



## UvA-DARE (Digital Academic Repository)

### Does a computer-aided detection algorithm in a second read paradigm enhance the performance of experienced computed tomography colonography readers in a population of increased risk?

de Vries, A.H.; Jensch, S.; Liedenbaum, M.H.; Florie, J.; Nio, C.Y.; Truyen, R.; Bipat, S.; Dekker, E.; Fockens, P.; Baak, L.C.; Stoker, J.

#### DOI

[10.1007/s00330-008-1215-3](https://doi.org/10.1007/s00330-008-1215-3)

#### Publication date

2009

#### Document Version

Final published version

#### Published in

European Radiology

[Link to publication](#)

#### Citation for published version (APA):

de Vries, A. H., Jensch, S., Liedenbaum, M. H., Florie, J., Nio, C. Y., Truyen, R., Bipat, S., Dekker, E., Fockens, P., Baak, L. C., & Stoker, J. (2009). Does a computer-aided detection algorithm in a second read paradigm enhance the performance of experienced computed tomography colonography readers in a population of increased risk? *European Radiology*, 19(4), 941-950. <https://doi.org/10.1007/s00330-008-1215-3>

#### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Ayso H. de Vries  
Sebastiaan Jensch  
Marjolein H. Liedenbaum  
Jasper Florie  
Chung Y. Nio  
Roel Truyen  
Shandra Bipat  
Evelien Dekker  
Paul Fockens  
Lubbertus C. Baak  
Jaap Stoker

## Does a computer-aided detection algorithm in a second read paradigm enhance the performance of experienced computed tomography colonography readers in a population of increased risk?

Received: 20 May 2008  
Revised: 26 August 2008  
Accepted: 27 September 2008  
Published online: 4 November 2008  
© The Author(s) 2008. This article is published with open access at Springerlink.com

L. C. Baak  
Department of Gastroenterology,  
Onze Lieve Vrouwe Gasthuis,  
Oosterpark 9,  
1091 AC Amsterdam, The Netherlands

A. H. de Vries (✉) · S. Jensch ·  
M. H. Liedenbaum · J. Florie ·  
C. Y. Nio · S. Bipat · J. Stoker  
Department of Radiology,  
Academic Medical Center,  
University of Amsterdam,  
Meibergdreef 9,  
1105 AZ Amsterdam, The Netherlands  
e-mail: Ayso.devries@amc.uva.nl  
Tel.: +31-20-56663028  
Fax: +31-20-56669119

S. Jensch  
Department of Radiology,  
Onze Lieve Vrouwe Gasthuis,  
Oosterpark 9,  
1091 AC Amsterdam, The Netherlands

R. Truyen  
Department of Healthcare Informatics,  
Philips Medical Systems,  
Veenpluis 6,  
5684 PC Best, The Netherlands

E. Dekker · P. Fockens  
Department of Gastroenterology,  
Academic Medical Center,  
University of Amsterdam,  
Meibergdreef 9,  
1105 AZ Amsterdam, The Netherlands

**Abstract** We prospectively determined whether computer-aided detection (CAD) could improve the performance characteristics of computed tomography colonography (CTC) in a population of increased risk for colorectal cancer. Therefore, we included 170 consecutive patients that underwent both CTC and colonoscopy. All findings  $\geq 6$  mm were evaluated at colonoscopy by segmental unblinding. We determined per-patient sensitivity and specificity for polyps  $\geq 6$  mm and  $\geq 10$  mm *without* and *with* computer-aided detection (CAD). The McNemar test was used for comparison the results without and with CAD. Unblinded colonoscopy detected 50 patients with lesions  $\geq 6$  mm and 25 patients with lesions  $\geq 10$  mm. Sensitivity of CTC without CAD for these size categories was

80% (40/50, 95% CI: 69–81%) and 64% (16/25, 95% CI: 45–83%), respectively. CTC with CAD detected one additional patient with a lesion  $\geq 6$  mm and two with a lesion  $\geq 10$  mm, resulting in a sensitivity of 82% (41/50, 95% CI: 71–93%) ( $p=0.50$ ) and 72% (18/25, 95% CI: 54–90%) ( $p=1.0$ ), respectively. Specificity without CAD for polyps  $\geq 6$  mm and  $\geq 10$  mm was 84% (101/120, 95% CI: 78–91%) and 94% (136/145, 95% CI: 90–98%), respectively. With CAD, the specificity remained (nearly) unchanged: 83% (99/120, 95% CI: 76–89%) and 94% (136/145, 95% CI: 90–98%), respectively. Thus, although CTC with CAD detected a few more patients than CTC without CAD, it had no statistically significant positive influence on CTC performance.

**Keywords** Colon ·  
CT colonography · Computer-aided  
detection · Virtual · Colonoscopy ·  
Flat lesions

### Introduction

Computed tomography colonography (CTC) has consistently shown to have a high accuracy for colorectal neoplasia, and has recently been included in the official guidelines for colorectal cancer screening [1].

In the past years, efforts have been made to in order to increase its accuracy, e.g., labeling fecal material with a contrast agent (fecal tagging), automatic insufflation and improvement of workstations. Despite these efforts, visible lesions are still missed, even by well-trained radiologists.

Computer-aided detection (CAD) is a promising technique [2–5] that could be helpful in reducing these false-negative findings [6, 7]. However, even if the CAD performance would be excellent, it does not automatically translate into equivalent reader performance [8, 9], i.e., CAD hits can be disregarded by the observer. This stresses the complex interaction between CAD and the observers.

Recent studies concluded that in a selected population CAD significantly improved per-polyp sensitivity for less experienced observers [10–13]. Though experienced observers benefited proportionally less from CAD [14, 15]. Therefore, the potential increase by CAD in accuracy for experienced observers is still controversial.

The additional value of CAD was tested in a selected and polyp-enriched population only. This may have a positive effect on the observer performance since observers may be more easily triggered to detect polyps. Secondly, the a priori chance that a finding is indeed a polyp has increased. Therefore, the additional value of CAD (that will have a similar detection pattern irrespective of the population) may be larger.

