Diagnostic accuracy of follow-up tests for detecting colorectal cancer recurrences in primary care

Liemburg, Geertje B; Brandenbarg, Daan; Berger, Marjolein Y; Duijts, Saskia F A; Holtman, Gea A; de Bock, Geertruida H; Korevaar, Joke C; Berendsen, Annette J

Published in:
European journal of cancer care

DOI:
10.1111/ecc.13432

Received: 11 November 2020 | Revised: 11 January 2021 | Accepted: 25 February 2021

DOI: 10.1111/ecc.13432

European Journal of Cancer Care WILEY

ORIGINAL ARTICLE

Diagnostic accuracy of follow-up tests for detecting colorectal cancer recurrences in primary care: A systematic review and meta-analysis

Geertje B. Liemburg[1] | Daan Brandenbarg[1] | Marjolein Y. Berger[1] | Saskia F.A. Duijts[1] | Gea A. Holtman[1] | Geertruida H. de Bock[2] | Joke C. Korevaar[3] | Annette J. Berendsen[1]

[1]Department of General Practice & Elderly Care Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[2]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[3]NIVEL Netherlands Institute for Health Services Research, Utrecht, The Netherlands

Correspondence
Daan Brandenbarg, Department of General Practice & Elderly Care Medicine, University Medical Center Groningen PO Box 196, 9700 AD Groningen, The Netherlands.
Email: d.brandenbarg@umcg.nl

Funding information
This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding) [grant number 11062, 2017-1]. The funding source had no role in study design, data collection, analysis and interpretation

Abstract

Introduction: Traditionally, follow-up of colorectal cancer (CRC) is performed in secondary care. In new models of care, the screening part care could be replaced to primary care. We aimed to synthesise evidence on the diagnostic accuracy of commonly used screeners in CRC follow-up applicable in primary care: carcinoembryonic antigen (CEA), ultrasound and physical examination.

Methods: Medline, EMBASE, Cochrane Trial Register and Web of Science databases were systematically searched. Studies were included if they provided sufficient data for a 2 × 2 contingency tables. QUADAS-2 was used to assess methodological quality. We performed bivariate random effects meta-analysis, generated a hypothetical cohort, and reported sensitivity and specificity.

Results: We included 12 studies (n = 3223, median recurrence rate 19.6%). Pooled estimates showed a sensitivity for CEA (≤ 5 μg/l) of 59% [47%–70%] and a specificity of 89% [80%–95%]. Only few studies reported sensitivities and specificities for ultrasound (36–70% and 97–100%, respectively) and clinical examination (23% and 27%, respectively).

Conclusion: In practice, GPs could perform CEA screening. Radiological examination in a hospital setting should remain part of the surveillance strategy. Personalised algorithms accounting for recurrence risk and changes of CEA-values over time might add to the diagnostic value of CEA in primary care.

KEYWORDS
Colorectal cancer, follow-up care, primary care, recurrences

1 | INTRODUCTION

In Western Europe, colorectal cancer (CRC) is the third most diagnosed cancer in males and the second most common in females (Arnold et al., 2015; Bray et al., 2018). In recent decades, improvements in staging and treatment have led to decreased mortality (Arnold et al., 2017; Bray et al., 2018; Grossmann et al., 2014; Torre et al., 2015), which combined with a rising incidence, has led to increasing numbers of CRC survivors (Arnold et al., 2017). This places organisational and financial burdens on follow-up care and

necessitates that we re-evaluate what constitutes the most effective approach (Campbell et al., 2002; Rubin et al., 2015).

Currently, most patients treated with curative intent enter a 5-year hospital-based follow-up program that aims to detect recurrence, monitor late effects of cancer treatment and provide psychological support (Marijnen CAM et al., 2014). Many countries, including the Netherlands, are currently debating whether follow-up can be moved from secondary to primary care (Nekhlyudov et al., 2017). The Dutch College of General Practitioners now supports this position, provided that evidence-based protocols can be provided (The Dutch College of General Practitioners (Nationaal Huisartsen Genootschap), 2014). Timely detection of recurrences, in particular, is thought to be a challenge for GPs. Most guidelines support the use of several tests to detect recurrences, including the carcinoembryonic antigen (CEA) blood test, ultrasound, radiological examinations and colonoscopy (Spronk et al., 2017; Steele et al., 2015), in which blood tests and imaging is mostly used for screening purposes, and colonoscopy for definitive diagnosis. These screening tests may be performed in primary care, to assess which patients to refer for further diagnostic workup in hospitals. CEA is because of its applicability and low costs a possible candidate. Possibly, echography, physical examination or some radiological examinations could be performed or ordered by general practitioners (GPs). Therefore, to assess feasibility of performing CRC follow-up in primary care, it is important to assess the diagnostic performance of these diagnostic routines. Earlier systematic reviews evaluated the role of CEA in CRC follow-up, but these included all available studies (Nicholson et al., 2015). Improved treatments and better staging of the disease may lead to lower recurrence rates, which in turn can effect diagnostic test outcomes.

In this systematic review and meta-analysis, we aimed to synthesise available evidence since 2010 on the diagnostic accuracy of tests commonly used in CRC follow-up. For that, we considered synthesised available evidence on the diagnostic accuracy of CEA, echography and physical examination to detect CRC recurrence in patients curatively treated for CRC.

2 | MATERIAL AND METHODS GRONINGEN

2.1 | Design, search strategy and information sources

This systematic review was registered in PROSPERO (number CRD42018096662) and performed according to the PRISMA guidelines (Moher et al., 2009) and Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi 2010; Reitsma et al., 2009). Up to March 2020, we searched the Medline, EMBASE, Cochrane Trials Register, Web of Science and Trial register databases from 2010 onward, using terms based on 'colorectal cancer', 'curative', 'follow-up' and 'recurrence'. The search strategy was first focused on diagnostic accuracy studies; however, this did not result in actual studies as the focus of the studies in this field are more clinical orientated. A librarian was consulted to develop the search strategy. All references were exported to RefWorks (ProQuest, Bethesda, MD, USA) and duplicates were removed. Reference lists were then hand searched for additional studies.

2.2 | Eligibility criteria

Studies were eligible for inclusion if the following criteria were met: 1) patients were ≥18 years, enrolled in follow-up and had completed CRC treatment with curative intent in any care setting; 2) the number of recurrences during the study period was reported; 3) sufficient data were available to construct or derive a 2 × 2 contingency table for CEA, echography and physical examination as follow-up tests (index test); 4) the reference standard was tumour recurrence by histological, radiological, clinical follow-up or repeated measurements; 5) a randomised controlled trial, clinical trial, cohort or case-control design was used; and 6) full-text articles were available in English, Dutch, German, French or Spanish.

We defined recurrence as loco-regional or distant recurrence during follow-up following previous complete remission. Disease-free survival was defined as no recurrence (negative test result) at the end of follow-up. Tumour stage was based on the TNM Classification of Malignant Tumors, 7th edition, produced by the American Joint Committee on Cancer. Studies reporting Dukes classification were converted into the TNM classification (Marijnen CAM et al., 2014).

2.3 | Data extraction

Two researchers (GBL and DB) independently screened titles and abstracts for eligibility. For full-text appraisal, GBL screened all papers and DB, JCK and SFAD shared the role of second assessor. An independent researcher (AJB) was contacted in case of disagreement. Agreement (percentage) and reliability (Cohen's κ) were calculated to measure interrater reliability (Sim & Wright, 2005).

Predefined data collection forms were used to extract data on study design, setting, patient characteristics, disease and treatment details, follow-up and index tests (with reference standard details), and outcome measures. We used the cut-off values for a positive test result as defined by the authors of the respective studies. All available data were extracted from studies reporting more than one set of data, and authors were contacted to obtain missing data.

2.4 | Risk of bias assessment

Study quality was assessed independently by two researchers (GBL and DB), using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) (Reitsma et al., 2009). Ratings were given on

TABLE 1 Characteristics of the included studies.

Author year	N	Study design	Age (years) mean ±SD / median (range)	% male	Type of cancer n (%)	Tumour staging* (TNM) n (%)		Therapy received n (%)
Augestad 2014	110	Randomised Controlled Trial	65.4 ± 8.1	59.1	Colon 110 (100%)	I: II: III	24 (21.8%) 55 (50.0%) 32 (29.0%)*	Surgery: 110 (100%)
Bhatti 2015	569	Prospective cohort study	70 (range 28–86)	51.0	Colon 336 (59%) Rectum 233 (41%)	0: I: II: III: IV:	31 (5%) 33 (6%) 78 (14%) 288 (51%) 137 (24%)	Surgery: 569 (100%) Chemotherapy: - Neoadjuvant: 106 (18.6%) - Adjuvant 204 (35.9%)
Chang 2017	357	Prospective cohort study	63.8 ± 11.5	57.7	Colon 259 (73%) Rectum 98 (28%)	I: II: III:	97 (27.2%) 125 (35.0%) 135 (37.8%)	Surgery: 357 (100%) Chemotherapy and/or radiotherapy: - Neoadjuvant: 81 (22.7%)
Gilardoni 2015	196	Retrospective cohort study	70 (range 40–89)	59.7	Colon 196 (100%)	I: II:	65 (33.2%) 131 (66.8%)	Surgery: 196 (100%)
Guo 2018	178	Retrospective cohort study	Metastasis: 59.7 ± 10.6 Non metastasis: 58.1 ± 11.8	66.3	Colon 79 (44%) Rectum 99 (56%)	0: I: II: III:	2 (1.1%) 11 (6.2%) 60 (33.7%) 105 (59.0%)	Surgery: 178 (100%) Chemotherapy: 66 (37.1%)
Hara 2011	127	Prospective cohort study	63.4 ± 9.4	55.1	Colon 85 (67%) Rectum 42 (33%)	III:	127 (100%)	Surgery: 127 (100%) Chemotherapy: - Adjuvant: 110 (86.6%)Radiotherapy: - Adjuvant: 1 (0.8%)
Jones 2015	118	Retrospective cohort study	Not reported	61.9	Colon 66 (56%) Rectum 52 (44%)	I: II: III:	26 (22.0%) 47 (39.8%) 45 (38.1%)	Surgery: 118 (100%) Chemotherapy: - Adjuvant: 73 (61.9%)
Kim 2013	336	Retrospective cohort	II: 60.4 ± 11.1 III: 60.4 ± 10.7	60.7	Colon 224 (67%) Rectum 112 (33%)	II: III:	189 (56%) 147 (44%)	Surgery: 336 (100%)
Moloney 2019	138	Retrospective cohort study	Mean 67 (range 37–95)	54.3	Colon 90 (65%) Rectum 48 (35%)	I: II: III:	58 (42.0%) 69 (50.0%) 11 (8.0%)	Surgery: 138 (100%)
Nicolini 2010	108	Prospective cohort study	Mean 60 (range 37–83)	Not reported	Colon 69 (64%) Rectum 39 (36%)	I: II: III:	29 (26.9%) 41 (38.0%) 38 (35.2%)*	Surgery: 108 (100%)

(Continues)

TABLE 1 (Continued)

Author year	N	Study design	Age (years) mean ±SD / median (range)	% male	Type of cancer n (%)	Tumour staging* (TNM) n (%)	Therapy received n (%)
Rodrigues 2017	404	Prospective cohort study	Mean 64.6 (range 25–80)	59.7	Colon 199 (49%) Rectum 205 (51%)	II: 177 (43.8%) III: 227 (56.2%)	Surgery: 404 (100%) Chemotherapy and/or radiotherapy: - Rectum: Neoadjuvant: 175 (85.4%) - Rectum: Adjuvant 196 (95.6%) - Colon: Adjuvant 86 (43.2)
Shinkins 2017	582	Randomised Controlled Trial	<65 years: 185 (31.8%) 65–79 years: 339 (58.2%) 80+ years: 58 (10.0%)	61.3	Subtype not specified	I: 110 (18.9%) II: 282 (48.5%) III: 166 (28.5%) U: 24 (4.1%)*	Surgery: 582 (100%)

Abbreviations:: TNM classification =Tumour, node and metastasis. U = unknown.

