CT Colonography: Preliminary Assessment of a Double-Read Paradigm That Uses Computer-aided Detection as the First Reader[1]

Gabriella Iussich, MD
Loredana Correale, PhD
Carlo Senore, MD
Nereo Segnan, MD
Andrea Laghi, MD
Franco Iafrate, MD
Delia Campanella, MD
Emanuele Neri, MD
Francesca Cerri, MD
Cesare Hassan, MD
Daniele Regge, MD

Purpose: To compare diagnostic performance and time efficiency of double-reading first-reader computer-aided detection (CAD) (DR FR CAD) followed by radiologist interpretation with that of an unassisted read using segmentally unblinded colonoscopy as reference standard.

Materials and Methods: The local ethical committee approved this study. Written consent to use examinations was obtained from patients. Three experienced radiologists searched for polyps 6 mm or larger in 155 computed tomographic (CT) colonographic studies (57 containing 10 masses and 79 polyps ≥6 mm). Reading was randomized to either unassisted read or DR FR CAD. Data sets were reread 6 weeks later by using the opposite paradigm. DR FR CAD consists of evaluation of CAD prompts, followed by fast two-dimensional review for mass detection. CAD sensitivity was calculated. Readers' diagnoses and reviewing times with and without CAD were compared by using McNemar and Student *t* tests, respectively. Association between missed polyps and lesion characteristics was explored with multiple regression analysis.

Results: With mean rate of 19 (standard deviation, 14; median, 15; range, 4–127) false-positive results per patient, CAD sensitivity was 90% for lesions 6 mm or larger. Readers' sensitivity and specificity for lesions 6 mm or larger were 74% (95% confidence interval [CI]: 65%, 84%) and 93% (95% CI: 89%, 97%), respectively, for the unassisted read and 77% (95% CI: 67%, 85%) and 90% (95% CI: 85%, 95%), respectively, for DR FR CAD ($P = .343$ and $P = .189$, respectively). Overall unassisted and DR FR CAD reviewing times were similar (243 vs 239 seconds; $P = .623$); DR FR CAD was faster when the number of CAD marks per patient was 20 or fewer (187 vs 220 seconds, $P < .01$). Odds ratio of missing a polyp with CAD decreased as polyp size increased (0.6) and for polyps visible on both prone and supine scans (0.12); it increased for flat lesions (9.1).

Conclusion: DR FR CAD paradigm had similar performance compared with unassisted interpretation but better time efficiency when 20 or fewer CAD prompts per patient were generated.

Supplemental material: *http://radiology.rsna.org/lookup/suppl/doi:10.1148/radiol.13121192/-/DC1*

[1] From the Radiology Unit, Institute for Cancer Research and Treatment, FPO, Strada Provinciale 142, Km 3,95, 10060 Candiolo, Italy (G.I., D.C., D.R.); im3D S.p.A., Turin, Italy (L.C.); AOU S Giovanni Battista-CPO Piemonte, SCDO Epidemiologia dei Tumori 2, Turin, Italy (C.S., N.S.); Department of Radiological Sciences Oncology and Pathology, University of Rome La Sapienza, Rome, Italy (A.L., F.I.); Diagnostic and Interventional Radiology, Department of Oncology, Transplants and Advanced Technologies in Medicine, University of Pisa, Pisa, Italy (E.N., F.C.); Gastroenterology and Endoscopy Unit, Nuovo Regina Margherita Hospital, Rome, Italy (C.H.). From the 2010 RSNA Annual Meeting. Received June 8, 2012; revision requested August 16; final revision received December 10; accepted January 13, 2013; final version accepted February 13. **Address correspondence to** G.I. (e-mail: *gabriella.iussich@ircc.it*).

Computed tomographic (CT) colonography has been included among the recommended options for colorectal cancer screening (1). However, a number of issues have limited the growth of CT colonography screening (2); high interobserver variability is still reported, mainly due to differences in readers' experience (3), and reviewing CT colonography is a long process, even when using newer, more user-friendly software (4). Furthermore, the cost of a CT colonographic examination needs to be reduced to become a cost-effective screening test (5). To reduce reading time, Mani et al (6) suggested using computer-aided detection (CAD) in a first-reading mode, in which CAD serves as detector, with radiologists serving to accept or reject CAD prompts. Theoretically, the high sensitivity of CAD algorithms to detect clinically relevant polyps (7–9) could eliminate the need for a reader to search the entire data set for polyps and to focus only on CAD prompts. However, as CAD systems are mainly designed to detect polyps, automatic mass detection poses challenges (10,11). We sought to overcome this limitation by adding to first-reader (FR) CAD a short two-dimensional (2D) review of unannotated areas of the colon, searching for possible masses or larger lesions missed by CAD. Masses are readily detected by experienced readers; therefore, it is expected that there would be a short addition of time for reviewing CT colonographic images. The purpose of this study was to compare diagnostic performance and time efficiency of the double-reading (DR) FR CAD paradigm, followed by radiologist interpretation with the unassisted read, using segmentally unblinded colonoscopy as reference standard.

Advances in Knowledge

- When CT colonographic studies are interpreted by experienced readers, diagnostic performance of double-reading first-reader computer-aided detection (CAD) (DR FR CAD) is not different from that of unassisted reading: Mean per-patient sensitivity was 77% versus 74%, and mean per-patient specificity was 93% versus 90%.
- Detection of low conspicuous polyps proved difficult also with CAD; indeed, about 29% (eight of 28) of missed polyps were flat lesions compared with 4% (two of 51) of detected polyps.
- DR FR CAD reduces time of unassisted interpretation by 15% when the number of CAD prompts per patient is 20 or fewer.

Materials and Methods

This study was approved by the local ethical committee, and written consent to use results of examinations was obtained from each individual; im3D (Turin, Italy) provided CAD hardware, technical support, and the viewing software used in this study. One author (L.C.) is a researcher at im3D, and two authors (G.I. and D.R.) are research consultants for im3D. Authors (C.H., C.S., N.S., and D.C.) who were not employees of or consultants for im3D had control of any data and information that might present a conflict of interest for those authors who are employees of or consultants for im3D.

CT Colonographic Examinations

Given a prestudy calculation (as detailed below), a total of 160 CT colonographic examinations were obtained from a database containing examinations from two previous trials (12,13), including mainly high-risk asymptomatic individuals. CT colonographic findings were verified by using colonoscopy. Figure 1 shows the flow diagram of study data set collection: 100 cases with negative results were randomly chosen among 408 patients without lesions of 6 mm or larger; 60 cases with positive results were randomly chosen from the subgroup of 209 patients with polyps of 6 mm or larger. We assumed simple random selection and did not stratify patients according to polyp features. CT colonographic protocol was the following: 120 kVp, less than 50 effective mAs, and section thickness of 2.5 mm or less. Bowel preparation consisted in the following: low-fiber diet starting 3 days before the procedure, oral intake of a 45-mL vial of sodium biphosphate and sodium phosphate (Phospho soda; C. B. Fleet, Lynchburg, Va) 14–18 hours before CT colonography, and oral intake of 150 mL of diatrizoate meglumine and diatrizoate sodium (Gastrografin; Bracco Diagnostics, Milan, Italy) on the afternoon before the examination. Pneumocolon was obtained in all cases by introducing CO_2 with an automatic insufflator (Protocol Colon Insufflator; Bracco).

Power Calculation

Specificity represents a critical parameter for a screening test, as the main objective of screening should be to

Implication for Patient Care

- The use of DR FR CAD for CT colonographic interpretation should be considered because, compared with unassisted read, it reduces reviewing time with no reduction of diagnostic performance; thus, this approach may be attractive in the context of a high volume of screening CT colonography.

Published online before print
10.1148/radiol.13121192 **Content code:** GI

Radiology 2013; 268:743–751

Abbreviations:
CAD = computer-aided detection
CI = confidence interval
DR = double reading
FR = first reader
OR = odds ratio
3D = three-dimensional
2D = two-dimensional

Author contributions:
Guarantors of integrity of entire study, C.S., F.C., C.H.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, G.I., L.C., F.I., E.N., C.H., D.R.; clinical studies, G.I., D.C., F.C., C.H., D.R.; experimental studies, G.I., N.S., A.L., F.I., E.N., F.C.; statistical analysis, L.C., C.S., N.S., C.H.; and manuscript editing, G.I., L.C., C.S., A.L., F.I., E.N., C.H., D.R.

Conflicts of interest are listed at the end of this article.

Figure 1

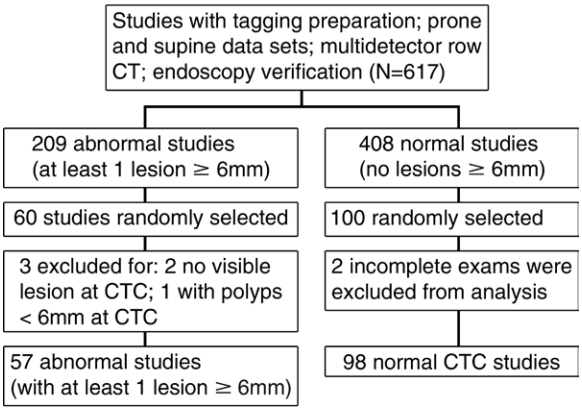

Figure 1: Flowchart shows selection of study data. Exclusion criteria were as follows: studies with incomplete data ($n = 2$), studies with endoscopic findings of 6 mm or larger not visible on CT images ($n = 2$), and studies with polyps smaller than 6 mm on CT images ($n = 1$). None of the selected studies were excluded because of technical inadequacy. *CTC* = CT colonography.

minimize procedure-related side effects for asymptomatic people attending the test. However, no data concerning specificity of a primary CAD were available at the time of writing. Thus, the power calculation was based on identifying a change of specificity. Detection of a 10% reduction in specificity ($\alpha = 5\%$, $\beta = 80\%$) between unassisted and CAD-assisted reads (90% vs 80%) for each reader required at least 90 patients without polyps of 6 mm or larger. With three readers and at least 52 patients containing polyps of 6 mm or less, the study had 80% power to detect a 10% difference in sensitivity between unassisted and CAD-assisted reads (80% vs 70%) for all readers combined (52 cases with positive results × three readers = 156 observations with positive results). The sample size was increased to 160 cases (100 cases with negative results and 60 cases with positive results) to allow for exclusion of patients with nondiagnostic tests, incomplete data, or findings invisible on CT images.

Reference Standard

Two radiologists (>200 CT colonographic validated cases reported) with knowledge of endoscopic and histologic results independently reviewed all cases to locate verified polyps on CT images. None of these radiologists participated in the subsequent reading study. Polyps were matched to the colonoscopy reference by using segment location (within the same or adjacent segment) and size (within a 50% margin of error) (14). All lesions were manually marked by tracing a three-dimensional (3D) box on the CT image to contain the lesion and were evaluated according to various characteristics used in previous studies (Table E1 [online]) (15,16). After lesion characterization, agreement between truthers (G.I. and D.C.) was verified: data were considered concordant if the 3D boxes were at least partially superimposed, measurements of diameter were within 2 mm, and other polyp features were the same. Disagreement was resolved by means of face-to-face discussion. Lesions identified at colonoscopy but not depicted on CT images ($n = 2$) were excluded because their spatial location could not be specified. Polyps of 5 mm or less ($n = 1$) were also excluded. All other lesions were defined positive by reference marking. Eighty-nine lesions (79 polyps ≥6 mm and 10 masses) were characterized. Forty-one lesions were 10 mm or larger and 48 were between 6 and 9 mm. Ten of the 79 polyps (13%) were flat, 27 were pedunculated (34%), and 42 were sessile (53%).

The CAD System

The commercially available CAD system used for the study (CAD-COLON-1.10; im3D) is described elsewhere (17–20). In brief, after electronic cleansing, the software extracts the colon from input CT images by applying a 3D region-growing algorithm. Colon surface voxels with suspicious curvature properties are then selected, and those whose shape and curvature indexes are within a predefined range are clustered according to spatial attenuation rules to yield initial marks. A linear classifier then attributes a score to each candidate. Candidate lesions with a score above a certain threshold (operating point) are then shown to the user. The corresponding CAD prompts are displayed on both 2D and 3D images with rectangular bounding boxes. None of the cases used in this study were used in system training. The operative point of the CAD, which was set to reach a sensitivity of 95% for lesions of 6 mm or larger at 11 false-positive results per series (22 false-positive results per patient), corresponds to the default setting of the CAD system. It was set during the training phase (19) to address the needs in a clinical situation for higher sensitivity.

Reviewing CT Colonography

Three additional radiologists (5–10 years of experience in CT colonography, results in >200 of 600–1200 examinations reported were verified by using colonoscopy) took part in the reading study. Each reader interpreted images from all examinations twice, at least 6 weeks apart, to minimize the recall bias. The images from the examinations were ordered randomly on a per-reader basis and were analyzed either as unassisted or with the DR FR CAD paradigm. Six weeks later, case ordering was randomized again, and images from the studies were interpreted by using the opposite paradigm. Readers were blinded to disease prevalence, and they were told to ignore polyps 5 mm or less. Polyps were measured as per CT Colonography Reporting and Data System guidelines (21). No time limit was imposed, but radiologists were expected to read images from a minimum of 25 cases in a day (8 hours per reading session). Regardless of the paradigm, radiologists were free to use the full functionality of the workstation when interpreting 2D axial images, multiplanar reconstructions, and endoluminal fly-through images. A primary 2D reading with 3D problem solving was used. When interpreting unassisted, radiologists analyzed images

from studies without CAD and characterized each abnormality, annotating its size, location, and morphologic characteristics. Interpretation time and a confidence score on a five-point scale (from definitely normal to definitely abnormal) were recorded. The use of the confidence score was limited to the generation of receiver operating characteristics. When using DR FR CAD, radiologists reviewed CAD prompts on both prone and supine scans, and lesions were classified as described above. After evaluating CAD prompts, radiologists scrolled through the 2D axial images searching for masses and/or evident lesions missed by CAD. Any additional finding was documented. For both reading paradigms and for each finding reported by readers, the 3D boxes on CT images containing the lesion (either manually traced or automatically estimated by CAD) were also recorded.

Figure 2

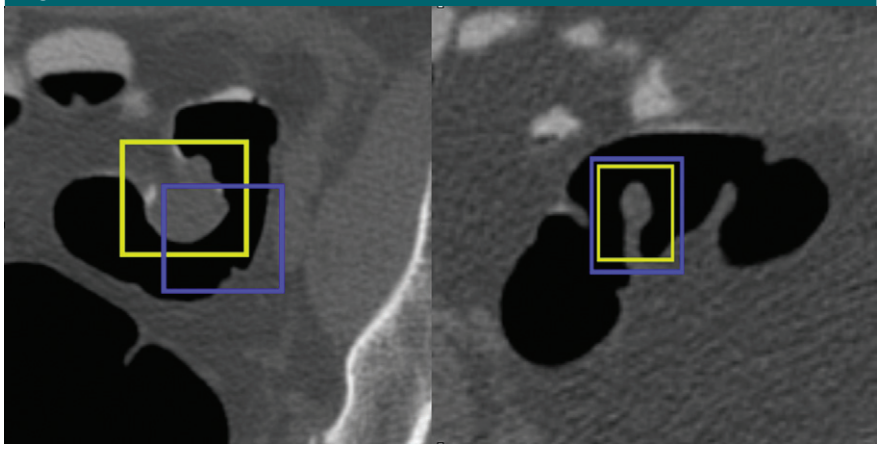

Figure 2: Two examples of correct reader detections (blue boxes) with their positive reference standard bounding boxes (yellow boxes) in a 63-year-old man. Detections were considered true-positive on the basis of the volume of the overlap with positive reference standard bounding boxes. Left: Detection overlapping with its positive reference standard bounding box of more than 25% Right: Detection containing all visible parts of its positive reference standard bounding box.

Automatic Lesion Matching

A detection was considered true-positive if any part of the detected bounding box matched any part of the reference standard bounding box (Fig 2); otherwise, it was considered false-positive (9). Lesions included in the reference standard but not observed were counted as false-negative. All detections were first automatically labeled with this criterion. Then, true-positive detections were sorted by their overlap with the reference standard bounding box, where the overlap was defined by the intersect volume divided by the union volume (22–24). True-positive detections overlapping with the reference standard less than 25% were reviewed by one of the study truthers (G.I. and D.C.) to ensure that the true lesion and not a neighboring structure was deemed as a true-positive detection. Detections that were found to match neighboring structures were eliminated as true-positive detections. As it was a detection problem, the threshold of 25% was set deliberately to low to account for inaccuracies in the bounding box data. Results for polyps were evaluated per patient, not per supine and prone position, so if a lesion was detected in one position it was considered a true-positive detection regardless of whether the reader identified the lesion in the complementary position. This algorithm was used only for the per-polyp analysis.

Data Analysis

First, the primary end point of the study was to compare per-patient specificity and sensitivity of the DR FR CAD paradigm to that of the unassisted read. A positive result for CT colonography was defined as detection of a polyp of 6 mm or larger. This determination was then deemed either true-positive or false-positive, depending on whether the reference standard confirmed that a polyp of 6 mm or larger was present. Negative results for CT colonography were similarly deemed true-negative or false-negative, depending on the results of the reference standard. Per-patient diagnoses given by each reader with and without CAD were compared by using the McNemar test. Sensitivity, specificity, and the 95% confidence intervals (CIs) of each were calculated for each reader and for each reading mode; the 95% CIs were calculated by taking a bootstrap approach (25). Nonparametric methods were used to estimate the receiver operating characteristic curves and the areas under the curve for each reader (26). Reviewing times were compared by using the paired *t* test; the relationship between increasing CAD prompts and reading time was investigated with linear regression.

Second, a per-polyp analysis was performed to identify polyp features associated with polyp detection. Polyps were separated according to a binary variable: those detected by none or by one of the study readers (ie, *missed polyps*) and those identified by two or three of the study readers (ie, *detected polyps*). Differences in characteristics between missed and detected polyps were tested by using the *t* test for continuous variables and by using the χ^2 test or the Fisher exact test, as appropriate. Multivariate analysis was performed (separately for each reading mode) by using logistic regression. Backward stepwise regression was used as a support to identify features more strongly associated with the outcome (27). Variables that were marginally significant ($P \leq .2$) with univariate comparison were initially entered, with the least significant variables ($P > .05$) sequentially eliminated. Results were expressed as odds ratios (ORs) of a polyp being missed and 95% CIs.

Table 1

Per-Patient Sensitivity and Specificity for Identifying Patients with 6 mm or Larger Polyps for Unassisted, FR CAD, and DR FR CAD Paradigm

Statistic and Reader	Unassisted Read		FR CAD		DR FR CAD		
	Percentage*	95% CI	Percentage*	95% CI	Percentage*	95% CI	*P* Value
Sensitivity							
Reader 1	74 (42/57)	60, 84	72 (41/57)	58, 83	72 (41/57)	58, 83	>.99
Reader 2	75 (43/57)	62, 86	79 (45/57)	66, 89	79 (45/57)	66, 89	.726
Reader 3	74 (42/57)	60, 84	79 (45/57)	66, 89	79 (45/57)	66, 89	.546
Average of all readers	74 (127/171)	65, 84	77 (131/171)	67, 85	77 (131/171)	67, 85	.343
Specificity							
Reader 1	91 (89/98)	83, 96	90 (88/98)	82, 95	89 (87/98)	80, 94	.815
Reader 2	95 (93/98)	88, 99	92 (90/98)	82, 95	92 (90/98)	82, 95	.727
Reader 3	95 (93/98)	88, 99	91 (89/98)	83, 96	91 (89/98)	83, 96	.344
Average of all readers	94 (275/294)	89, 97	91 (267/294)	85, 96	90 (266/294)	85, 95	.189

* Numbers in parentheses were used to calculate the percentages.

Table 2

Area under the Receiver Operating Characteristic Curve Estimate

Observer	Unassisted Read	DR FR CAD	Difference	*P* Value
Reader 1	0.84 (0.76, 0.92)	0.82 (0.74, 0.90)	−0.02 (−0.04,0.04)	.519
Reader 2	0.87 (0.78,0.90)	0.90 (0.84,0.96)	0.03 (−0.08,0.02)	.3
Reader 3	0.84 (0.76,0.92)	0.89 (0.79, 0.99)	0.05 (−0.11,0.007)	.083

Note.—The receiver operating characteristic curves without and with CAD were generated for each reader on the basis of association between the confidence scores (values ranging from one to five) and correct case classification (normal versus abnormal). Numbers in parentheses are 95% CIs.

Statistical analysis was performed by using statistical software (R; the R Foundation, Vienna, Austria) (28–33). Two-sided *P* values less than .05 were considered to indicate significance.

Results

Of the 160 selected cases, five were excluded from the analysis for incomplete examination ($n = 2$), lesion with a largest diameter less than 6 mm ($n = 1$), and patients with polyps identified at colonoscopy but invisible on CT colonographic scans ($n = 2$). All studies were deemed to be technically adequate (ie, bowel preparation and distention were of sufficient quality to exclude polyps ≥6 mm in normal segments). In the final 155 cases, 83 patients were men (average age, 57 years; range, 39–71 years) and 72 were women (average age, 56 years; range, 37–75 years). Fifty-seven patients had a total of 89 lesions of 6 mm or larger, including 79 polyps and 10 masses. Forty-eight patients had polyps and no masses, eight had masses and no polyps, and one had both masses and polyps. Forty-four patients had at least one advanced neoplasia of 6 mm or larger (ie, lesions ≥10 mm or with a 20% or more villous component or with high-grade dysplasia).

CAD Stand-alone Performance

Per-polyp CAD sensitivity was 90% (71 of 79; 95% CI: 81%, 96%) and 97% (30 of 31; 95% CI; 90%, 98%) for polyps of 6 mm or larger and 10 mm or larger, respectively. In nine patients with 10 masses, CAD detected all masses, for a sensitivity of 100%. The sensitivity for pedunculated, sessile, and flat lesions was 89% (24 of 27; 95% CI: 71%, 98%), 95% (40 of 42; 95% CI: 84%, 95%), and 70% (seven of 10; 95% CI: 35%, 96%), respectively. The mean number of false-positive detections per series was 9.6 (standard deviation, seven; median, eight; range, 2–64; interquartile range, 6–12); these data correspond to a per-patient mean number of 19 detections (range, 4–127 detections).

Per-Patient Analysis

Results of per-patient analysis are shown in Table 1. Mean sensitivity of unassisted and of CAD-assisted readings did not differ significantly (74% vs 77%, $P = .343$), and all masses were detected both with and without CAD. All lesions of 6 mm or larger detected at the DR FR CAD reading were initially correctly identified by the CAD system. No differences were found for detection of advanced neoplasia (81% vs 82% for DR FR CAD; $P > .99$). The mean specificity of the unassisted read was 93%; an insignificant reduction of specificity was observed using CAD (90%, $P = .189$). No specificity differences were observed when comparing studies with less than 15 CAD prompts per scan with those that had more than 15 CAD prompts per scan (OR, 1.1; 95% CI: 0.69, 3.5; $P = .303$). Area under the curve increased marginally by using CAD for two readers and decreased for one (Table 2). None of the individual area under the curve differences achieved significance.

Reading Time

The average reading time of unassisted and DR FR CAD was 243 seconds (95% CI: 229, 257) and 239 seconds (95% CI: 229, 257), respectively ($P = .623$). The mean reading time for the evaluation of CAD prompts was 186 seconds (95% CI: 170, 202); an additional 53 seconds (95% CI: 47, 60) was required for the 2D review. A positive correlation was observed between the increasing number of CAD prompts and interpretation time both with and without CAD (regression coefficient, 5.3 seconds [95% CI: 4.3, 6.3] and 1.6 seconds [95% CI: 0.6, 2.6]; $P = .02$ and $P < .001$, respectively). Compared with

Table 3

Per-Polyp Sensitivity for Each Reader and for All Readers Combined Stratified for Lesion Size

Observer	Unassisted Read		FR CAD		DR FR CAD		*P* Value
	Sensitivity (%)*	95% CI	Sensitivity (%)*	95% CI	Sensitivity (%)*	95% CI	
	For ≥6 mm Polyps						
Reader 1	70 (62/89)	59, 79	66 (59/89)	54, 76	66 (59/89)	54, 76	.438
Reader 2	71 (63/89)	60, 80	67 (60/89)	57, 77	67 (60/89)	57, 77	.654
Reader 3	63 (56/89)	52, 73	70 (62/89)	59, 79	70 (62/89)	59, 79	.382
Average of all readers	68 (181/267)	59, 76	68 (181/267)	59, 76	68 (181/267)	59, 66	>.99
	For ≥10 mm Polyps						
Reader 1	95 (39/41)	83, 99	90 (37/41)	77, 97	90 (37/41)	77, 97	.625
Reader 2	93 (38/41)	80, 98	88 (36/41)	74, 96	88 (36/41)	74, 96	.625
Reader 3	80 (33/41)	88. 99	85 (35/41)	83, 96	85 (35/41)	83, 96	.625
Average of all readers	89 (110/123)	83, 94	88 (108/123)	81, 93	88 (108/123)	81, 93	.774

* Numbers in parentheses were used to calculate the percentages.

unassisted interpretation, DR FR CAD significantly shortened interpretation time for studies generating 20 or fewer CAD prompts ($P < .01$) (Fig 3).

Per-Polyp Analysis

Table 3 reports results of per-polyp analysis. The pooled sensitivity for lesions of 6 mm or larger was 68% for both reading modes (95% CI: 59%, 76%; $P > .99$). No differences were observed even when comparing sensitivities at the 10-mm threshold ($P = .774$). During unassisted reading, 49 polyps were detected either by three ($n = 40$) or two readers ($n = 9$). The remaining 30 polyps were detected by one reader ($n = 13$) or by neither ($n = 17$). Using CAD, 51 polyps were detected either by two ($n = 16$) or three readers ($n = 35$). The remaining 28 polyps, of which 20 were correctly prompted by CAD, were either detected by one reader ($n = 17$) or by neither ($n = 13$). Interreader agreement of polyp detection varied substantially (Appendix E1 [online], Table E2 [online]). Appendix E2 (online) and Table E3 (online) report the characteristics of missed polyps (ie, detected by no reader or by one reader) and those of detected polyps (ie, detected by two or three readers). According to multivariate analysis (Appendix E2 [online]), the odds of missed detection decreased with increasing polyp size (OR, 0.6; 95% CI: 0.46, 0.9) and for

Figure 3

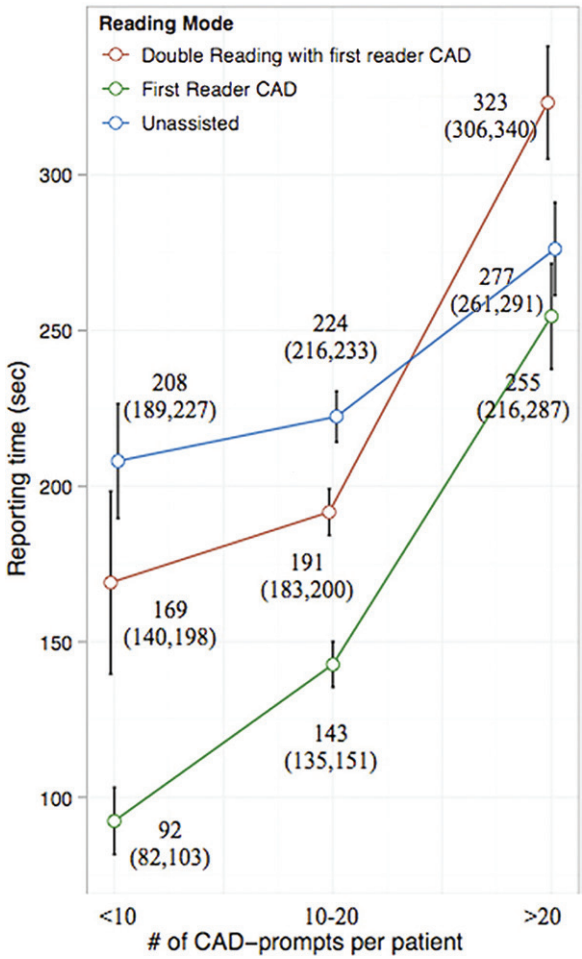

Figure 3: Graph of mean reviewing times as a function of the number of CAD prompts per patient according to reading strategy. Data in parentheses are 95% CIs.

Table 4

Results of Multiple Logistic Regression Modeling

Variable	Unassisted Read	*P* