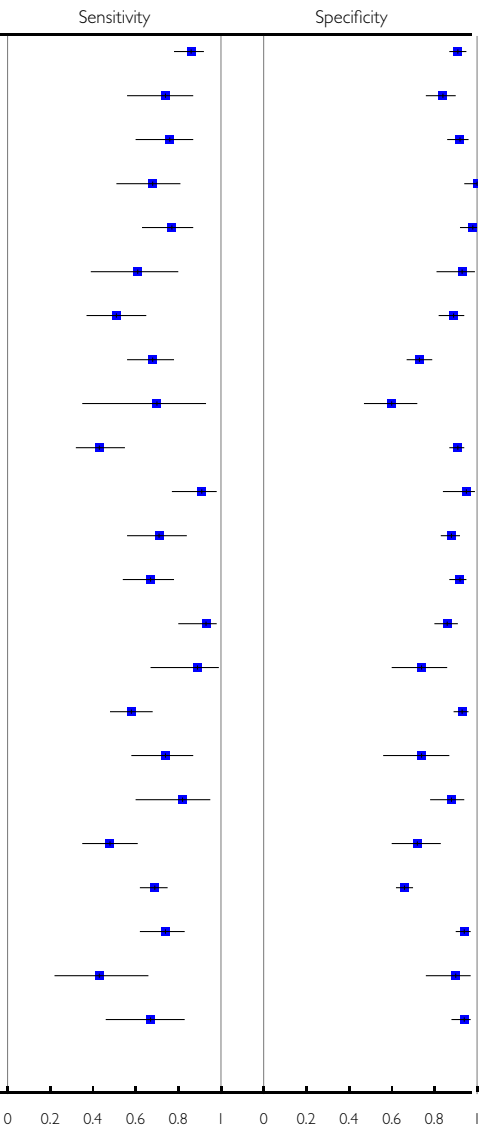


Test 3. CEA at 5µg/L.

Review: Blood CEA levels for detecting recurrent colorectal cancer

Test: 3 CEA at 5 g/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Banaszkiewicz 2011	96	20	16	208	0.86 [0.78, 0.92]	0.91 [0.87, 0.95]
Beart 1981	25	18	9	97	0.74 [0.56, 0.87]	0.84 [0.76, 0.90]
Bjerkeset 1988	34	9	11	108	0.76 [0.60, 0.87]	0.92 [0.86, 0.96]
Carpelan-Holmström 2004	27	0	13	62	0.68 [0.51, 0.81]	1.00 [0.94, 1.00]
Carriquiry 1999	40	2	12	88	0.77 [0.63, 0.87]	0.98 [0.92, 1.00]
Deveney 1984	14	3	9	39	0.61 [0.39, 0.80]	0.93 [0.81, 0.99]
Hara 2008	26	14	25	109	0.51 [0.37, 0.65]	0.89 [0.82, 0.94]
Hara 2010	50	61	24	168	0.68 [0.56, 0.78]	0.73 [0.67, 0.79]
Kanellos 2006a	7	25	3	38	0.70 [0.35, 0.93]	0.60 [0.47, 0.72]
Kim 2013	34	23	45	234	0.43 [0.32, 0.55]	0.91 [0.87, 0.94]
Koizumi 1992	32	2	3	40	0.91 [0.77, 0.98]	0.95 [0.84, 0.99]
Li Destri 1998	32	23	13	171	0.71 [0.56, 0.84]	0.88 [0.83, 0.92]
Lucha 1997	44	18	22	201	0.67 [0.54, 0.78]	0.92 [0.87, 0.95]
Luporini 1979	37	22	3	136	0.93 [0.80, 0.98]	0.86 [0.80, 0.91]
Mach 1978	17	12	2	35	0.89 [0.67, 0.99]	0.74 [0.60, 0.86]
McCall 1994	57	15	41	198	0.58 [0.48, 0.68]	0.93 [0.89, 0.96]
Mittal 2011	29	9	10	25	0.74 [0.58, 0.87]	0.74 [0.56, 0.87]
Ohtsuka 2008	18	9	4	66	0.82 [0.60, 0.95]	0.88 [0.78, 0.94]
Peng 2013	30	18	32	47	0.48 [0.35, 0.61]	0.72 [0.60, 0.83]
Steele 1982	150	175	68	341	0.69 [0.62, 0.75]	0.66 [0.62, 0.70]
Tang 2009	56	14	20	215	0.74 [0.62, 0.83]	0.94 [0.90, 0.97]
Tobaruela 1997	9	4	12	35	0.43 [0.22, 0.66]	0.90 [0.76, 0.97]
Wang 1994	18	9	9	131	0.67 [0.46, 0.83]	0.94 [0.88, 0.97]