*Tumour stage was expressed according to the Dukes classification and is converted into the TNM tumour stage.

four domains: Patient selection, index test, reference standard, and flow and timing. The signalling questions are enclosed in Supplement 2. Discrepancies were resolved by consensus, with a third researcher (AJB) contacted if needed.

2.5 | Statistical analysis for meta-analysis

Data were imported to Review Manager 5.3 (RevMan, Copenhagen, The Cochrane Collaboration, 2014) to calculate the sensitivity, specificity and corresponding 95% confidence intervals (95%CIs) for each test and for each study separately. Heterogeneity was explored by visual examination of forest plots, not by statistical analysis (Macaskill et al., 2010). If heterogeneity was observed, we visually evaluated the following possible explanations: study design, methodological quality, sample size, mean age, gender, CRC subtype, percentage of recurrences, follow-up duration, follow-up protocol, tumour recurrence site, testing frequency, test threshold and the reference standards. Bivariate random effects models were used to calculate pooled estimates of sensitivity and specificity if at least five studies were included for a diagnostic test (Diaz, 2015; Reitsma et al., 2009), using the METANDI module in STATA version 15 (College Station, Texas, USA). Subgroup analyses were performed in which we included prospective studies only.

2.6 | Hypothetical cohort

If possible, we will construct a hypothetical cohort with the aim of showing what the pooled estimates mean in practice for missed recurrences and false positives. We will calculate the median recurrence rate based on the included studies. We will apply the calculated pooled sensitivity and specificity from our meta-analysis, to devise 2x2 tables for 100 patients.

3 | RESULTS

3.1 | Article selection

The initial search yielded 3232 articles. After removing duplicates and screening titles and abstracts, full-text assessment of 73 studies led to the inclusion of 37 articles. Contingency tables could initially be generated for 12 studies (Supplement 1), but authors did provide data for one additional study upon request. Thus, 12 studies were included (Augestad et al., 2014; Bhatti et al., 2015; Chang et al., 2017; Gilardoni et al., 2015; Guo et al., 2018; Hara et al., 2011; Jones et al., 2015; Kim & Lee, 2013; Moloney et al., 2019; Nicolini et al., 2010; Rodrigues et al., 2017; Shinkins et al., 2017). Agreement between researchers was 'very good' for title and abstract selection (98%; κ, 0.889) and 'good' for the full-text assessment (86%; κ, 0.703).

TABLE 2 Diagnostic value of follow-up routines for detecting recurrence, as measured by contingency tables.

Author +year	N	% recurrences	Duration of follow-up (months)	Index test	Frequency*	Reference standard#	TP	FP	FN	TN
CEA										
Augestad 2014	110	12.7	Median 17	CEA threshold 5 μg/l	3–6 monthly[a]	Contrast-enhanced ultrasound, PET scan, CT scan thorax/abdomen or colonoscopy	3	6	4	97
Bhatti 2015	569	26.2	Median 40 (range 6–50)	CEA threshold 5 μg/l	6-monthly	CT scan and/or PET scan with or without biopsy	123	23	26	397
Chang 2017	357	18.8	Median 31 (range 3–64)	CEA threshold 5 μg/l	3–6 monthly[b]	CT scan: abdominal or chest	44	59	23	231
Hara 2011	127	36.2	Not reported	CEA threshold 5 μg/l	3-monthly	Radiological examination†	31	31	15	50
Guo 2018	178	51.7	Not reported	CEA threshold 5 μg/l	2–6 monthly[c]	Radiological examination with X-ray and abdominopelvic CT scan, image-guided biopsy, or exploratory laparotomy	66	36	26	50
Kim 2013	336	23.5	Median 45 (range 36–134)	CEA threshold 5 μg/l	3–6 monthly[a]	Biopsy and radiological examination†	34	23	45	234
Moloney 2019	138	4.3	Mean 25.3 ± 18.5	CEA threshold not reported	4–6 monthly[d]	CT scan, ultrasound or colonoscopy	2	5	4	127
Rodrigues 2017	404	12.9	Mean 37 (range 3–79)	CEA 3 μg/l for non-smokers and 5 μg/l for smokers	3–6 monthly[b]	Radiological examinations (liver and lung)† or biopsy	23	36	29	316
Shinkins 2017	582	17.9	Not reported	CEA threshold 5 μg/l	3–6 monthly[a]	CT scan thorax/abdomen/pelvis, clinical examination or colonoscopy	51	12	53	466
Ultrasound										
Gilardoni 2015	196	5.6	At least 60	Contrast-enhanced ultrasound: abdomen	Annually	Contrast-enhanced CT scan	4	5	7	180
Nicolini 2010	108	20.4	Mean 99 ± 57 (Range 13–179)	Ultrasound: abdomen	8-monthly	CT scan or MRI scan and if necessary cytohistological	14	1	6	744
Clinical examination										
Jones 2015	118	22.0	Median 36	Clinical examination	3 monthly annually[e]	CT scan	6	16	20	6

Abbreviations:: CEA, Carcinoembryonic antigen; CT scan, Computed Tomography; FN, false negatives; FP, false positives; MRI scan, Magnetic resonance imaging scan; PET scan, Positron-emission tomography scan; TN, true negatives; TP, true positives.