To our knowledge, the effect of CAD on the performance of observers has not been prospectively evaluated in an unselected patient population of increased risk for colorectal cancer. Therefore, the purpose of this study was to determine whether CAD in a second read paradigm could improve the performance characteristics in a practical setting. Based on indirect comparison of two experienced observers and CAD [16], we hypothesized that CAD could still improve experienced observer performance.

## Materials and methods

The institutional review board of both hospitals approved the study. All patients gave written informed consent.

### Study population

Consecutive patients with a personal or family history of colorectal polyps or cancer were invited to participate from February 2006 until July 2007. All patients were scheduled to undergo a routine colonoscopy at one or other of the two participating hospitals. Exclusion criteria were: age under 18 years, pregnancy, personal history of inflammatory bowel disease, familial adenomatous polyposis, Peutz-Jeghers syndrome, hereditary non-polyposis colorectal cancer, prior allergic reaction to iodine contrast, untreated hyperthyroidism, known colorectal polyps that were not removed at an earlier endoscopy.

Patients ingested 4 l polyethylene glycol electrolyte solution (KleanPrep; Helsinn Birex Pharmaceuticals, Dublin, Ireland) on the day before and the day of the examinations. If contraindicated, other regimes were used. Patients ingested 50 ml oral iodine contrast

(ioxithalamate, 300 mg ml<sup>-1</sup>) (Telebrix, Guerbet, Roissy, France) with each liter of polyethylene glycol electrolyte solution.

All CT-examinations were performed on two different CT systems. The CT parameters for the four-slice CT were 120 kV, 50 mAs (abdominal circumference ≤ 103 cm) or 70 mAs (>103 cm), effective slice thickness 3.2 mm, pitch 1.25 and reconstruction interval 1.6 mm. The CT-parameters for the 64-slice CT were 120 kV, 58 mAs (abdominal circumference ≤ 103 cm) or 82 mAs (>103 cm), effective slice thickness 0.9 mm, pitch 0.984 and reconstruction interval 0.7 mm. Procedural details and baseline characteristics are listed in Table 1.

### Observers

The CTC examinations were evaluated by one observer of a group of five observers; one board certified abdominal radiologist, two radiology residents (2nd and 4th year) and two radiology research fellows. The observers read the CTC examinations in a quiet environment not pressured to provide rapid reports, although they knew that the colonoscopy would be performed within 3 h. They were blinded to clinical data.

Although their experience varied, all had seen at least 100 CTC examinations verified by colonoscopy, often combined with additional examinations without direct feedback (Table 2).

Just prior to the study, all had passed a test of 25 selected CTC examinations [17] by scoring above a predefined per-polyp sensitivity threshold of 90%. In 12 of these 25 patients, 19 polyps ≥ 6 mm (one flat lesion) and 10 polyps larger than 10 mm could be detected.

### CTC image analysis

The observers were blinded to the CAD results during the initial reading. All patients were evaluated with a primary three-dimensional (3D) method (Endo 3D Unfolded, ViewForum, Philips Medical Systems, Best, The Netherlands). This validated method [18] was used to increase surface visibility and reduce reading time.

Additional two-dimensional (2D) displays with instant on-screen correlation were used for problem solving. Stool subtraction software was not used. The observers digitally recorded size (mm), morphology (pedunculated, sessile, flat) and colon segment (cecum, ascending colon, transverse colon, descending colon, sigmoid colon or rectum).

After their unassisted reading they were able to access the CAD results. Readers were permitted to discard unassisted findings after CAD application. The incorporated commercially available CAD algorithm (ColonCAD, Philips Medical Systems, Best, The Netherlands) had a fixed sensitivity threshold that was not changed during the

**Table 1** The table displays the baseline patient characteristics and procedural details of CTC ( $n=170$ )

Male/female	98/72
Age in years: mean $\pm$ SD	57 $\pm$ 12
Symptomatic (according to referring physician)	45
Abdominal pain	12
Weight loss	2
Changed bowel habits	25
Rectal bleeding	16
Hospital patients:	
Hospital 1	137
Hospital 2	33
History of colonoscopy or sigmoidoscopy 10 years prior to CTC	
Yes	107
No	53
Unknown	10
Patients:	
Without polyps $\geq 6$ mm	120
With polyps $\geq 6$ mm	50
With a polyp $\geq 10$ mm	25
Bowel preparation	
PEG	169
Sodium phosphate	1
CT system used	
Four-slice <sup>a</sup>	7
64-slice <sup>b</sup>	163
Spasmolytics	
20 mg butyl scopolamine bromide <sup>c</sup>	136
1 mg glucagon <sup>d</sup>	28
No spasmolytics	6
Abdominal circumference	
<103 cm	82
$\geq 103$ cm	88
Mean volume of insufflated CO <sub>2</sub> <sup>e</sup>	4.5 l (range: 2.2–7.6)
Mean scanner room examination time	21 min (range 12–35)
Number of complications	0

<sup>a</sup>Mx 8000, Philips Medical Systems, Best, The Netherlands

<sup>b</sup>Brilliance 64, Philips Medical Systems, Best, The Netherlands

<sup>c</sup>Boehringer Ingelheim, Ingelheim, Germany

<sup>d</sup>Glucagon; Novo Nordisk A/S, Bagsvaerd, Denmark

<sup>e</sup>ProtoCO2l, E-Z-EM, Lake Success, N.Y., USA

study. The CAD algorithm was trained on annotated polyp data from 13 patients from a comparative study of 249 patients [19]. These datasets had been verified by colonoscopy and contained a total of 80 polyps  $\geq 5$  mm. In this study, by mouse-clicking a listed candidate, corresponding 3D, 2D axial and 2D MPR views were shown with a mark on the polyp candidate (Fig. 1).