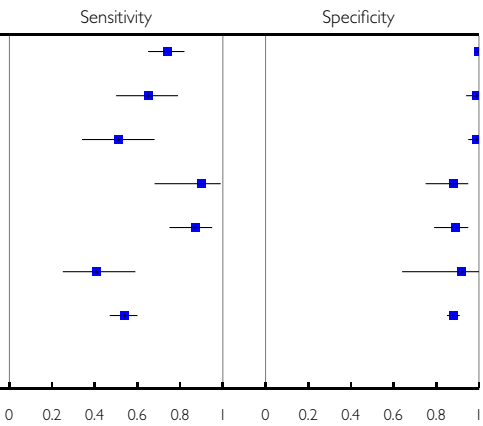


Test 4. CEA at 10µg/L.

Review: Blood CEA levels for detecting recurrent colorectal cancer

Test: 4 CEA at 10 g/L

Study	TP	FP	FN	TN	Sensitivity	Specificity
Banaszkiewicz 2011	83	0	29	228	0.74 [0.65, 0.82]	1.00 [0.98, 1.00]
Irvine 2007	30	1	16	92	0.65 [0.50, 0.79]	0.99 [0.94, 1.00]
Korner 2007	19	1	18	115	0.51 [0.34, 0.68]	0.99 [0.95, 1.00]
Mariani 1980	18	6	2	43	0.90 [0.68, 0.99]	0.88 [0.75, 0.95]
Miles 1995	46	8	7	64	0.87 [0.75, 0.95]	0.89 [0.79, 0.95]
Ochoa-Figueroa 2012	14	1	20	12	0.41 [0.25, 0.59]	0.92 [0.64, 1.00]
Steele 1982	117	62	101	454	0.54 [0.47, 0.60]	0.88 [0.85, 0.91]



APPENDICES

Appendix 1. Cochrane Central Register of Controlled Trials search strategy

#1	MeSH descriptor: [Colorectal Neoplasms] explode all trees
#2	(colorectal near/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*)):ti,ab,kw (Word variations have been searched)
#3	(colon* near/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*)):ti,ab,kw (Word variations have been searched)
#4	(bowel near/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*)):ti,ab,kw (Word variations have been searched)
#5	(rectal near/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*)):ti,ab,kw (Word variations have been searched)
#6	(rectum near/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*)):ti,ab,kw (Word variations have been searched)
#7	#1 or #2 or #3 or #4 or #5 or #6
#8	MeSH descriptor: [Carcinoembryonic Antigen] explode all trees
#9	cea:ti,ab,kw (Word variations have been searched)

(*Continued*)

#10	(carcinoembryonic near/3 antigen*):ti,ab,kw (Word variations have been searched)
#11	(carcinoembryonic near/3 antibod*):ti,ab,kw (Word variations have been searched)
#12	(carcino-embryonic near/3 antigen*):ti,ab,kw (Word variations have been searched)
#13	(carcino-embryonic near/3 antibod*):ti,ab,kw (Word variations have been searched)
#14	#8 or #9 or #10 or #11 or #12 or #13
#15	#7 and #14