*Specified index test frequency: [a]Year 1 + 2: 3 monthly, year 3–5: 6-monthly; [b]Year 1–3: 3 monthly, >3 years: 6 monthly; [c]Year 1 + 2: 2 monthly, year 3–6: 6-monthly; [d]Year 1 + 2: 4-monthly, year 3–5: 6-monthly; [e]Year 1: 3–monthly, year 2: 6-monthly, year 3–5: annually. #Studies used more than one reference standard, so not all patients received the same reference standard to confirm a recurrence.
†Radiological examination not further specified.

FIGURE 1 Risk of bias and applicability concerns by QUADAS-2 domain for each study.

3.2 | Patients

Overall, 3223 patients (males, 58.8%; age range 25–95 years) were included (Table 1). All studies were from secondary care and included 108–569 patients. Most patients were diagnosed with colon cancer (62.6%); although this was typically stage II or III, two studies included patients with stage IV cancer who were considered cured after resecting liver metastases. The central tendencies for the reported follow-up durations ranged from median 17 to 99 months, and the median recurrence rates based on all included studies was 19.6% (4.3%–51.7%), which appeared independent of follow-up duration.

3.3 | Index tests and reference standards

The most commonly reported follow-up tests were CEA (9 studies), ultrasound (2 studies) and physical examination (1 study). For CEA, all but one (Rodrigues et al., 2017) study used a cut-off value of 5 μg/l for all patients and one study did not report their threshold (Moloney et al., 2019). Reference standards varied widely, consisting mostly of different radiological and histopathological examinations. Multiple reference standards were used in seven studies (Table 2).

3.4 | Methodological quality of included studies

Risk of bias was highest in the patient selection (5 studies) and reference standard (6 studies) domains

(Figure 1). However, risk of bias was unclear in the index test domain. Issues in the patient selection domain resulted from inappropriate exclusions, while issues with the flow and timing domain resulted from variations in the reference standards (Bhatti et al., 2015; Chang et al., 2017; Guo et al., 2018).

3.5 | Diagnostic accuracy

Forest plots are shown for the different follow-up tests in Figure 2.

For CEA (9 studies, 2,801 patients), the sensitivity and specificity for detecting CRC recurrence were 33–83% and 58%–97%, respectively. The pooled sensitivity was 59% (95%CI: 47%–70%), and the pooled specificity was 89% (95%CI: 80%–95%) (Figure 3). For sensitivity, we observed two outliers, reporting a lower sensitivity with broad confidence intervals (Augestad et al., 2014; Moloney et al., 2019). Also, two outliers were observed for specificity (Guo et al., 2018; Hara et al., 2011), reporting a lower specificity. A subgroup analysis without retrospective cohort studies (6 studies, 2149 patients), showed a pooled sensitivity of 62% (95%CI: 48%–74%) and a pooled specificity of 90% (95%CI: 80%–96%) (data not shown).

For ultrasound (2 studies, 901 patients), the ranges for sensitivity and specificity were 36%–70% and 97%–100%, respectively. Although the forest plots showed narrow CIs for specificity, they were wide for sensitivity. For clinical examination (one study, 118 patients), the sensitivity was 23% and the specificity 27%.

3.6 | Hypothetical cohort

We were able to construct a hypothetical cohort based on the 9 studies on CEA. The overall median prevalence of tumour recurrence was 21% given a median follow-up between 2 and 5 years. Using CEA to detect recurrence misclassified 18 cases, with 9 of the 21 recurrences being missed and 9 of 79 patients receiving unnecessary follow-up testing (Table 3).

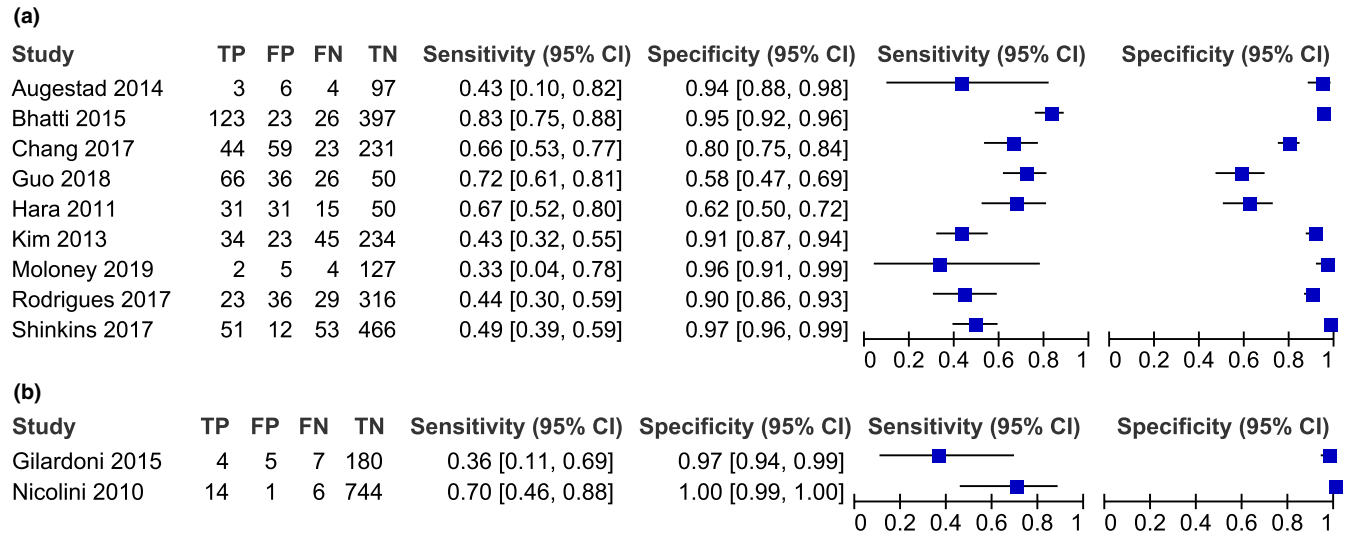

FIGURE 2 Forest plots displaying the diagnostic accuracy of different follow-up routines: CEA testing (a) and Ultrasound (b). Diagnostic accuracy and the 2 × 2 table are displayed with true positives (TP), false positives (FP), false negatives (FN) and true negative (TN). Sensitivity and specificity with the 95%-confidence intervals (CI) are given.

