

Ayso H. de Vries
Marjolein H. Liedenbaum
Shandra Bipat
Roel Truyen
Iwo W. O. Serlie
Rutger H. Cohen
Saskia G. C. van Elderen
Anneke Heutinck
Oskar Kesselring
Wouter de Mony 
Lambertus te Strake
Tjeerd Wiersma
Jaap Stoker

Primary uncleaned 2D versus primary electronically cleansed 3D in limited bowel preparation CT-colonography. Is there a difference for novices and experienced readers?

Received: 27 August 2008
Revised: 1 January 2009
Accepted: 26 January 2009
Published online: 20 March 2009
  The Author(s) 2009. This article is published with open access at Springerlink.com

A. H. de Vries ( ) ·
M. H. Liedenbaum · S. Bipat ·
A. Heutinck · J. Stoker
Department of Radiology Academic
Medical Centre,
University of Amsterdam,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands
e-mail: Ayso.devries@amc.uva.nl
e-mail: aysodevries@hotmail.com
Tel.: +31-020-56663028
Fax: +31-020-56669119

R. Truyen · I. W. O. Serlie
Department of Healthcare Informatics,
Philips Medical Systems 3,
Veenpluis 6,
5684 PC Best, The Netherlands

R. H. Cohen
Department of Radiology,
Ziekenhuis Amstelland,
Laan van de Helende Meesters 8,
1186 AM
Amstelveen, The Netherlands

S. G. C. van Elderen
Department of Radiology,
Leids Universitair Medisch Centrum,
Albinusdreef 2,
2333 ZA Leiden, The Netherlands

O. Kesselring · T. Wiersma
Department of Radiology,
Ziekenhuis Rijnstate,
Wagnerlaan 55,
6815 AD Arnhem, The Netherlands

W. de Mony 
Department of Radiology,
Kennemer Gasthuis,
Boerhaavelaan 22,
2035 RC Haarlem, The Netherlands

L. te Strake
Department of Radiology,
Middlemore Hospital,
Private Bag 93311,
Otahuhu, Auckland, 6, New Zealand

Abstract The purpose of this study was to compare a primary uncleaned 2D and a primary electronically cleansed 3D reading strategy in CTC in limited prepped patients. Seventy-two patients received a low-fibre diet with oral iodine before CT-colonography. Six novices and two experienced observers reviewed both cleansed and uncleaned examinations in randomized order. Mean per-polyp sensitivity was compared between the methods by using generalized estimating equations. Mean per-patient sensitivity, and

specificity were compared using the McNemar test. Results were stratified for experience (experienced observers versus novice observers). Mean per-polyp sensitivity for polyps 6 mm or larger was significantly higher for novices using cleansed 3D (65%; 95% CI 57–73%) compared with uncleaned 2D (51%; 95% CI 44–59%). For experienced observers there was no significant difference. Mean per-patient sensitivity for polyps 6 mm or larger was significantly higher for novices as well: respectively 75% (95% CI 70–80%) versus 64% (95% CI 59–70%). For experienced observers there was no statistically significant difference. Specificity for both novices and experienced observers was not significantly different. For novices primary electronically cleansed 3D is better for polyp detection than primary uncleaned 2D.

Keywords CT-colonography · Sensitivity and specificity · Colonic polyps · Electronic cleansing · Colorectal neoplasms

Introduction

CT-colonography (CTC) has consistently been shown to have a high accuracy for the detection of colorectal neoplasia, and has recently been included in the official guidelines for colorectal cancer screening [1]. An im-

portant disadvantage of the technique is that many patients find the bowel preparation burdensome [2]. Therefore efforts have been made to prepare patients for CTC with a less extensive bowel preparation [3–7]. Minimizing bowel preparation may increase patient compliance [8–10], but will result in larger amounts of residual faeces in the colon.

A prerequisite is that faecal material is labelled with oral contrast (i.e. faecal tagging) in order to differentiate faecal material from colonic structures.

To our knowledge, all limited prepared CTC studies have been performed using primary two-dimensional (2D) display methods [3–5]. The rationale for this approach is that submerged segments can be better assessed in 2D.

Previous studies in patients with extensive bowel preparation have indicated that primary three-dimensional (3D) reading may result in less false negative findings compared with primary 2D reading [11, 12]. If a similar empty endoluminal view could be achieved by electronic removal of tagged material (“electronic cleansing”) in patients who have undergone limited bowel preparation, primary 3D could be a method of choice.

However, specific artefacts of electronic cleansing were described that potentially reduced the accuracy of CTC [13]. This may be the reason for the paucity of papers on the use of electronic cleansing. A specifically noticeable problem is posed by the distracting ‘ridges’ or ‘pseudopolyps’ emanating from locations where air, soft tissue and tagged material meet.

We hypothesized that especially for inexperienced observers a primary electronically cleansed 3D (PEC3D) method may have advantages for evaluation of the colon: polyps are visible for longer than in a 2D method and PEC3D provides a more intuitive reproduction of reality. In this study we used a cleansing algorithm that was devised to improve 3D image quality at the junctions of air, soft tissue and tagged material.

Therefore the purpose of this study was to assess whether there was any difference in accuracy between two different reading strategies for the detection of colorectal polyps in a patient population that had undergone a reduced bowel preparation. The results of primary uncleaned 2D (PU2D) and primary electronically cleansed 3D (PEC3D) were stratified for reader experience.

Materials and methods

Study population

The institutional review board of our hospital approved the study. All patients gave written informed consent.

The CT datasets used in this study were a consecutive series of FOBT (faecal occult blood test) positive patients that were included in the framework of a previous comparative study of two different faecal occult blood tests [14].

Bowel preparation

Bowel preparation started 2 days before CTC and consisted of seven 50-ml aliquots of meglumine ioxithalamate

(Telebrix Gastro 300 mg I/ml; Guerbet, Cedex, France) administered orally (undiluted) with each meal (breakfast, lunch and dinner). The use of oral contrast was combined with a low-fibre diet. The evening and morning before the CTC examination no solid foods were allowed. Explicit instructions about fluid intake were not given. No laxatives were used in order to minimize patient discomfort.

Scan parameters

CTC was executed according to state-of-the-art techniques. Patients were examined in prone and supine position after the intravenous administration of bowel relaxants (Buscopan; Boehringer Ingelheim, Germany or, if contraindicated, Glucagon; Novo-Nordisk, Bagsvaerd, Denmark). CO₂ was automatically insufflated (PROTOCO₂L, EZ-E-M). Patients were not given intravenous contrast medium. Examinations were performed on a 64-slice multidetector CT system (Brilliance, Philips Medical Systems, Best, the Netherlands) with a reference mA s value of 40 mA s (z-axis tube modulation and automatic current selection was used). Collimation was 64×0.625 mm, pitch 1.2, slice thickness 0.9 mm, rotation time 0.4 s and tube voltage 120 kV.