**Table 2** Table lists the number of read CTC cases per observer and observer experience with and without colonoscopic verification

Observer	No. of read cases	No. of endoscopic examinations with colonoscopic verification	No. of endoscopic examinations without colonoscopic verification <sup>b</sup>
1	31/170	500 CTC	225 CTC
2	36/170	300 CTC and 300 MRC	225 CTC
3	29/170	300 CTC	75 CTC
4	36/170	230 CTC and 30 MRC <sup>a</sup>	100 CTC and 240 MRC
5	51/170	100 CTC	25

<sup>a</sup>Including matching polyps in 200 CTC studies of patients of increased risk

<sup>b</sup>Including 25 test patients

If the observers identified CAD lesions that were not detected in the unassisted evaluation, these could be added to the initial list of findings.

#### Interpretation time and image quality

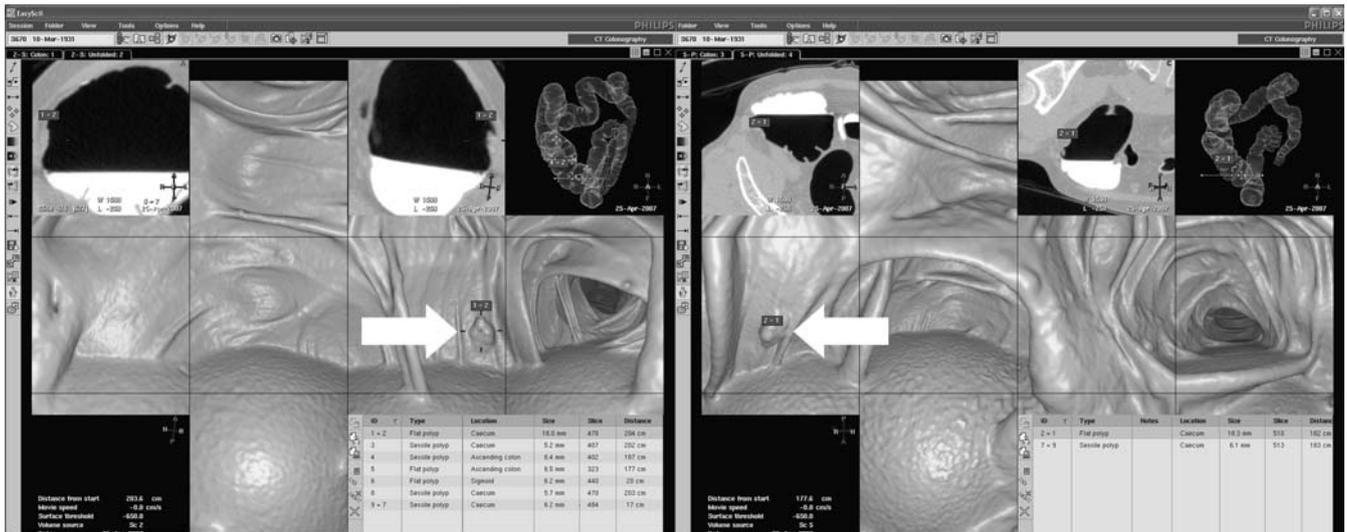
Interpretation times for the unassisted read and for the evaluation of CAD candidates were recorded with a stopwatch for both the prone and supine positions.

When the reading was completed, the observer assessed the degree of colonic distension and quality of the fecal tagging on a four-point Likert-scale (good, sufficient, moderate, poor). The overall quality of the examination was assessed as “diagnostic” or “non-diagnostic”. If the quality was assessed as “non-diagnostic” by the examining physician, the patient was excluded.

#### Colonoscopy

All patients underwent colonoscopy within 3 h after CTC. A gastroenterologist (>200 colonoscopies), or fellow or nurse under direct supervision of a staff member performed the colonoscopy with a standard colonoscope (CF-140L; Olympus, Tokyo, Japan). Chromoendoscopy or narrow-band imaging to improve flat polyp detection was not performed. Patients received on request midazolam (Dormicum, Roche, Basel, Switzerland) and fentanyl (Hameln Pharmaceuticals, Hameln, Germany) or propofol (Fresenius Kabi, Uppsala, Sweden) and fentanyl. The examination was digitally recorded.

Polyp characteristics (size, morphology and segmental location) were documented on a case record form by an attending research nurse. Polyp size was measured with open biopsy forceps (8 mm). The determination of the



**Fig. 1** Screenshot of both monitors of the workstation displaying the patient in the supine (*left*) and prone positions (*right*). The *white arrows* mark a 16 mm sessile polyp that was detected by both CAD and the observer

morphology of polyps was done by the gastroenterologist based on the endoscopic classification of superficial neoplastic lesions [20]. In this classification, flat polyps were defined as lesions with a maximum height of 2.5 mm (closed cups of biopsy forceps). Segmental unblinding was performed for CT lesions 6 mm or larger. Histology was obtained at colonoscopy, except in those cases in which polyp removal was technically impossible or when material was lost during the procedure.

#### Determination of lesion status

Observers were instructed that only hyperplastic, adenomatous (advanced and not-advanced) and potentially malignant lesions were considered true-positive lesions. This qualification was based on the histology report or—if histology was not acquired—based on the endoscopic report.

For CTC, a polyp was considered true-positive, if: (1) its appearance resembled the corresponding polyp at colonoscopy, (2) its segment or adjacent segment corresponded with the reference standard segment and (3) the polyp size as estimated by the endoscopist corresponded with size as measured on CTC, considering a margin of error of 50%. Since the colonoscopy measurement is subject to inaccuracy [21, 22] this criterion could be overruled by the first two criteria.

Polyps  $\geq 6$  mm at colonoscopy that were not identified by the observer without or with CAD, were re-evaluated with knowledge of the colonoscopic findings by a research fellow with experience of more than 300 CTC examinations verified by colonoscopy. In this re-evaluation the nature of all detection errors  $\geq 6$  mm (false-negative

findings) was assessed and differentiated between perception errors (visible in retrospect) and non-perception errors (lesions not visible in retrospect).