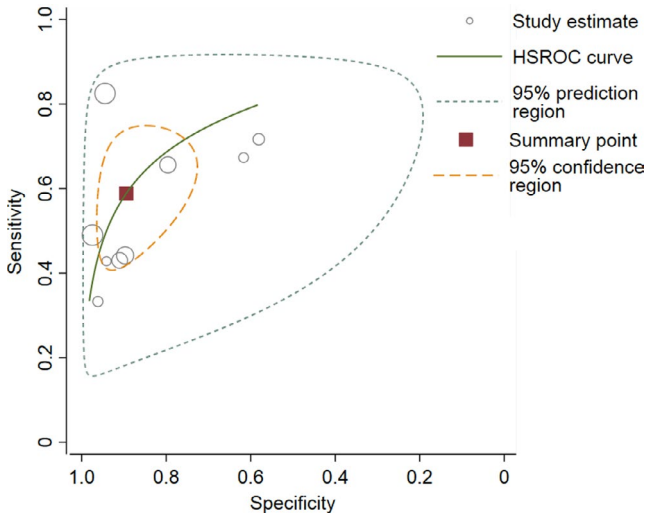

FIGURE 3 Hierarchical summary receiver operator curves Pooled estimates are shown for CEA. Abbreviations: CEA, Carcinoembryonic antigen; HSROC =hierarchical summary receiver operating characteristic.

TABLE 3 Pooled estimates for CEA follow-up tests for detecting recurrence in a hypothetical cohort

	CEA
Number of studies	9[a]
Prevalence (%)*	21%
Sensitivity (95%CI)	59% (47–70)
Specificity (95%CI)	89% (80–95)
Misclassified	18
Missed cancers	9
Unnecessary diagnostic follow-up testing	9

The hypothetical cohort comprised 100 survivors of colorectal cancer and was based on the pooled diagnostic accuracy. Abbreviations: CEA, Carcinoembryonic antigen.

[a]based on 2801 patients.

4 | DISCUSSION

4.1 | Main findings

Most of the included studies evaluated follow-up testing by CEA. The pooled estimates showed a sensitivity for CEA of 59% and a specificity of 89%. A hypothetical cohort of 100 patients, based on the pooled characteristics of all included studies, revealed that CEA misclassified 18 of 100 cases, with 9 of 21 recurrences being missed. For ultrasound, sensitivity and specificity ranged from 36%–70% to 97%–100%, respectively. For clinical examination, the sensitivity was 23% and the specificity 27%.

4.2 | Limitations

This comprehensive review provides an overview of the diagnostic accuracy in secondary care of different follow-up tests for detecting CRC recurrence that potentially can be applied in primary care. Because we only included studies from the past ten years, the diagnostic properties of the tests we evaluated correspond to the current recurrence rates. This focus also posed a limitation, however, the displayed meta-analysis (Figure 3) should be interpreted with caution because, ideally, at least ten studies should be included to achieve balance across the estimates (Diaz, 2015). Furthermore, studies applied different reference standards, but this effect should be negligible given that both histological and radiological standards are used widely and considered reliable. In addition, included studies showed different durations of follow-up, which may have influenced recurrence rates. However, heterogeneity could not be explained by this. Studies reporting a low sensitivity for CEA generally showed a lower recurrence rate and less advanced cancer stages. In contrast, studies reporting low specificity for CEA showed higher recurrence rates and cancer stages. Finally, the overall methodological quality of the included studies was low. However, it should be noted that none of the studies aimed to evaluate the characteristics of the index tests exclusively.

4.3 | Comparison with existing literature

Authors of a systematic review in 2015 reported a slightly higher pooled sensitivity (64.5%) and a comparable specificity (89.5%) for CEA at a threshold of 5 μg/L (Nicholson et al., 2015) when compared with our results (59% and 89%, respectively). The difference in sensitivity may be explained by the higher prevalence of CRC, and therefore difference in population when compared to our study. Given the advance in treatments in recent decades, the population of patient with recurrent CRC in our study period may be different to that historically diagnosed with a CRC recurrence. Another systematic review from 2016 showing results comparable to Nicholson et al. also included patients from before 2010 (Sorensen et al., 2016).

Interestingly, a recent randomised controlled trial, reporting a low prevalence of recurrences (16.6%), suggests that frequent monitoring of CEA is as good as intensified imaging with computed tomography (CT) (Primrose et al., 2014). In that study, the patient's general practice physician referred the patient urgently to the local hospital if a patient's blood CEA level was 7 μg/L or more above the level at trial entry, and a second test result was also greater than this threshold.

Although we only included a small number of studies based on ultrasound, this follow-up test had a high specificity, but appeared to be lacking sensitivity. Clinical examination did not seem feasible to be used as a diagnostic strategy for detecting recurrences given the low sensitivity and specificity. We found no systematic reviews of ultrasound or clinical examination being used to detect recurrence during follow-up.