Appendix 2. MEDLINE search strategy

1	colorectal neoplasms/ or exp adenomatous polyposis coli/ or exp colonic neoplasms/ or colorectal neoplasms, hereditary nonpolyposis/ or exp rectal neoplasms/	142383
2	(colorectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	69267
3	(colon* adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	56720
4	(bowel adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	3988
5	(rectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	18409
6	(rectum adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	4598
7	1 or 2 or 3 or 4 or 5 or 6	179150
8	Carcinoembryonic Antigen/	13372
9	cea.ti,ab.	16371
10	(carcinoembryonic adj3 antigen?).ti,ab.	11442
11	(carcinoembryonic adj3 antibod*).ti,ab.	622
12	(carcino-embryonic adj3 antigen?).ti,ab.	431

(*Continued*)

13	(carcino-embryonic adj3 antibod*).ti,ab.	13
14	8 or 9 or 10 or 11 or 12 or 13	23958
15	Neoplasm Recurrence, Local/	79823
16	Recurrence/	155149
17	recur*.ti,ab.	381384
18	relaps*.ti,ab.	116217
19	treatment failure/	25585
20	Reoperation/	63998
21	Follow-Up Studies/ and Postoperative Care/	5767
22	reoperat*.ti,ab.	23840
23	((local or distant) adj2 failure).ti,ab.	3371
24	((therap* or treatment or surg*) adj3 fail*).ti,ab.	58705
25	((therap* or treatment or surg*) adj3 (respond* or response*)) .ti,ab	116904
26	((postoperat* or post-operat* or postsurg* or post-surg* or post-treat* or post-treat* or posttherap* or post-therap*) adj5 follow up).ti,ab	16723
27	((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) adj5 surveillance).ti,ab	1277
28	((postoperat* or post-operat* or postsurg* or post-surg* or post-treat* or post-treat* or posttherap* or post-therap*) adj5 monitor*).ti,ab	3604
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	802827
30	7 and 14 and 29	1993
31	7 and 14	6353
32	limit 31 to "reviews (maximizes specificity)"	41

(*Continued*)

33	30 not 32	1966

Appendix 3. Embase search strategy

1	exp colon cancer/ or exp rectum cancer/	172220
2	(colorectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	97898
3	(colon* adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	75721
4	(bowel adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	5761
5	(rectal adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	26610
6	(rectum adj3 (neoplas* or cancer? or tumour? or tumor? or carcinoma?)).ti,ab	5978
7	1 or 2 or 3 or 4 or 5 or 6	234787
8	carcinoembryonic antigen/	25911
9	cea.ti,ab.	22520
10	(carcinoembryonic adj3 antigen?).ti,ab.	13394
11	(carcinoembryonic adj3 antibod*).ti,ab.	657
12	(carcino-embryonic adj3 antigen?).ti,ab.	617
13	(carcino-embryonic adj3 antibod*).ti,ab.	21
14	8 or 9 or 10 or 11 or 12 or 13	36255
15	cancer recurrence/ or tumor recurrence/	119064
16	recurrent disease/ or relapse/	192303
17	recur*.ti,ab.	523223
18	relaps*.ti,ab.	174290

(*Continued*)

19	exp treatment failure/	82867
20	Reoperation/	53394
21	follow up/ and (postoperative care/ or postoperative period/)	38038
22	reoperat*.ti,ab.	31321
23	((local or distant) adj2 failure).ti,ab.	4986
24	((therap* or treatment or surg*) adj3 fail*).ti,ab.	83522
25	((therap* or treatment or surg*) adj3 (respond* or response*)) .ti,ab	167374
26	((postoperat* or post-operat* or postsurg* or post-surg* or post-treat* or post-treat* or posttherap* or post-therap*) adj5 follow up).ti,ab	23063
27	((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) adj5 surveillance).ti,ab	1797
28	((postoperat* or post-operat* or postsurg* or post-surg* or post-treat* or post-treat* or posttherap* or post-therap*) adj5 monitor*).ti,ab	4961
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	1107887
30	7 and 14 and 29	2994
31	(meta-analysis or systematic review or MEDLINE).tw.	144743
32	7 and 14 and 31	78
33	30 not 32	2952

Appendix 4. Science Citation Index & Conference Proceedings Citation Index - Science search strategy:

#1	**TOPIC:** ((colorectal NEAR/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*))) *OR* **TOPIC:** ((colon* NEAR/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*))) *OR* **TOPIC:** ((bowel NEAR/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*))) *OR* **TOPIC:** ((rectal NEAR/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*))) *OR* **TOPIC:** ((rectum NEAR/3 (neoplas* or cancer* or tumour* or tumor* or carcinoma*))) Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	189,742
#2	**TOPIC:** (cea) *OR* **TOPIC:** ((carcinoembryonic NEAR/3 antigen*)) *OR* **TOPIC:** ((carcinoembryonic NEAR/3 antibod*)) *OR* **TOPIC:** ((carcino-embryonic NEAR/3 antigen*)) *OR* **TOPIC:** ((carcino-embryonic NEAR/3 antibod*)) Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	23,879
#3	**TOPIC:** (recur*) *OR* **TOPIC:** (relaps*) *OR* **TOPIC:** (reoperat*) Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	511,568
#4	**TOPIC:** (((local or distant) NEAR/2 failure)) *OR* **TOPIC:** (((therap* or treatment or surg*) NEAR/3 fail*)) *OR* **TOPIC:** (((therap* or treatment or surg*) NEAR/3 (respond* or response*))) Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	200,865
#5	**TOPIC:** (((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) NEAR/5 "follow up")) *OR* **TOPIC:** (((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) NEAR/5 surveillance)) *OR* **TOPIC:** (((postoperat* or post-operat* or postsurg* or post-surg* or posttreat* or post-treat* or posttherap* or post-therap*) NEAR/5 monitor*)) Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	17,719
#6	#5 OR #4 OR #3 Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	699,223
#7	#6 AND #2 AND #1 Indexes=SCI-EXPANDED, CPCI-S Timespan=All years	1,518

Appendix 5. 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		
A: Risk of Bias		
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		
B: Applicability		
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		
A: Risk of Bias		
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		

(*Continued*)

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		
B: Applicability		
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		
A: Risk of Bias		
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		
B: Applicability		

(*Continued*)

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		
A: Risk of Bias		
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?		
	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

NWR and BDN devised the search strategy.

BDN and IP reviewed titles, abstracts, and full-text articles, and extracted all data.

BS acted as moderator at all stages.

BDN, IP, and BS performed the QUADAS-2 assessment.

BS, BDN, and DM devised the statistical analysis.

BS conducted statistical analyses in R, Stata, and SAS.

BDN and BS wrote the initial draft of the review

DM, TJJ, SM, IP, JP, and RP provided comments and edited the draft.

DECLARATIONS OF INTEREST

None

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).

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We stated we would contact the principal investigators to clarify methodological queries and ask for any unpublished data relevant to this review. This has not yet been done, and we have stated this in the Methods section.

We were unable to apply the Hamza method which allows data for multiple thresholds from a single study to be incorporated in the meta-analysis. This method requires 2 x 2 data at consistent thresholds across studies, but in our review accuracy has been reported at a wide range of inconsistent thresholds.

In terms of sensitivity analyses, we did not feel it necessary to remove each study in turn from the analyses as our review includes such a large number of studies, of which none is notably larger than the others, making it high unlikely that one particular study would heavily skew the overall pooled estimates.

INDEX TERMS

Medical Subject Headings (MeSH)

Carcinoembryonic Antigen [*blood]; Colorectal Neoplasms [blood; *diagnosis]; Neoplasm Recurrence, Local [blood; *diagnosis]; Sensitivity and Specificity

MeSH check words

Humans