Lesions not confirmed by colonoscopy  $\geq 6$  mm (false positives) were assessed by consensus by two experienced research fellows (300 colonoscopy verified CTC). The consensus panel determined whether the finding was related to bowel preparation.

#### Power calculation

Based on a prior feasibility study [16], we expected a 15% increase in sensitivity. In order to determine a statistically significant increase of 15% for polyps  $\geq 6$  mm, at least 39 lesions were required. For this approach, a McNemar test with continuity correction and a  $p$  value of 0.05 to indicate statistical significance was used. Based on prior studies in this patient population, we assumed that the prevalence of patients with polyps  $\geq 6$  mm would be 25% [23]. We therefore required a minimal number of  $39/0.25 = 156$  patients. The total number of patients determined was 170.

#### Outcome parameters per patient

Sensitivity, specificity, positive- and negative-predictive values of CTC without and with CAD were calculated. Sensitivity and number of false-positive findings were calculated for CAD without interaction of the observers (stand alone). Furthermore, sensitivity was calculated for unblinded colonoscopy. The outcome parameters were determined for polyps  $\geq 6$  mm and  $\geq 10$  mm.

A patient was considered true-positive if CTC detected at least one polyp seen at colonoscopy, based on the matching criteria described previously. A patient was categorized as false negative if CTC detected no polyps (although present at the reference standard) or only those of a lower size category in comparison to the reference standard.

We used the McNemar test to compare per-patient sensitivity and specificity values between CTC without and with CAD.

#### Outcome parameters per polyp

We calculated the per-polyp sensitivity for CTC without and with CAD, CAD (stand alone) and blinded colonoscopy for lesions 6–9 mm and  $\geq 10$  mm. In this study, more than one polyp was detected in some patients. Therefore, generalized estimating equations (GEE) (SPSS, 15.0, Statistics, Chicago, USA) was used to revise the data clustering and dependency. In the GEE, the adjusted confidence intervals with regard to per-polyp sensitivity for CTC and blinded colonoscopy (i.e., before unblinding of CTC results) were assessed for CTC without and with CAD. In this same GEE method, regression analyses were done to compare the sensitivity values.

#### Outcome time parameters

The median interpretation time of the CTC reading without CAD and the median time to evaluate all CAD results were calculated.

#### Prevalence of flat polyps stratified for endoscopic colon examination

Because a relatively high number of flat polyps were detected in this population we retrospectively determined whether a colon examination 10 years or less prior to the CTC in the patient's history could effect the prevalence of these polyps. The rationale for this retrospective study was an article published by MacCarty et al. [24] that suggested a higher prevalence of polyps in patients who had undergone a previous endoscopic colon examination. We did not specify the type of colon examination in colonoscopy, sigmoidoscopy or proctoscopy because it was not always clear which part of the colon was examined. The arbitrary period of 10 years was chosen since we assumed this would be the period from a polyp to grow into a tumor and a colon examination executed earlier may have effect the prevalence of flat lesions at CTC.

The prevalence of polyps in the group that had undergone a colon examination and the group that had not were compared with the McNemar test and stratified for size.

## Results

Of 448 eligible patients that were scheduled to undergo optical colonoscopy during the inclusion period, 170 "diagnostic" examinations were included in this study (Fig. 2). The baseline characteristics and procedural details are listed in Table 1.

The degree of bowel distention was assessed as "good" or "average" in 161 patients (95%), "moderate" in eight patients (5%) and "poor" in one patient (1%).

Fecal tagging was assessed as "good" or "average" in 144 patients (85%), "moderate" in 24 (14%) and "poor" in two patients (1%).

#### Reference standard

Unblinded colonoscopy revealed that 50 out of 170 patients (29%) harbored one or more polyps  $\geq 6$  mm and 25 of 170 patients (15%) one or more polyps  $\geq 10$  mm. Table 3 displays the histological and morphological characteristics. One colorectal carcinoma (50 mm) was found.

#### Per-patient analysis

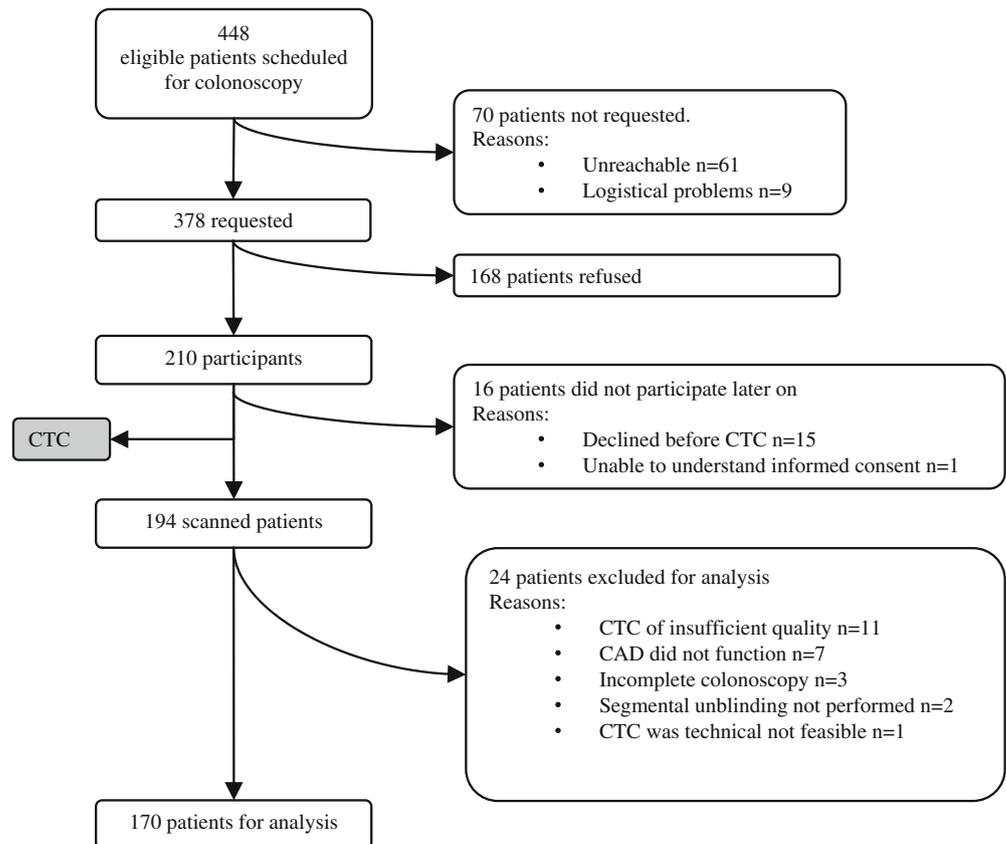
The per-patient sensitivity and specificity is displayed in Table 4. CAD did not significantly alter per-patient sensitivity and specificity for lesions  $\geq 6$  mm and  $\geq 10$  mm. Assisted by CAD, the observers detected one additional patient with a lesion  $\geq 6$  mm and two additional patients with a lesions  $\geq 10$  mm, resulting in a sensitivity of 82% ( $p=1.0$ ) and 72% ( $p=0.5$ ), respectively.