4.4 | Implications

Ideal follow-up routines are cost-efficient, sensitive for detecting recurrence and specific for identifying patients without recurrence. At

present, the survival benefit of follow-up protocols remains a topic of debate. Some studies indicate that intensified follow-up based on CEA monitoring leads to earlier detection of recurrences and to higher cure rates (Primrose et al., 2014; Verberne et al., 2015), whereas others have shown that intensive follow-up routines confer no survival benefits (Jeffery et al., 2016; Mant et al., 2017; Rosati et al., 2016; Wille-Jorgensen et al., 2018). Only one study compared follow-up using CEA or CT to the value of self-reported detection, and this showed worse survival (Verberne et al., 2017). Furthermore, recent studies suggest that individualised CEA levels, using multiple measurements and accounting for pre-operative levels, increase the sensitivity of CEA and therefore its use as a screening marker (Hida et al., 2017; Jeon et al., 2013; Saito et al., 2017; Shinkins et al., 2018). If this evidence proves to be robust, there is scope for involvement of primary care in follow-up, since CRC survivors already consult their GP more often (Brandenbarg et al., 2017).

The finding that 42% of all recurrences is detected during non-scheduled follow-up visits, mostly based on clinical symptoms (Duineveld et al., 2016), adds to the relevance of involving primary care in CRC follow-up. In countries like the Netherlands, in which the general practitioner acts as a gatekeeper to secondary care, these patients are likely to present in primary care. Furthermore, GPs already play a role in the other goals of follow-up care; provision of psychological support and monitoring treatment-related side effects. To implement a possible shared care protocol between settings, we must have a reliable and highly sensitive test strategy that balances missing recurrences and producing false positives that may cause psychological distress (van der Velde et al., 2017). We recommend that future studies focus on evaluating the diagnostic accuracy of test combinations, investigating strategies to increase the accuracy of CEA (e.g. developing personalised algorithms that account for recurrence risk factors and changes over time), and exploring the value of using reported clinical symptoms during check-ups to increase accuracy (Rose et al., 2019).

5 | CONCLUSION

This systematic review and meta-analysis showed that none of the investigated follow-up tests, applicable in primary care, was adequate for detecting recurrence when used in isolation. The use of CEA with a threshold of 5 μg/L results in missing about half of the recurrences, and therefore lacks sensitivity to be used as a single screener. The other diagnostic tests available in primary care, ultrasound and clinical examination, lack diagnostic accuracy to be used as follow-up diagnostic tests, so radiological and possibly endoscopic examination in a hospital setting should remain part of the surveillance strategy. In practice, GPs could perform CEA tests, so this might be implemented in primary care in a shared care model as the target population with curative treated CRC stays the same. We suggest that our data can be used to guide further review of test options for CRC recurrence and before such an implementation is discussable, future studies should investigate the influence of using personalised algorithms that account for recurrence risk factors and changes over time for repeated measurements.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank Prof Dr L. Gianotti for providing us with additional data for the paper of Gilardoni et al., 2015. We would also like to thank all authors who responded to our data request for their efforts to look whether they were able to provide additional data. We would like to thank Sjoukje van der Werf for her help with formulating the search string and the actual literature search. Dr Robert Sykes (www.doctored.org.uk) provided technical editing and writing services for the final draft of this manuscript.

ETHICAL STATEMENT

Ethical approval is not required for the study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Daan Brandenbarg https://orcid.org/0000-0001-8054-7337
Saskia F.A. Duijts https://orcid.org/0000-0001-6025-5361

REFERENCES

Arnold, M., Karim-Kos, H. E., Coebergh, J. W., Byrnes, G., Antilla, A., Ferlay, J., … Soerjomataram, I. (2015). Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *European Journal of Cancer*, *51*(9), 1164–1187. https://doi.org/10.1016/j.ejca.2013.09.002.

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, *66*(4), 683–691. https://doi.org/10.1136/gutjnl-2015-310912.

Augestad, K. M., Norum, J., Rose, J., & Lindsetmo, R. O. (2014). A prospective analysis of false positive events in a national colon cancer surveillance program. *BMC Health Services Research*, *14*, 137–6963-14-137. https://doi.org/10.1186/1472-6963-14-137.

Bhatti, I., Patel, M., Dennison, A. R., Thomas, M. W., & Garcea, G. (2015). Utility of postoperative CEA for surveillance of recurrence after resection of primary colorectal cancer. *International Journal of Surgery (London, England)*, *16*(Pt A), 123–128.S1743-9191(15)00080-1 [pii].

Brandenbarg, D., Roorda, C., Groenhof, F., de Bock, G. H., Berger, M. Y., & Berendsen, A. J. Primary healthcare use during follow-up after curative treatment for colorectal cancer. *European Journal of Cancer Care*, *26*(3), e12581. https://doi.org/10.1111/ecc.12581.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492.

Campbell, N. C., MacLeod, U., & Weller, D. (2002). Primary care oncology: Essential if high quality cancer care is to be achieved for all. *Family Practice*, *19*(6), 577–578.

Chang, Y. T., Yeh, Y. S., Ma, C. J., Huang, C. W., Tsai, H. L., Huang, M. Y., & Wang, J. Y. (2017). Optimization of a multigene biochip for detection of relapsed and early relapsed colorectal cancer. *The Journal of Surgical Research*, *220*, 427–437.S0022-4804(17)30416-X [pii].

Diaz, M. (2015). Performance measures of the bivariate random effects model for meta-analses of diagnostic accuracy. *Computational Statistics & Data Analysis*, *83*, 82–90.

Duineveld, L. A., van Asselt, K. M., Bemelman, W. A., Smits, A. B., Tanis, P. J., van Weert, H. C., & Wind, J. (2016). Symptomatic and asymptomatic colon cancer recurrence: A multicenter cohort study. *Annals of Family Medicine*, *14*(3), 215–220. https://doi.org/10.1370/afm.1919.

Gilardoni, E., Bernasconi, D. P., Poli, S., Garancini, M., Luperto, M., Zucchini, N., & Gianotti, L. (2015). Surveillance for early stages of colon cancer: Potentials for optimizing follow-up protocols. *World Journal of Surgical Oncology*, *13*, 260–015-0674-7. https://doi.org/10.1186/s12957-015-0674-7.

Grossmann, I., Doornbos, P. M., Klaase, J. M., de Bock, G. H., & Wiggers, T. (2014). Changing patterns of recurrent disease in colorectal cancer. *European Journal of Surgical Oncology : the Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, *40*(2), 234–239. https://doi.org/10.1016/j.ejso.2013.10.028.

Guo, Y., Chen, F., & Cui, W. (2018). Usefulness of plasma D-dimer level for monitoring development of distant organ metastasis in colorectal cancer patients after curative resection. *Cancer Management and Research*, *10*, 4203–4216. https://doi.org/10.2147/CMAR.S177274.

Hara, M., Sato, M., Takahashi, H., Takayama, S., & Takeyama, H. (2011). Accuracy of monitoring serum carcinoembryonic antigen levels in postoperative stage III colorectal cancer patients is limited to only the first postoperative year. *Surgery Today*, *41*(10), 1357–1362. https://doi.org/10.1007/s00595-010-4519-2.

Hida, K., Okamura, R., Park, S. Y., Nishigori, T., Takahashi, R., Kawada, K., & Sakai, Y. (2017). A new prediction model for local recurrence after curative rectal cancer surgery: Development and validation as an asian collaborative study. *Diseases of the Colon and Rectum*, *60*(11), 1168–1174. https://doi.org/10.1097/DCR.0000000000000896.

Jeffery, M., Hickey, B. E., Hider, P. N., & See, A. M. (2016). Follow-up strategies for patients treated for non-metastatic colorectal cancer. *The Cochrane Database of Systematic Reviews*, *11*, CD002200. https://doi.org/10.1002/14651858.CD002200.pub3.

Jeon, B. G., Shin, R., Chung, J. K., Jung, I. M., & Heo, S. C. (2013). Individualized cutoff value of the preoperative carcinoembryonic antigen level is necessary for optimal use as a prognostic marker. *Annals of Coloproctology*, *29*(3), 106–114. https://doi.org/10.3393/ac.2013.29.3.106.

Jones, R. P., McWhirter, D., Fretwell, V. L., McAvoy, A., & Hardman, J. G. (2015). Clinical follow-up does not improve survival after resection of stage I-III colorectal cancer: A cohort study. *International Journal of Surgery*, *17*(5), 67–71.

Kim, H. S., & Lee, M. R. (2013). Diagnostic accuracy of elevated serum carcinoembryonic antigen for recurrence in postoperative stage II colorectal cancer patients: Comparison with stage III. *Annals of Coloproctology*, *29*(4), 155–159. https://doi.org/10.3393/ac.2013.29.4.155.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. Chapter 10: Analysing and presenting results. In Deeks J. J., bossuyt P. M., gatsonis C. (editors), *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0. the cochrane collaboration*, *2010*. Available from: http://Srdta.cochrane.org/.

Mant, D., Gray, A., Pugh, S., Campbell, H., George, S., Fuller, A., & Primrose, J. (2017). A randomised controlled trial to assess the cost-effectiveness of intensive versus no scheduled follow-up in patients who have undergone resection for colorectal cancer with curative intent. *Health Technology Assessment (Winchester, England)*, *21*(32), 1–86. https://doi.org/10.3310/hta21320.

Marijnen, C. A. M., Van Vegchel, T., Sleutel, A. Y., Beets-Tan, R. G. H., Nagtegaan, I. D., Peters, F. T. M. et al (2014). Landelijke richtlijn colorectaal carcinoom. Retrieved from http://www.oncoline.nl/colorectaalcarcinoom.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Moloney, J., Partridge, C., Delanty, S., Lloyd, D., & Nguyen, M. H. (2019). High efficacy and patient satisfaction with a nurse-led colorectal cancer surveillance programme with 10-year follow-up. *ANZ Journal of Surgery*, *89*(10), 1286–1290. https://doi.org/10.1111/ans.15333.

Nekhlyudov, L., O'malley, D. M., & Hudson, S. V. (2017). Integrating primary care providers in the care of cancer survivors: Gaps in evidence and future opportunities. *The Lancet Oncology*, *18*(1), e30–e38.S1470-2045(16)30570-8 [pii].

Nicholson, B. D., Shinkins, B., Pathiraja, I., Roberts, N. W., James, T. J., Mallett, S., & Mant, D. (2015). Blood CEA levels for detecting recurrent colorectal cancer. *The Cochrane Database of Systematic Reviews*, (12):CD011134. https://doi.org/10.1002/14651858.CD011134.pub2.

Nicolini, A., Ferrari, P., Duffy, M. J., Antonelli, A., Rossi, G., Metelli, M. R., … Miccoli, P. (2010). 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 (Chicago, Ill.: 1960)*, *145*(12), 1177-1183. https://doi.org/10.1001/archsurg.2010.251

Primrose, J. N., Perera, R., Gray, A., Rose, P., Fuller, A., Corkhill, A., & FACS Trial Investigators, (2014). 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*, *311*(3), 263–270. https://doi.org/10.1001/jama.2013.285718.

Reitsma, J., Rutjes, A., Whiting, P., Vlassov, V., Leeflang, M., & Deeks, J. (2009). Chapter 9: Assessing methodological quality. In: Deeks J. J., Bossuyt P. M., Gatsonis C. (editors), *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0*. The Cochrane Collaboration. Available from: http://Srdta.cochrane.org/.

Rodrigues, R. V., Pereira da Silva, J., Rosa, I., Santos, I., Pereira, N., Soares, C., & Pereira, A. D. (2017). Intensive follow-up after curative surgery for colorectal cancer. *Acta Medica Portuguesa*, *30*(9), 633–641. https://doi.org/10.20344/amp.7889.

Rosati, G., Ambrosini, G., Barni, S., Andreoni, B., Corradini, G., Luchena, G., & GILDA working group. (2016). A randomized trial of intensive versus minimal surveillance of patients with resected dukes B2-C colorectal carcinoma. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *27*(2), 274–280. https://doi.org/10.1093/annonc/mdv541.

Rose, J., Homa, L., Kong, C. Y., Cooper, G. S., Kattan, M. W., Ermlich, B. O., & Meropol, N. J. (2019). Development and validation of a model to predict outcomes of colon cancer surveillance. *Cancer Causes & Control: CCC*, *30*(7), 767–778. https://doi.org/10.1007/s10552-019-01187-x.

Rubin, G., Berendsen, A., Crawford, S. M., Dommett, R., Earle, C., Emery, J., & Zimmermann, C. (2015). The expanding role of primary care in cancer control. *The Lancet. Oncology*, *16*(12), 1231–1272. https://doi.org/10.1016/S1470-2045(15)00205-3.

Saito, G., Sadahiro, S., Kamata, H., Miyakita, H., Okada, K., Tanaka, A., & Suzuki, T. (2017). Monitoring of serum carcinoembryonic antigen levels after curative resection of colon cancer: Cutoff values determined according to preoperative levels enhance the diagnostic accuracy for recurrence. *Oncology*, *92*(5), 276–282. https://doi.org/10.1159/000456075.

Shinkins, B., Nicholson, B. D., Primrose, J., Perera, R., James, T., Pugh, S., & Mant, D. (2017). The diagnostic accuracy of a single CEA blood test in detecting colorectal cancer recurrence: Results from the FACS trial. *PLoS One*, *12*(3), e0171810. https://doi.org/10.1371/journal.pone.0171810.

Shinkins, B., Primrose, J. N., Pugh, S. A., Nicholson, B. D., Perera, R., James, T., & Mant, D. (2018). Serum carcinoembryonic antigen trends for diagnosing colorectal cancer recurrence in the FACS randomized clinical trial. *The British Journal of Surgery*, *105*(6), 658–662. https://doi.org/10.1002/bjs.10819.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268.

Sorensen, C. G., Karlsson, W. K., Pommergaard, H. C., Burcharth, J., & Rosenberg, J. (2016). The diagnostic accuracy of carcinoembryonic antigen to detect colorectal cancer recurrence - A systematic review. *International Journal of Surgery (London, England)*, *25*, 134–144. https://doi.org/10.1016/j.ijsu.2015.11.065.

Spronk, I., Korevaar, J. C., Schellevis, F. G., Albreht, T., & Burgers, J. S. (2017). Evidence-based recommendations on care for breast cancer survivors for primary care providers: A review of evidence-based breast cancer guidelines. *British Medical Journal Open*, *7*(12), e015118. 2016-015118. https://doi.org/10.1136/bmjopen-2016-015118.

Steele, S. R., Chang, G. J., Hendren, S., Weiser, M., Irani, J., Buie, W. D., … Clinical Practice Guidelines Committee of the American Society of Colon and Rectal Surgeons. (2015). Practice guideline for the surveillance of patients after curative treatment of colon and rectal cancer. *Diseases of the Colon and Rectum*, *58*(8), 713–725. https://doi.org/10.1097/DCR.0000000000000410.

The Dutch College of General Practitioners (Nationaal Huisartsen Genootschap). (2014). NHG-standpunt oncologische zorg in de huisartsenpraktijk. Retrieved from https://www.nhg.org/themas/publicaties/nhg-standpunt-oncologische-zorg-de-huisartsenpraktijk

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, *65*(2), 87–108. https://doi.org/10.3322/caac.21262.

van der Velde, J. L., Blanker, M. H., Stegmann, M. E., de Bock, G. H., Berger, M. Y., & Berendsen, A. J. (2017). A systematic review of the psychological impact of false-positive colorectal cancer screening: What is the role of the general practitioner? *European Journal of Cancer Care*, *26*(3), https://doi.org/10.1111/ecc.12709. Epub 2017 May 11.

Verberne, C. J., Zhan, Z., van den Heuvel, E., Grossmann, I., Doornbos, P. M., Havenga, K., & Wiggers, T. (2015). Intensified follow-up in colorectal cancer patients using frequent carcino-embryonic antigen (CEA) measurements and CEA-triggered imaging: Results of the randomized "CEAwatch" trial. *European Journal of Surgical Oncology: the Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, *41*(9), 1188–1196. https://doi.org/10.1016/j.ejso.2015.06.008.

Verberne, C. J., Zhan, Z., van den Heuvel, E. R., Oppers, F., de Jong, A. M., Grossmann, I., & Wiggers, T. (2017). Survival analysis of the CEAwatch multicentre clustered randomized trial. *The British Journal of Surgery*, *104*(8), 1069–1077. https://doi.org/10.1002/bjs.10535.

Wille-Jorgensen, P., Syk, I., Smedh, K., Laurberg, S., Nielsen, D. T., Petersen, S. H., … COLOFOL Study Group (2018). Effect of more vs less frequent follow-up testing on overall and colorectal cancer-specific mortality in patients with stage II or III colorectal cancer: The COLOFOL randomized clinical trial. *JAMA*, *319*(20), 2095–2103. https://doi.org/10.1001/jama.2018.5623.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Liemburg GB, Brandenbarg D, Berger MY, et al. Diagnostic accuracy of follow-up tests for detecting colorectal cancer recurrences in primary care: A systematic review and meta-analysis. *Eur J Cancer Care*. 2021;00:e13432. https://doi.org/10.1111/ecc.13432