Reading methods

The examinations were read in random order i.e. the PU2D and PEC3D datasets were interspersed. The observers were blinded to the results of the reference standard. To avoid recall bias, we aimed to maximize the interval between the PU2D and PEC3D reading of the same patient. This period varied per observer (mean 33 days, range 7–66) (Table 1). All detected lesions were recorded in a digital database. The method of detection (PU2D or PEC3D), colon segment and size of the lesion (as measured in the primary review method) of all findings were documented.

Primary uncleaned 2D (PU2D)

The PU2D interface is illustrated in Fig. 1 (ViewForum 6.1, Philips Medical Systems), using axial CT images (*W* 1,250, *L* 50). Observers were free to adjust the window setting when appropriate. To further elucidate suspected findings on 2D, an uncleaned 3D display and a 3D colon overview could be viewed for problem solving. The dual-screen interface simultaneously displayed both supine and prone scan positions.

Primary electronically cleansed 3D (PEC3D)

The PEC3D interface is illustrated in Fig. 2. The unfolded cube review method (ViewForum 6.1, Philips Medical

Table 1 Individual observer performance in PU2D and PEC3D

| Observer | Number Experience | Number of primary 2D CTCs verified by colonoscopy | Number of primary 3D CTCs verified by colonoscopy | Review times (s) | | Mean interval (days) between both review methods (SD) |
|---------------------------|-------------------|---|---|------------------|------|---|
| | | | | 2D | 3D | |
| 2 | Experienced | 500 | 100 | 574 | 593 | 32 (15) |
| 6 | Experienced | 300 | 50 | 432 | 579 | 66 (38) |
| Mean experienced observer | | | | 503* | 586* | |
| 1 | Novice | 40 | 10 | 765 | 870 | 11 (4) |
| 3 | Novice | 40 | 10 | 756 | 964 | 36 (19) |
| 4 | Novice | 40 | 10 | 404 | 376 | 7 (3) |
| 5 | Novice | 40 | 10 | 259 | 419 | 60 (38) |
| 7 | Novice | 40 | 10 | 634 | 795 | 30 (23) |
| 8 | Novice | 40 | 10 | 403 | 447 | 20 (9) |
| Mean novice observers | | | | 537* | 645* | |

SD standard deviation

*Denotes statistically significant difference between the review methods of both groups

Systems) was developed to maximize the area of visible colon surface and was previously validated [15]. The unfolded cube display method was simultaneously displayed with corresponding original—uncleansed—2D multiplanar reformatted images and a 3D colon overview for problem solving. The dual-screen interface simultaneously displayed both supine and prone examinations.

At the time of study the cleansing algorithm (View-Forum, Philips Medical Systems, Best, the Netherlands) [16] was not yet commercially available. In short, the algorithm assumes that the measured density in a voxel arises due to a combination of three materials: soft tissue, air and tagged material. Initially, the percentage of materials in each voxel is determined. Subsequently, the

partial volume of tagged material is replaced by air and the new density is calculated. Finally, a 3D method visualizes the colon from an endoluminal perspective as if there were no faecal remains.

During the study the algorithm was not yet integrated into the system and was therefore processed on a separate computer (Precision 690, Dell, Round Rock, USA). Afterwards, the cleansed data were reloaded on the workstation.

Observers

Eight physicians participated in this study: two experienced observers (two research physicians working full time on

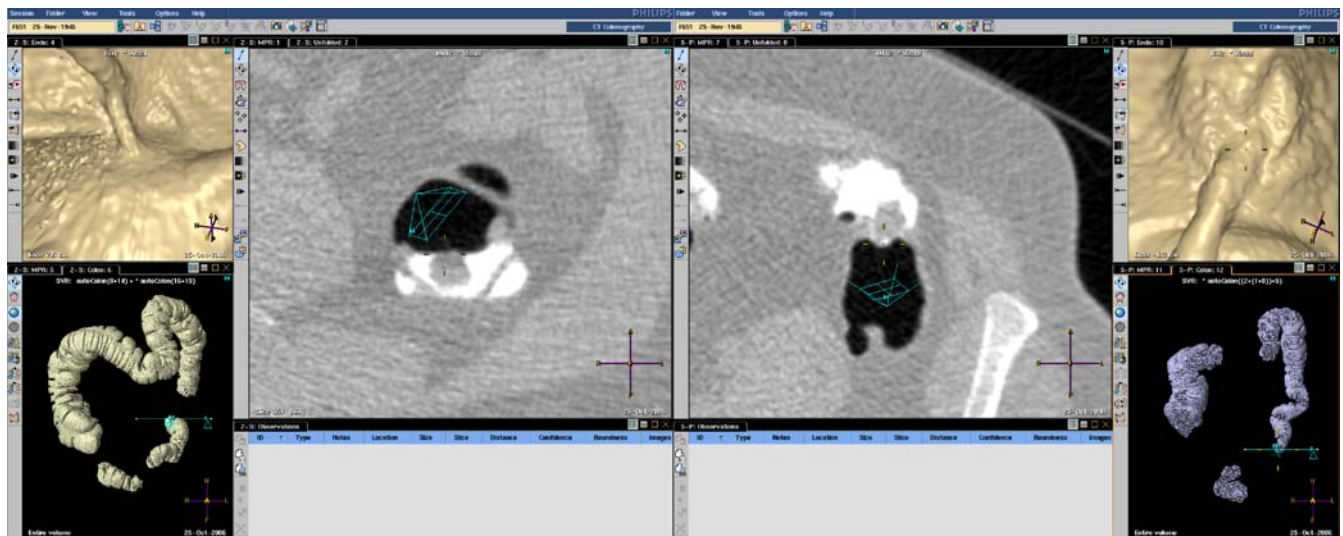


Fig. 1 Interface of the PU2D method. Uncleansed 3D images were used for problem solving. In the corresponding scan positions a large stalked polyp can be seen in the descending colon

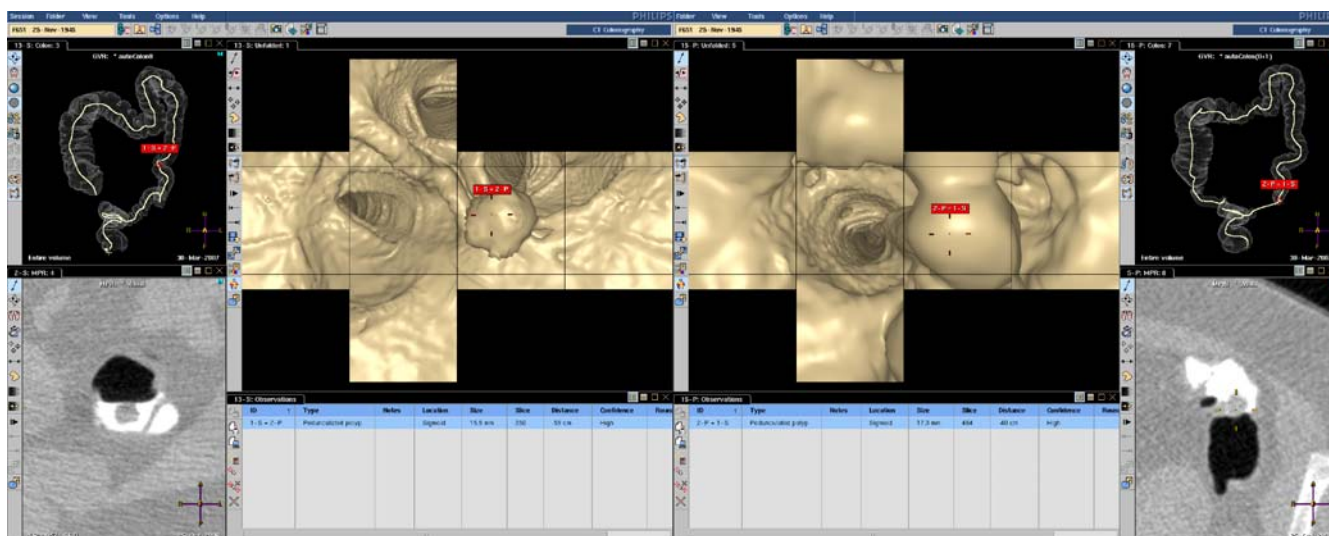


Fig. 2 Interface of the PEC3D method. Uncleansed 2D images were used for problem solving. In the corresponding scan positions the same large stalked polyp can be seen as in Fig. 1. This submerged polyp has become visible in 3D after electronic cleansing

CTC) and six novice observers (five radiologists, one recently qualified physician).

The two experienced observers had seen over 350 CTC examinations verified by colonoscopy (among these the 75 patients included in this study). All novice observers, without any notable prior experience with CTC, had undergone the same learning curve. The learning curve consisted of 50 selected CTC examinations from a publicly available database [17]. Forty examinations were read using a primary 2D method and 10 using a primary 3D method. The ratio for this distribution was the assumption that detecting lesions in 2D is more difficult than in 3D, although the assessment of a suspected lesion is similar in both review methods. Personal feedback on false negative and false positive findings for polyps 6 mm or larger was provided by a radiology research fellow (more than 500 CTC examinations verified by colonoscopy).

The observers were aware of the fact that the actual study was done on FOBT positive patients.

Reference standard

The reference standard was based on the findings of the initial CTC reading, which was double read by two experienced observers (more than 200 CTC) and verified by colonoscopy with segmental unblinding. Colonoscopy was performed within 2 weeks after CTC.

A polyp seen during CTC was considered true positive if (1) its appearance resembled the corresponding adenomatous or nonadenomatous polyp at colonoscopy, (2) its segment or adjacent segment corresponded with the segment of the reference standard and (3) the polyp size as estimated by the endoscopist corresponded with the CTC size, considering a margin of error of 50%. Since the

colonoscopy measurement is subject to inaccuracy [18, 19] this criterion could be overruled by the first two criteria. All other annotations were considered false positives.

The relation of polyps to faecal material

To illustrate the influence of faecal material on the visibility of polyps, a research fellow determined whether each polyp was completely covered by faecal material (i.e. completely submerged in both positions), partially covered by faecal material (i.e. not covered in both scan positions but at least partially covered in at least one position) or not covered by faecal material at all.

Power analysis

A power calculation was performed based on an assumed 15% difference between the methods in sensitivity for polyps 6 mm or larger (i.e. 70% versus 85%). The number of visible polyps required to detect a statistically significant difference by using the McNemar test was 75 ($p=0.05$). As we expected some of the patients to be excluded due to insufficient diagnostic quality, we included a total of 75 examinations of FOBT positive patients that had 84 visible polyps.

Performance

Per polyp To investigate differences between PU2D and PEC3D, we calculated the mean per-polyp sensitivity for both experienced and novice readers. Statistical differences between the review methods were assessed by using

generalized estimating equations (GEE) (SPSS, 15.0, Statistics, Chicago, USA) to revise for data clustering and dependency. In this GEE method, regression analyses were performed to compare the mean sensitivity values of the two methods [20]. Since the per-polyp specificity cannot be calculated because it is a nonexistent entity, we confined the per-polyp results to the number of false positive findings.

Per patient Other main outcome per-patient parameters were per-patient sensitivity and specificity. Statistical differences in mean per-patient sensitivity and mean specificity measures were assessed with the McNemar test.

Both mean per-polyp and per-patient outcome measures were analysed according to cutoff values of 6 and 10 mm. P values <0.05 were considered statistically significant.

Review time

The review time, defined as the time measured with a stopwatch to review a complete examination, was compared for both methods. The review time did not include the time required for processing the images. These procedures are highly dependent on calculation power, are performed semiautomatically and require no reviewer interaction.

Differences in mean review time of experienced and novice observers were assessed with a paired Student's t test. P values <0.05 were considered statistically significant.

Image quality

The image quality of the examinations was rated (after reading) on a four-point Likert scale: diagnostic without artefacts; diagnostic with a small number of artefacts, polyps 6 mm or larger cannot be missed; diagnostic with many artefacts, polyps 6–9 mm can be missed; not diagnostic, polyps 10 mm or larger can be missed.

Firstly, if at least four observers rated the examination as “not diagnostic” the patient was excluded.

Secondly, we determined the percentage of PU2D and PEC3D examinations in the various rating categories. The percentage represented the mean rating of all eight observers.

Thirdly, we assessed per-observer differences in quality between PU2D and PEC3D using ordinal regression analysis with PU2D as reference standard. A lower relative diagnostic odds ratio (RDOR) implies inferior image quality of the PEC3D compared with the PU2D. Confidence intervals not reaching 1 indicate significant inferiority.

Fourthly, each observer determined the presence of artefacts per examination. The impact of each artefact on

the “readability” of the examination was assessed on a four-point Likert scale: not disturbing; disturbing, but cannot hinder detection of polyps 6 mm or larger; disturbing, can cover polyps 6–9 mm; and disturbing, can cover polyps 10 mm or larger. We report the number of patients with artefacts observed by at least four observers and the number of patients with artefacts classified by at least four observers as “disturbing, can cover polyps”. These were analysed according to cutoff values of 6 and 10 mm.

Results

All 75 patients were scanned between October 2006 and January 2007 and underwent colonoscopy within 9.4 days (SD 6.6 days). We excluded three of the 75 patients because more than four observers rated the diagnostic quality of three PEC3D examinations “not diagnostic”. In 2D these examinations were rated by one to three observers as “not diagnostic”. The remaining 72 patients consisted of 38 men and 34 women (mean age 59.5 years, SD 6.4 years, range 50–73). Bleeding during polypectomy was reported for three patients, none of whom required reintervention. No adverse events were reported to occur in any of the CT examinations.

A total of 90 polyps 6 mm or larger were detected: 17 polyps were 20 mm or larger (median size 25 mm, range 20–40 mm), 36 polyps were 10–19 mm (median size 11 mm, range 10–17 mm) and 37 polyps were 6–9 mm (median size 7 mm, range 6–9 mm). The histology revealed 82 adenomatous polyps, 4 nonadenomatous polyps and 4 colorectal carcinomas.

Polyps 6 mm or larger were found in 50 out of 72 patients (69%) and polyps 10 mm or larger were found in 34 out of 72 patients (47%). As shown by Table 2 a substantial part of most polyps was covered by faecal material.

Performance

Per polyp Mean per-polyp sensitivity for novices and experienced observers for both review methods are listed in Table 3. Novice observers had a significantly higher mean sensitivity when using PEC3D for polyps 6 mm or larger (+14%, $p<0.001$), 6–9 mm (+19%, $p<0.001$) and 10 mm or larger (+9%, $p<0.001$). Within the group of novice observers there was a considerable difference in sensitivity for both review methods (Tables 3 and 4), despite the fact that all observers had undergone a similar training protocol.

For experienced observers there was no significant difference in mean sensitivity between both methods for polyps 6 mm or larger ($p=0.755$), 6–9 mm ($p=0.170$) and 10 mm or larger ($p=0.207$).

Table 2 Relation of polyps to faecal material

| | 6–9 mm | | ≥10 mm | |
|---|----------|-----|----------|-----|
| | <i>n</i> | % | <i>n</i> | % |
| Polyps completely covered by faecal material in both scan positions | 1 | 3 | 2 | 4 |
| Polyps partially covered by faecal material in one or both scan positions or completely covered in one position | 13 | 35 | 28 | 53 |
| Polyps not covered by faecal material at all in both scan positions | 18 | 49 | 22 | 42 |
| In retrospect not visible in both scan positions | 5 | 14 | 1 | 2 |
| Total number of polyps | 37 | 100 | 53 | 100 |

Table shows the number (*n*) and percentage of visible polyps either *completely* covered by faecal material, *partially* covered by faecal material or *not* covered by faecal material at all in two scan positions. Polyps that are not visible at all are reported as well

Per patient The per-patient performance characteristics for PU2D and PEC3D are shown in Table 4. Novice observers had a significantly higher mean sensitivity when using PEC3D for polyps 6 mm or larger (+11%, $p < 0.001$) and 10 mm or larger (+6%, $p = 0.033$) compared with PU2D. For experienced observers there was no significant difference in mean sensitivity between both methods for polyps 6 mm or larger ($p = 0.549$) and 10 mm or larger ($p = 0.125$).

Specificity for novice observers when using PEC3D was not significantly lower for polyps 6 mm or larger ($p = 0.057$) and 10 mm or larger ($p = 0.36$) compared with PU2D. For experienced observers there was no significant difference for polyps 6 mm or larger ($p = 0.5$) and 10 mm or larger ($p = 1.0$) as well. Thus, specificity did not significantly differ between both methods in any size category for both experienced and novice observers.

Review time

For novice observers mean review time for PU2D was 8.9 min (range 4.3–12.8 min) compared with 10.8 min (range 6.3–16.1 min) for PEC3D ($p < 0.001$). For experienced observers the review times were respectively 8.4 min (range 7.2–9.6 min) and 9.8 min (range 9.6–9.9 min) ($p < 0.001$).

For most observers PU2D was faster. One observer evaluated the examinations faster in primary cleansed 3D method (Table 1).

Diagnostic quality

The mean rating of the diagnostic quality is displayed in Table 5. Figure 3 shows that all observers rated the image quality of PEC3D significantly lower than PU2D, as the confidence intervals did not reach 1.

Table 3 Per-polyp sensitivity and false positives rate of experienced observers and novices

| Observer | Experience | Per-polyp sensitivity (%) | | | | | | Number of false positives | | | |
|----------------------------|-------------|---------------------------|-------------|-------------|-------------|-------------|-------------|---------------------------|--------|----|----|
| | | 2D | | | 3D | | | 2D | | 3D | |
| Number | | ≥6 mm | 6–9 mm | ≥10 mm | ≥6 mm | 6–9 mm | ≥10 mm | ≥6 mm | ≥10 mm | 2D | 3D |
| 2 | Experienced | 80 (70–87) | 65 (47–79) | 91 (79–96) | 82 (74–88) | 76 (61–86) | 87 (77–93) | 11 | 7 | 2 | 0 |
| 6 | Experienced | 78 (69–85) | 57 (43–70) | 92 (79–98) | 78 (67–86) | 65 (48–79) | 87 (73–94) | 9 | 12 | 1 | 2 |
| Mean experienced observers | | 79 (70–86) | 61 (46–74) | 92 (79–97) | 80 (72–86) | 70 (57–81) | 87 (75–93) | 10 | 10 | 2 | 1 |
| 1 | Novice | 63 (53–71) | 42 (27–58) | 77 (66–85) | 77 (67–85) | 67 (49–81) | 85 (74–91) | 2 | 4 | 1 | 1 |
| 3 | Novice | 59 (50–67) | 38 (24–54) | 74 (59–84) | 63 (54–72) | 46 (31–62) | 75 (61–86) | 7 | 8 | 1 | 3 |
| 4 | Novice | 56 (45–66) | 27 (14–45) | 75 (59–87) | 66 (55–75) | 41 (23–61) | 83 (68–92) | 8 | 10 | 3 | 4 |
| 5 | Novice | 53 (44–62) | 24 (12–42) | 74 (61–83) | 68 (58–76) | 46 (29–64) | 83 (74–90) | 2 | 9 | 1 | 3 |
| 7 | Novice | 29 (21–39) | 8 (3–23) | 45 (31–57) | 53 (43–64) | 32 (19–50) | 68 (49–82) | 4 | 13 | 4 | 7 |
| 8 | Novice | 49 (35–64) | 22 (9–43) | 70 (54–82) | 64 (52–75) | 49 (32–65) | 76 (61–86) | 14 | 18 | 7 | 2 |
| Mean novice observers | | 51* (44–59) | 27* (17–39) | 69* (59–77) | 65* (57–73) | 46* (33–60) | 78* (68–86) | 6 | 10 | 3 | 3 |

Individual per-polyp sensitivity is stratified for polyp size; 95% confidence intervals are between brackets

*Denotes statistically significant difference between the review methods

Table 4 Per-patient sensitivity and specificity of experienced observers and novices

| Observer | Experience | Per-patient sensitivity (%) | | | | Specificity (%) | | | |
|-----------------------------------|-------------|-----------------------------|-------------|-------------|-------------|-----------------|---------------|---------------|---------------|
| | | 2D | | 3D | | 2D | | 3D | |
| Number | | ≥6 mm | ≥10 mm | ≥6 mm | ≥10 mm | ≥6 mm | ≥10 mm | ≥6 mm | ≥10 mm |
| 2 | Experienced | 88 (79–97) | 97 (91–100) | 86 (76–96) | 91 (81–100) | 95 (87–100) | 97 (92–100) | 91 (79–100) | 100 (100–100) |
| 6 | Experienced | 84 (74–94) | 97 (91–100) | 80 (69–91) | 91 (82–100) | 95 (87–100) | 100 (100–100) | 91 (79–100) | 95 (88–100) |
| Mean values experienced observers | | 86 (79–93) | 97 (93–100) | 83 (76–90) | 91 (84–98) | 95 (89–100) | 99 (96–100) | 91 (82–99) | 97 (94–100) |
| 1 | Novice | 71 (58–84) | 82 (69–95) | 85 (75–95) | 91 (81–100) | 100 (100–100) | 97 (92–100) | 91 (79–100) | 97 (92–100) |
| 3 | Novice | 74 (62–86) | 85 (73–97) | 72 (60–84) | 79 (66–93) | 91 (79–100) | 100 (100–100) | 95 (87–100) | 95 (88–100) |
| 4 | Novice | 68 (55–81) | 85 (73–97) | 74 (62–86) | 91 (82–100) | 95 (87–100) | 95 (88–100) | 100 (100–100) | 95 (88–100) |
| 5 | Novice | 70 (57–83) | 88 (77–99) | 78 (67–89) | 91 (82–100) | 100 (100–100) | 97 (92–100) | 91 (79–100) | 97 (92–100) |
| 7 | Novice | 42 (28–56) | 59 (42–75) | 72 (60–84) | 85 (73–97) | 100 (100–100) | 97 (92–100) | 91 (79–100) | 95 (88–100) |
| 8 | Novice | 62 (48–76) | 81 (67–95) | 68 (55–81) | 81 (67–95) | 95 (87–100) | 95 (88–100) | 77 (60–95) | 95 (88–100) |
| Mean values novice observers | | 64 (59–70) | 80 (74–86) | 75* (70–80) | 86* (82–91) | 97 (94–100) | 97 (95–99) | 91 (86–96) | 96 (93–98) |

Individual per-patient sensitivity and specificity are stratified for polyp size; 95% confidence interval are between brackets)

*Denotes statistically significant difference between the review methods

Floating debris (Fig. 4) and holes in the colon wall (Fig. 5) were important causes of artefacts in PEC3D (Table 6). According to the observers, floating debris in particular may hinder the diagnostic accuracy by covering polyps 6 mm or larger. In PU2D virtually no disturbing artefacts were reported.

Discussion

This study shows that novice observers (compared with experienced observers) have a higher sensitivity in limited prepared patients when using PEC3D compared with PU2D. The higher sensitivity comes *without* a statistically significant lower specificity. On average more review time was needed for PEC3D. Paradoxically, despite its superior performance in polyp detection, the image quality of PEC3D was rated significantly less than for PU2D.

Recently, two comparative studies of primary 2D and primary 3D review methods have addressed the merit of both techniques [11, 12]. As in this study, more polyps of 6 mm or larger were detected using primary 3D, although in only one study the difference was statistically significant [11]. The reason for this superior sensitivity may be that abnormalities are visible to the observer for longer. Secondly, polyps that may have a similar appearance to folds on 2D are easier to distinguish from folds in 3D. However, the fact that primary 3D is often not used in CTC is probably based on practical grounds such as long review time and high computer requirements associated with this review technique in the past.

Compared with the abovementioned previous studies, in this study patients underwent a limited bowel preparation. There is a risk that the reported superiority of 3D in polyp detection would be neutralized by the reduction of visible colonic surface. In this study we report a large number of at least partially submerged polyps. This increases the risk of overlooking polyps [21, 22]. This is the reason why electronic cleansing was used in the 3D examinations.

Electronic cleansing has been subject to study for several years now [16, 23–27]. Recently two comparative studies of electronic cleansing were published. In both studies electronic cleansing had an additional value in terms of sensitivity for some observers [23, 24]. In this study we assumed it had an additional value as well: nearly half of the polyps were in at least one position at least partially covered by faecal material (Table 2).

Specific artefacts of electronic cleansing are described in the literature [13] e.g. ridges, pseudopolyps due to partial volume effect and floating debris due to untagged faecal material. These may be the reason that electronic cleansing for primary CTC evaluation has not often been used. The electronic cleansing algorithm we used in this study was specially designed to overcome artefacts of distracting ‘ridges’ emanating from locations where air, soft tissue and tagged material meet [16]. These ridges were in fact noted by none of the observers in this study.

Floating debris, though, was detected in the majority of patients examined in PEC3D (Table 6). An important cause of debris is noise due to heterogeneously or insufficiently tagged stool. The three patients were excluded because of these artefacts. These artefacts stress the fact that more than just a good cleansing algorithm is important in order to

Table 5 Mean rating of diagnostic quality

| Image quality of the examinations | 2D Percentage of the examinations | 3D Percentage of the examinations |
|---|---|---|
| Not diagnostic, polyps 10 mm or larger can be missed | 1 | 4 |
| Diagnostic with many artefacts, polyps 6–9 mm can be missed | 4 | 18 |
| Diagnostic with a small number of artefacts, polyps 6 mm or larger cannot be missed | 20 | 53 |
| Diagnostic without artefacts | 69 | 19 |
| Unknown | 5 | 5 |

Table displays the percentage of PU2D and PEC3D examinations in the various rating categories. The percentage represented the mean rating of all eight observers in 72 patients

achieve good 3D image quality i.e. a good tagging regimen, good patient compliance and good CT parameters. However, the three excluded patients were rated in PU2D by one to three observers “not diagnostic” as well. This stresses the fact that although 3D is more susceptible to tagging artefacts, 2D suffers from heterogeneous or insufficient tagging as well.

Artefacts were seen in the majority of patients reviewed in PEC3D. To be able to easily distinguish artefacts from polyps, it is important to be able to correlate electronically cleansed 3D images with complementary *original* uncleaned 2D images. This combination has limited the number of false positive findings when using PEC3D (Table 3). Using PEC3D did not statistically decrease specificity for any observer group at any size per-patient threshold i.e. true negative patients were not erroneously classified using PEC3D.

We have not studied a primary 3D reading method *without* electronic cleansing or primary 2D *with* electronic cleansing. Although interesting from a methodological point of view, we think that these approaches are not meaningful; the former is not since a large number of (partially) submerged polyps are prevented from being detected because they are otherwise covered by faecal material; the second approach is not since there is no need to electronically remove faecal material that can already be distinguished from colonic structures. However, one study [23] demonstrated an additional value in terms of polyp detection of cleansing in a 2D approach. An important difference compared with our study is the nature of the preparation: barium instead of iodine and no low-fibre diet. This results in more adherent and solid stool that is “mentally” more difficult to read than the quiet homogeneous fluid levels seen in our patient population. Therefore, electronic cleansing may prevent reader fatigue in this patient population.

The mean difference between PU2D and PEC3D was in accordance with the expected difference between both techniques. However, the expected baseline sensitivity for polyps 6 mm or larger was higher (70%) than the actual measured sensitivity (51%). This had consequences for the statistical power of the comparison; however, statistical significance was still reached for the group of inexperienced readers.

The higher per-polyp sensitivity of PEC3D mainly concerned polyps 6–9 mm (Table 3). The prevalence of adenomas with advanced features (i.e. villous components or high-grade dysplasia) in this size category tends to be low [28]. The joint guideline of the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer and the American College of Radiology recommends colonoscopy and polypectomy for polyps 6 mm or larger [1]. Thus, polyps in this size category may not be neglected.

The novice observers were trained according to the recommendations of the American College of Radiology and the European Society of Gastrointestinal and Abdominal Radiology [29, 30]; 50 CTC studies with 20–50% prevalence with personal feedback on all false positive and negative findings for polyps 6 mm or larger [29]. The response to training, though, is unpredictable and competence cannot be assumed after 50 cases [31]. In this study the two experienced observers (350 CTCs or more) outperformed the six novice observers. Thus, it is likely that the optimum number of training cases is more than 50, as suggested earlier [32]. However, the difference between these levels of experience in PEC3D is less than PU2D. So,

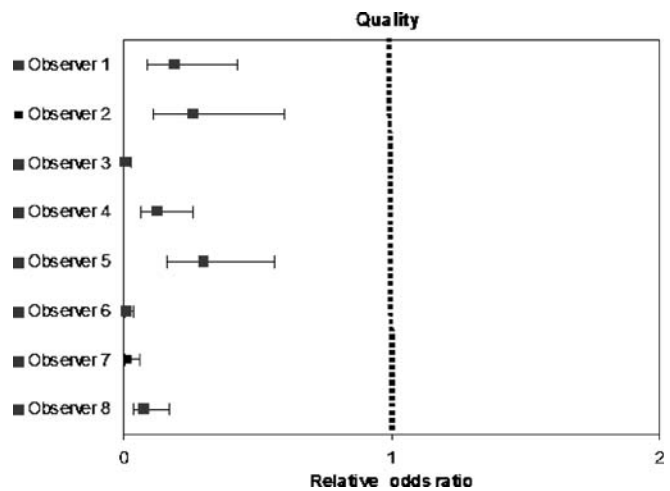


Fig. 3 Image quality assessment showing regression coefficients of primary electronically cleansed 3D (PEC3D). It estimates the change in the log transformed diagnostic odds ratio (DOR) compared with primary uncleaned 2D (PU2D). A lower relative diagnostic odds ratio (RDOR) implies inferior image quality. Confidence intervals not reaching 1 indicate significant inferiority. Thus, this figure shows that all observers rated the image quality of PEC3D significantly less qualitative than PU2D

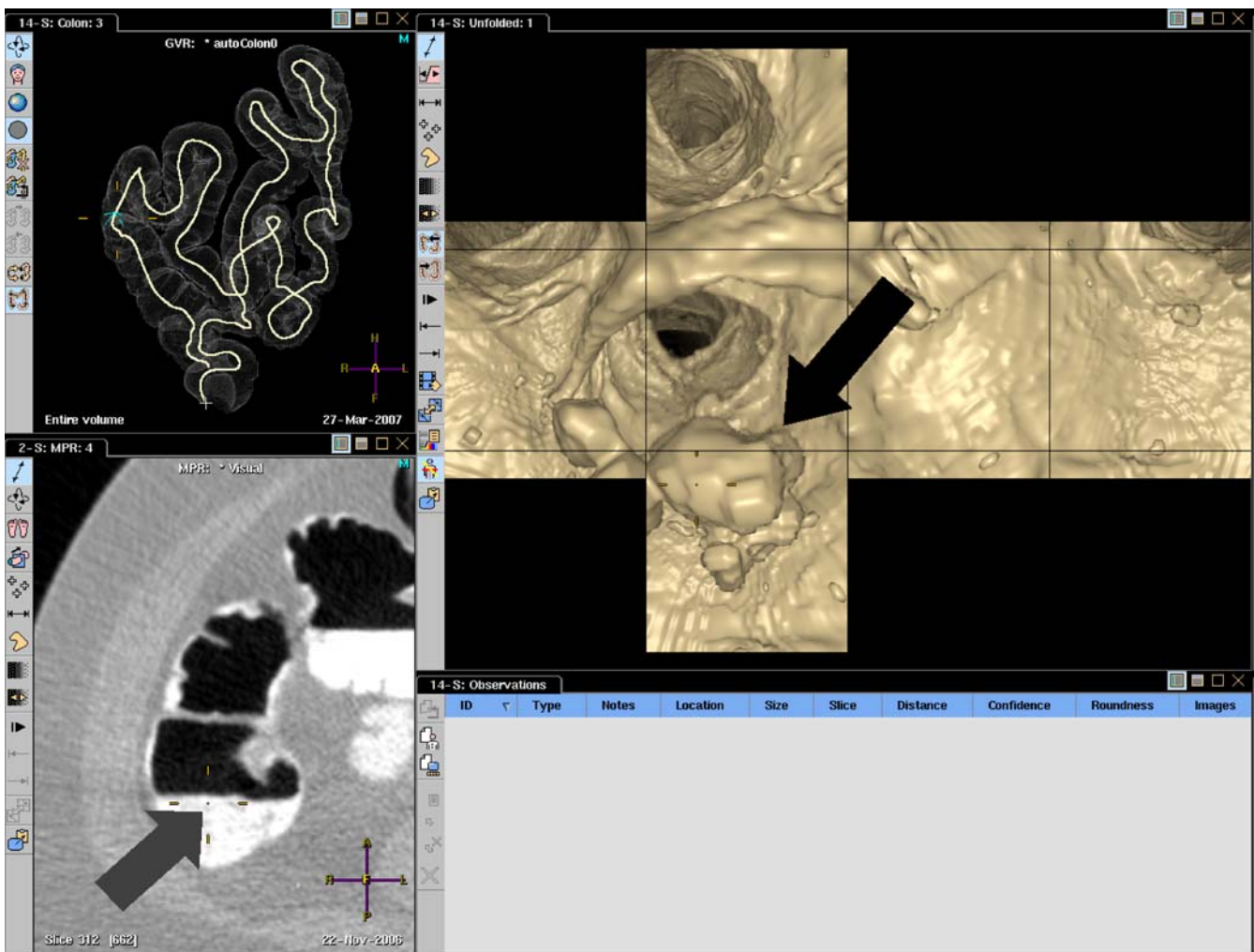


Fig. 4 Floating debris (*black arrow*) as a result of incomplete electronic cleansing of inhomogeneously tagged faecal material (*grey arrow* on 2D image)

in the phase of familiarization with CTC primary cleansed 3D is advantageous.

The group of experienced readers consisted of two observers compared with the group of inexperienced readers that consisted of six observers. In general, experienced observers show less difference in polyp detection between review methods [33] and will perform better when compared with inexperienced readers. Therefore, we expected the differences between both review methods to be insignificant in a very experienced observer group. This was confirmed by the two experienced reviewers in this study who performed very well using both methods. Thus, we think that although interesting from a methodological point of view adding four extra reviewers would not change the conclusion of this study.

A computer-aided detection algorithm has not been used in this study. Its effect will be twofold when applied to electronically cleansed data: the number of detectable polyps will increase as well as the number of detected

artefacts. Since these effects are not yet balanced, this will be subject to further research.

This study has limitations. First the prevalence of polyps in this FOBT positive population was fairly high compared with an asymptomatic screening population. This may limit the generalization to an average risk screening population.

Secondly, in this patient population we used a low-dose scan protocol combined with sub-millimetre slices. Noise in the images may have limited the quality of the cleansing since faecal material appeared less homogeneous. Still, the quality of the images was rated diagnostic in the vast majority of the cases i.e. only three cases were excluded.

Thirdly, all patients had been prepared with oral iodine tagging resulting in a fairly homogeneously tagged colon content. Probably any electronic cleansing algorithm will perform optimal with homogeneously tagged stool [27]. Accordingly, we expect the algorithm to be better suited for removing pools of (iodine) fluid compared with adherent heterogeneous faecal residue encountered in barium tagging.

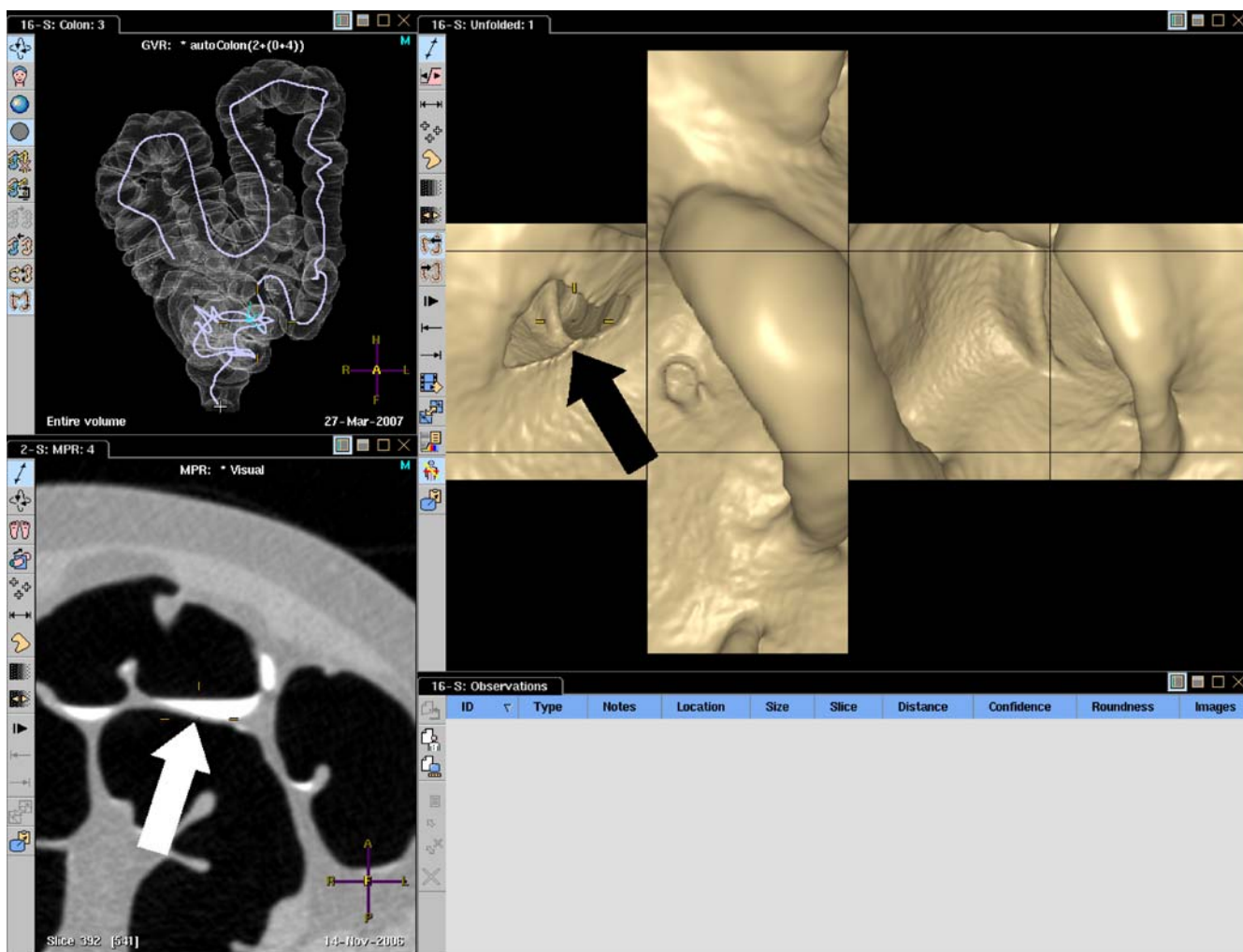


Fig. 5 Hole in the wall on the 3D image (*black arrow*). The *white arrow* indicates the corresponding colon wall in 2D. After electronic cleansing, the colon wall has become so thin that virtual holes appear between two air-containing structures

Table 6 Number of patients with artefacts

| Artefacts | Number of patients | Number of patients having artefacts classified as “disturbing, can cover polyps 6 mm or larger” | Number of patients having artefacts classified as “disturbing, can cover polyps 10 mm or larger” |
|---|--------------------|---|--|
| Primary electronically cleansed 3D | | | |
| Floating debris | 55 (76%) | 18 (25%) | 1 (1%) |
| Holes in the colon wall | 39 (54%) | 2 (3%) | 0 (0%) |
| Windmill artefacts | 1 (1%) | 0 (0%) | 0 (0%) |
| Ridges | 0 (0%) | 0 (0%) | 0 (0%) |
| Primary uncleaned 2D | | | |
| Windmill artefacts | 2 (3%) | 0 (0%) | 0 (0%) |
| Insufficient faecal tagging | 2 (3%) | 1 (1%) | 0 (0%) |

Table displays the number of patients with artefacts observed by at least four observers, the number of patients having artefacts classified by at least four observers as “disturbing, can cover polyps 6 mm or larger” and the number of patients having artefacts classified by at least four observers as “disturbing, can cover polyps 10 mm or larger”

Fourthly, before this study the experienced observers had evaluated all patients in the framework of a comparative study of colonoscopy and CTC (http://rsna2008.rsna.org/event_display.cfm?em_id=6012336). These patients were evaluated at least 1 year before this study with a primary 2D review method. In the period in between both studies at least 100 other CTC examinations were read. So, it is not likely that this has influenced the performance characteristics of the experienced observers.

Fifthly, in this study we have used an enhanced 3D display i.e. the unfolded cube display. The advantage of this technique is that it covers nearly all colonic mucosa without image distortion in a single fly-through [15], compared with the conventional 'endoscopic' view that needs a bidirectional fly-through to cover nearly all colonic mucosa. Therefore, our approach is a more time efficient method than a conventional 3D technique [15]. This may limit the generalizability of difference in review time, however *not* in accuracy.

In summary, we conclude that novice observers have a significantly higher sensitivity for the detection of clinically relevant polyps when using primary electronically cleansed 3D compared with primary 2D. For experienced observers, who performed better overall, there is no difference between both methods. Specificity is not affected when using primary electronically cleansed 3D. Therefore we recommend primary electronically cleansed 3D for novice observers in evaluating CTC in patients that have undergone limited bowel preparation.

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.

Acknowledgments A grant was received from Philips Healthcare.

References

- 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
- van Gelder RE, Birnie E, Florie J et al (2004) CT colonography and colonoscopy: assessment of patient preference in a 5-week follow-up study. *Radiology* 233:328–337
- Iannaccone R, Laghi A, Catalano C et al (2004) Computed tomographic colonography without cathartic preparation for the detection of colorectal polyps. *Gastroenterology* 127:1300–1311
- 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
- Lefere P, Gryspeerdt S, Marrannes J, Baekelandt M, van HB (2005) CT colonography after fecal tagging with a reduced cathartic cleansing and a reduced volume of barium. *AJR Am J Roentgenol* 184:1836–1842
- Iafrate F, Hassan C, Zullo A et al (2008) CT colonography with reduced bowel preparation after incomplete colonoscopy in the elderly. *Eur Radiol* 18:1385–1395
- Gryspeerdt S, Lefere P, Herman M et al (2005) CT colonography with fecal tagging after incomplete colonoscopy. *Eur Radiol* 15:1192–1202
- Gluecker TM, Johnson CD, Harmsen WS et al (2003) Colorectal cancer screening with CT colonography, colonoscopy, and double-contrast barium enema examination: prospective assessment of patient perceptions and preferences. *Radiology* 227:378–384
- Rex DK (2002) Current colorectal cancer screening strategies: overview and obstacles to implementation. *Rev Gastroenterol Disord* 2(Suppl 1):S2–S11
- Weitzman ER, Zapka J, Estabrook B, Goins KV (2001) Risk and reluctance: understanding impediments to colorectal cancer screening. *Prev Med* 32:502–513
- 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
- van Gelder RE, Florie J, Nio CY et al (2006) A comparison of primary two- and three-dimensional methods to review CT colonography. *Eur Radiol* 17:1181–1192
- Pickhardt PJ, Choi JH (2003) Electronic cleansing and stool tagging in CT colonography: advantages and pitfalls with primary three-dimensional evaluation. *AJR Am J Roentgenol* 181:799–805
- van Rossum L, van Rijn A, Laheij R et al (2008) Random comparison of Guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology* 135(1):82–90
- 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
- Serlie IW, Vos FM, Truyen R, Post FH, van Vliet LJ (2007) Classifying CT image data into material fractions by a scale and rotation invariant edge model. *IEEE Trans Image Process* 16:2891–2904
- Walter Reed Army Medical Center. Walter Reed Army Medical Center (2005) <http://wramc.vcscreen.com/training/index.html>. Accessed 23 Nov 2005
- 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

19. 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
20. Pepe MS (2003) The statistical evaluation of medical tests for classification and prediction. Oxford University Press, New York, pp 35–65
21. Doshi T, Rusinak D, Halvorsen RA, Rockey DC, Suzuki K, Dachman AH (2007) CT colonography: false-negative interpretations. *Radiology* 244:165–173
22. 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
23. Johnson CD, Manduca A, Fletcher JG et al (2008) Noncathartic CT colonography with stool tagging: performance with and without electronic stool subtraction. *AJR Am J Roentgenol* 190:361–366
24. Juchems MS, Ernst A, Johnson P, Virmani S, Brambs HJ, Aschoff AJ (2008) Electronic colon-cleansing for CT colonography: diagnostic performance. *Abdom Imaging*. doi: 10.1007/s00261-008-9386-6
25. Serlie IW, de Vries AH, Vos FM et al (2008) Lesion conspicuity and efficiency of CT colonography with electronic cleansing based on a three-material transition model. *AJR Am J Roentgenol* 191:1493–502
26. Wang Z, Liang Z, Li X et al (2006) An improved electronic colon cleansing method for detection of colonic polyps by virtual colonoscopy. *IEEE Trans Biomed Eng* 53:1635–1646
27. Zalis ME, Perumpillichira JJ, Magee C, Kohlberg G, Hahn PF (2006) Tagging-based, electronically cleansed CT colonography: evaluation of patient comfort and image readability. *Radiology* 239:149–159
28. Kim DH, Pickhardt PJ, Taylor AJ (2007) Characteristics of advanced adenomas detected at CT colonographic screening: implications for appropriate polyp size thresholds for polypectomy versus surveillance. *AJR Am J Roentgenol* 188:940–944
29. ACR practice guideline for the performance of computed tomography (CT) colonography in adults (2006) http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/dx/gastro/ct_colonography.aspx. Accessed 8 July 2008
30. Taylor SA, Laghi A, Lefere P, Halligan S, Stoker J (2007) European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. *Eur Radiol* 17:575–579
31. Taylor SA, Halligan S, Burling D et al (2004) CT colonography: effect of experience and training on reader performance. *Eur Radiol* 14:1025–1033
32. Spinzi G, Belloni G, Martegani A, Sangiovanni A, Del FC, Minoli G (2001) Computed tomographic colonography and conventional colonoscopy for colon diseases: a prospective, blinded study. *Am J Gastroenterol* 96:394–400
33. Johnson CD, Chen MH, Toledano AY et al (2008) Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med* 359:1207–1217