Two patients were erroneously classified as having a lesion  $\geq 6$  mm after accepting a CAD hit and no patients without a lesion  $\geq 10$  mm were wrongly added to the list.

CAD on a stand-alone basis detected 74% (37/50) and 64% (16/25) of the patients with lesions  $\geq 6$  mm and  $\geq 10$  mm. There was no statistically significant difference between the observers and CAD in the respective size categories ( $p=0.375$  and  $p=1.0$ ). CAD had a median number of nine hits per-patient (25–75% quartiles: 5–15). Blinded colonoscopy detected 96% (48/50) and 100% (25/25) of the patients in the respective size categories. As displayed in Table 4, both the positive- and negative-predictive values of the observers with and without CAD were nearly unchanged.

#### Per-polyp analysis

Per-polyp sensitivity for polyps of 6–9 mm and  $\geq 10$  mm are displayed in Table 4. CAD detected one lesion 6–9 mm and three polyps  $\geq 10$  mm initially missed by the observer, but it did not significantly increase sensitivity of the

**Fig. 2** The flowchart of this study

observer for the respective size categories ( $p=0.31$  and  $p=0.08$ ). No true-positive CAD hits were erroneously dismissed by observers. Per-polyp sensitivity was better for polyps 6–9 mm than for polyps  $\geq 10$  mm, without CAD as well as with CAD. To a large extent this can be explained by the difficulty of the observers and CAD in

**Table 3** Table displays the histology and morphology of polyps at seen and removed during colonoscopy

	6–9 mm ( $n=58$ )	$\geq 10$ mm ( $n=30$ )
Non-adenomatous lesions	24	4
Adenomatous lesions classified as “advanced”	4	20
Adenomatous lesions classified as “not advanced”	23	0
Unknown histology	7	4
CRC	0	1
Sessile	40	7
Pedunculated	5	9
Flat	13	13
Tumor	0	1

detecting the relative prevalent number of undetected flat lesions  $\geq 10$  mm; 23% (3/13) of the flat lesions  $\geq 10$  mm were detected without CAD and 31% (4/13) with CAD (Fig. 3). Figure 4 shows that the largest part (6/10) of the missed flat large lesions (either without or with CAD) were not visible in retrospect (non-perception errors). Since these polyps cannot be detected, it is difficult to assess exactly why these non-perception errors were missed.

The three polypoid perception errors 6–9 mm (Fig. 4) were missed because they were situated on a fold ( $n=1$ ), were clearly smaller than 6 mm when measured on CT ( $n=1$ ) or could be defined as flat on CT ( $n=1$ ). The two polypoid perception errors  $\geq 10$  mm were missed because they were situated on a fold ( $n=1$ ) or because of unclear reasons ( $n=1$ ).

Of the 53 false positive lesions detected by the reader without CAD, 23 (43%) findings were according to consensus related to bowel preparation. None of the four false-positive findings suggested by CAD and incorporated in the final list by the observer were related to bowel preparation.

Although CAD did not significantly increase sensitivity, it did not significantly alter specificity either: three extra false-positive lesions 6–9 mm and one extra lesion  $\geq 10$  mm

**Table 4** Results on a per-patient and per-poly basis (%) (95% CI in parentheses)

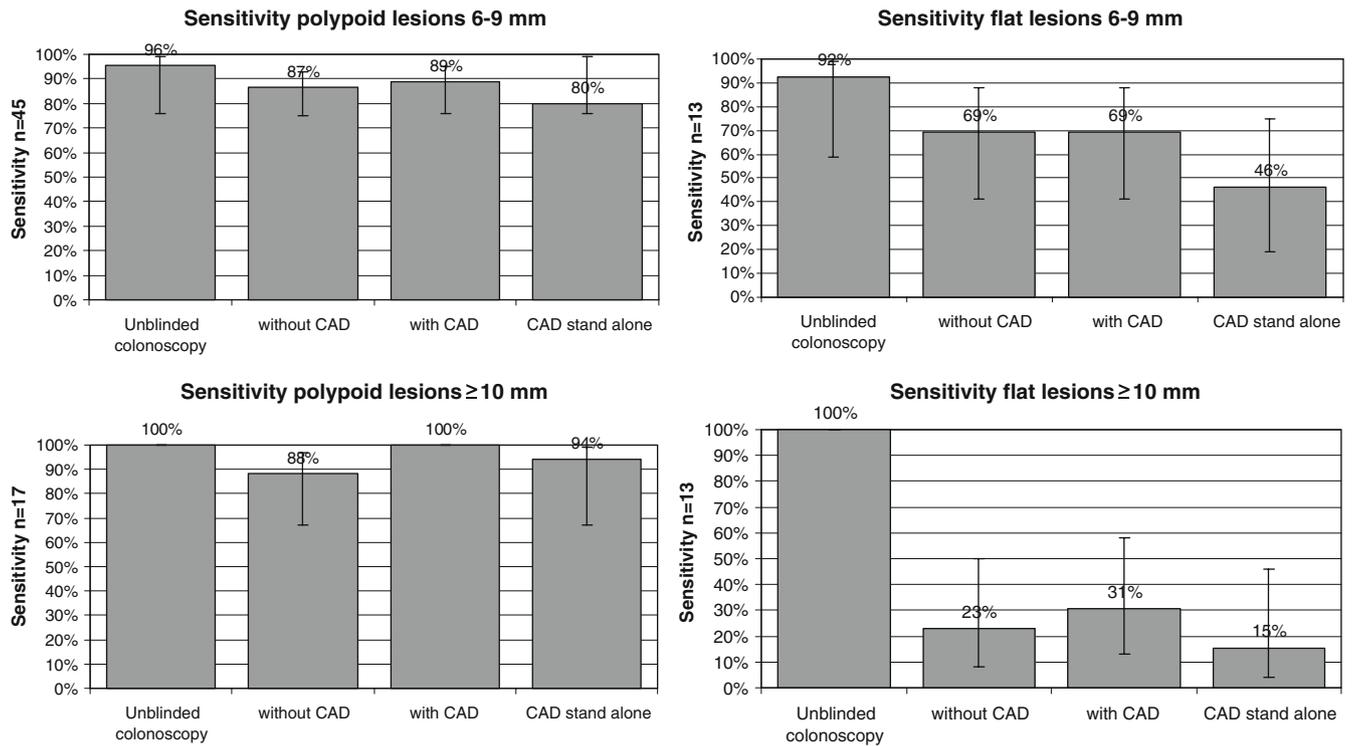
			CTC without CAD	CTC with CAD
Per patient	Sensitivity	≥6 mm	80% (69–91) 40/50	82% (71–93) 41/50
		≥10 mm	64% (45–83) 16/25	72% (54–90) 18/25
	Specificity	≥6 mm	84% (78–91) 101/120	83% (76–89) 99/120
		≥10 mm	94% (90–98) 136/145	94% (90–98) 136/145
	Positive-predictive value	≥6 mm	68% (56–80) 40/59	66% (54–78) 41/62
		≥10 mm	64% (45–83) 16/25	67% (48–84) 18/27
Negative-predictive value	≥6 mm	91% (86–96) 101/111	92% (86–97) 99/108	
	≥10 mm	94% (90–98) 136/145	95% (92–98) 136/143	
Per poly	Sensitivity	6–9 mm	83% (72–90) 48/58	84% (73–92) 49/58
		≥10 mm	60% (43–75) 18/30	70% (51–84) 21/30
	Total number of false-positive lesions	6–9 mm	42	45
		≥10 mm	11	12

in 170 patients were added to the list of the observer (Table 4).

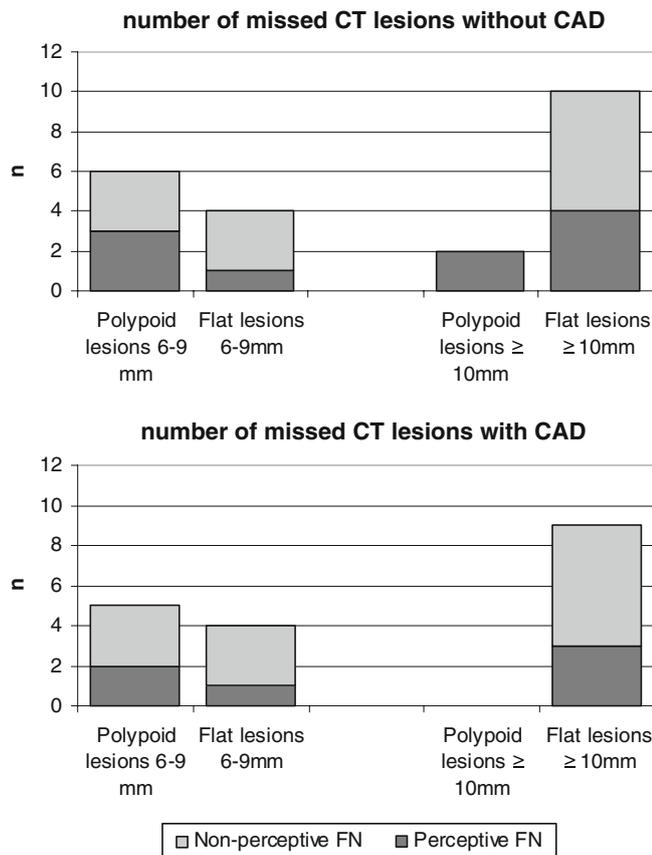
CAD on a stand-alone basis detected 72% (42/58) of the polyps 6–9 mm and 60% (18/30) of the polyps ≥10 mm. Blinded colonoscopy detected 95% (55/58) of the polyps 6–9 mm and 100% (30/30) of the polyps ≥10 mm.

Interpretation time

The observers had an a median interpretation time of 16 min 00 s (25–75% quartiles: 11 min 35 s–23 min 6 s) to complete the examination and a median time of 1 min 26 s (25–75% quartiles: 28 s–2 min 46 s) to evaluate all CAD results after the initial reading.



**Fig. 3** The left histograms show the sensitivity of polypoid lesions of 6–9 mm and ≥10 mm, the right histograms show the sensitivity of lesions with a flat morphology. The difference in sensitivity between polypoid and flat lesions is more striking at lesions ≥10 mm



**Fig. 4** The number of false-negative findings and distribution of perceptive and non-perceptive errors among flat and non-flat lesions

#### Prevalence of flat polyps stratified for endoscopic colon examination

Sixty-three percent (107/170) of the patients had undergone an endoscopic colon examination prior to CTC, 31% (53/170) had not. For ten patients, the history could not be retrieved. Of the polyps 6–9 mm, in patients with a history of endoscopy 27% (12/44) were flat, in contrast to 8% (1/12) of the polyps in patients without a history of endoscopy ( $p=0.259$ ). Of the polyps  $\geq 10$  mm, in patients with a history of endoscopy 60% (9/15) were flat, in contrast to 33% (4/12) of polyps in patients without a history of endoscopy ( $p=0.168$ ). Thus, the prevalence of flat polyps in both size categories was higher in the group that had undergone colon examination, though statistical significance was not reached.

#### Discussion

Although CAD in a second-read paradigm detected one additional patient with a lesion  $\geq 6$  mm and two patients with a lesion  $\geq 10$  mm, it did not significantly improve

per-patient sensitivity in this increased risk patient population.

Several CTC studies in which the additional value of CAD was evaluated (after the interaction with the observer) have reported good results in terms of polyp detection [12, 25–27]. All concluded that the observers detected statistically more polyps with CAD.

In contrast to these studies, we did not find a significant additional value for CAD. The study design of these studies differs from our study in a number of aspects: patient selection, inclusion and exclusion criteria and reference standard. However, we think that the most important difference between the aforementioned studies and our study is the fact that the observers in our study were more experienced, i.e., more than 100 CTC cases verified by colonoscopy. Since there is good evidence that experience in CTC results in less false-negative findings [28, 29], it is logical that it is more difficult to substantially increase the sensitivity of the observer with CAD. In studies that report data about the additional value of CAD on experienced observers [30, 31], experienced observers benefited proportionately less from CAD when compared with the inexperienced readers. This finding is supported by the results of this study.

Although a CAD algorithm has the potential to decrease the number of perceptual errors by exposing the observer to candidate lesions, it cannot account for interpretative errors. In the above-mentioned papers, a significant increase of false positives have been reported. Though the specificity in this study was not significantly increased, there were only two patients erroneously classified as having a lesion. Both false-positive lesions measured 6–9 mm, none was larger than 10 mm.

Even though the sensitivity of the observers was low (i.e., 72% for lesions  $\geq 10$  mm), CAD was not able to increase their performance. In our opinion, the reported sensitivity requires looking for causes in the population itself. All nine polyps  $\geq 10$  mm missed by the observer with CAD had a flat morphology (Fig. 4). These flat polyps are an important cause of false-negative findings [32, 33]. In this population, 13 of the 30 lesions  $\geq 10$  mm were flat and therefore an important explanation of the moderate sensitivity, not only for the observer but for CAD as well.

The unexpectedly high number of flat lesions may be related to the history of patients; MacCarty et al. [34] reported in a prospective study of 75 consecutive patients that more than 50% of the false negatives missed by experienced readers were not even visible in retrospect in a population that had been screened by colonoscopy 5 years prior to CTC. Nearly all these polyps were flat. In this study population, the prevalence of flat lesions was higher (although not statistically significant) in the group of patients that had undergone a colonoscopy or sigmoidoscopy prior to CTC as well (Table 1).

We concur with MacCarty and coworkers that previous screening could adversely affect CTC sensitivity in two

ways: first, it is likely that many easy-to-see polyps would be detected and removed at the initial screening, and fewer hard-to-see polyps would be detected and removed; second, endoscopic polypectomy may have been incomplete. Remnants of polyps are flatter than the original intact lesions, and would, therefore, be more difficult to detect. So we think that the selection of patients has an important influence on the test characteristics.

The type of bowel preparation, i.e., extensive or reduced, with or without oral contrast (iodine and/or barium) may influence the performance in terms of polyp detection and number of false-positive findings of CAD; polyps can be covered by fecal material or fecal remains may simulate polyps. In this study, we used PEG as an extensive bowel preparation for colonoscopy, combined with Telebrix that has a laxative effect as well. Although we have not evaluated the nature of all CAD candidates, the additional value of CAD did not seem to be impaired by this type of bowel preparation used in this study, since none of the false-positive lesions incorporated in the final list of the observers were prep-related and only two of 22 false-negative findings were covered by fecal material (although we think this is not the reason why they were missed).

This study has limitations; due to the small time-frame between the CTC and colonoscopy, the patients could be read by only one observer out of a group of five different observers. Each observer had a different level of experience. Therefore, although no statistically significant difference in sensitivity was measured between the observers, and none of the observers had a significant improvement in performance after CAD (data not shown), the best-performing observers could have leveled out the sensitivity of the least performing observers. Still, the situation as described in this paper is similar to the practical setting of many hospitals; each examination will not be read by five different radiologists but by only one of a pool of experts.

Secondly, the level of experience of the group of five readers was relatively high. It is likely that the additional value of CAD would be larger if the data were read by a relatively inexperienced reader group [35, 36]. Therefore, our conclusion may not apply to relatively untrained readers.

Thirdly, we evaluated CAD using a primary 3D reading paradigm. Although the discussion as to whether to read the data in 2D or 3D is still not settled, a (slight) superiority in terms of polyp detection with 3D is reported by some studies [37, 38]. Since the reported sensitivities of a primary 2D paradigm tend to be lower, CAD may have a larger additional value when used in a 2D reading protocol.

Fourthly, the relatively large number of flat polyps may limit the generalization of the results of this study. However, if we leave out all 26 flat polyps, and we only consider the remaining 62 polypoid lesions, we still cannot demonstrate a significant contribution of CAD to the sensitivity of the observer. This number was still more than the 39 polyps that were needed to demonstrate a 15% sensitivity difference according to our power analysis.

In conclusion, although CTC with CAD in a second read paradigm detected a few more lesions than CTC without CAD, CAD has no statistically significant positive influence on CTC performance in an increased-risk population when used by a relative experienced group of observers.

**Acknowledgements** We thank Prof. M.J. van de Vijver, MD, PhD for assessing the histology of the polyps. We thank Mrs. A.C. Dobben, MSc, PhD for her statistical support. This work was supported by a grant from Philips Medical Systems (Best, The Netherlands).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Levin B, Lieberman DA, McFarland B et al (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology* 134:1570–1595
2. Summers RM, Yao J, Pickhardt PJ et al (2005) Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology* 129:1832–1844
3. Yoshida H, Nappi J, MacEneaney P, Rubin DT, Dachman AH (2002) Computer-aided diagnosis scheme for detection of polyps at CT colonography. *Radiographics* 22:963–979
4. Jerebko AK, Summers RM, Malley JD, Franaszek M, Johnson CD (2003) Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees. *Med Phys* 30:52–60
5. Mani A, Napel S, Paik DS et al (2004) Computed tomography colonography: feasibility of computer-aided polyp detection in a “first reader” paradigm. *J Comput Assist Tomogr* 28:318–326
6. Summers RM, Jerebko AK, Franaszek M, Malley JD, Johnson CD (2001) Complementary role of computer-aided diagnosis for detecting colonic polyps with CT colonography. *Radiology* 120:391–399
7. Taylor SA, Halligan S, Burling D et al (2006) Computer-assisted reader software versus expert reviewers for polyp detection on CT colonography. *AJR Am J Roentgenol* 186:696–702
8. Halligan S, Altman DG, Mallett S et al (2006) Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 131:1690–1699
9. Taylor SA, Charman SC, Lefere P et al (2008) CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. *Radiology* 246:463–471
10. Halligan S, Altman DG, Mallett S et al (2006) Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 131:1690–1699

11. Baker ME, Bogoni L, Obuchowski NA et al (2007) Computer-aided detection of colorectal polyps: can it improve sensitivity of less-experienced readers? Preliminary findings. *Radiology* 245:140–149
12. Nicholas Petrick P, Maruf Haider M, Ronald M, Summers MP et al (2008) CT colonography with computer-aided detection as a second reader: observer performance study. *Radiology* 246:148–156
13. Mang T, Peloschek P, Plank C et al (2007) Effect of computer-aided detection as a second reader in multidetector-row CT colonography. *Eur Radiol* 17:2598–2607
14. Mang T, Peloschek P, Plank C et al (2007) Effect of computer-aided detection as a second reader in multidetector-row CT colonography. *Eur Radiol* 17:2598–2607
15. Taylor SA, Charman SC, Lefere P et al (2008) CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. *Radiology* 246:463–471
16. de Vries AH, Peters J, Grigorescu S, Truyen R, Gerritsen FA, Stoker J (2005) The performance of CAD versus the performance of an expert in CT colonography. *Proc Sixth Int Symp Virtual Colonoscopy*, Boston, p 189
17. Pickhardt PJ, Choi JR, Hwang I et al (2003) Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 349:2191–2200
18. Vos FM, van Gelder RE, Serlie IW et al (2003) Three-dimensional display modes for CT colonography: conventional 3D virtual colonoscopy versus unfolded cube projection. *Radiology* 228:878–885
19. van Gelder RE, Nio CY, Florie J et al (2004) Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer. *Gastroenterology* 127:41–48
20. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002 (2003). *Gastrointest Endosc* 58:S3–43
21. Fennerty MB, Davidson J, Emerson SS, Sampliner RE, Hixson LJ, Garewal HS (1993) Are endoscopic measurements of colonic polyps reliable? *Am J Gastroenterol* 88:496–500
22. Gopalswamy N, Shenoy VN, Choudhry U et al (1997) Is in vivo measurement of size of polyps during colonoscopy accurate? *Gastrointest Endosc* 46:497–502
23. Jensch S, de Vries AH, Peringa J et al (2008) CT colonography with limited bowel preparation: performance characteristics in an increased-risk population. *Radiology* 247:122–132
24. MacCarty RL, Johnson CD, Fletcher JG, Wilson LA (2006) Occult colorectal polyps on CT colonography: implications for surveillance. *AJR Am J Roentgenol* 186:1380–1383
25. Baker ME, Bogoni L, Obuchowski NA et al (2007) Computer-aided detection of colorectal polyps: can it improve sensitivity of less-experienced readers? Preliminary findings. *Radiology* 245:140–149
26. Halligan S, Altman DG, Mallett S et al (2006) Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 131:1690–1699
27. Mang T, Peloschek P, Plank C et al (2007) Effect of computer-aided detection as a second reader in multidetector-row CT colonography. *Eur Radiol* 17:2598–2607
28. Burling D, Halligan S, Altman DG et al (2006) CT colonography interpretation times: effect of reader experience, fatigue, and scan findings in a multicentre setting. *Eur Radiol* 16:1745–1749
29. Johnson CD, Toledano AY, Herman BA et al (2003) Computerized tomographic colonography: Performance evaluation in a retrospective multicenter setting. *Gastroenterology* 125:688–695
30. Mang T, Peloschek P, Plank C et al (2007) Effect of computer-aided detection as a second reader in multidetector-row CT colonography. *Eur Radiol* 17:2598–2607
31. Taylor SA, Charman SC, Lefere P et al (2008) CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. *Radiology* 246:463–471
32. Fidler JL, Fletcher JG, Johnson CD et al (2004) Understanding interpretive errors in radiologists learning computed tomography colonography. *Acad Radiol* 11:750–756
33. Gluecker TM, Fletcher JG, Welch TJ et al (2004) Characterization of lesions missed on interpretation of CT colonography using a 2D search method. *AJR Am J Roentgenol* 182:881–889
34. MacCarty RL, Johnson CD, Fletcher JG, Wilson LA (2006) Occult colorectal polyps on CT colonography: implications for surveillance. *AJR Am J Roentgenol* 186:1380–1383
35. Mang T, Peloschek P, Plank C et al (2007) Effect of computer-aided detection as a second reader in multidetector-row CT colonography. *Eur Radiol* 17:2598–2607
36. Taylor SA, Charman SC, Lefere P et al (2008) CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. *Radiology* 246:463–471
37. van Gelder RE, Florie J, Nio CY et al (2007) A comparison of primary two- and three-dimensional methods to review CT colonography. *Eur Radiol* 17:1181–1192
38. Pickhardt PJ, Lee AD, Taylor AJ et al (2007) Primary 2D versus primary 3D polyp detection at screening CT colonography. *AJR Am J Roentgenol* 189:1451–1456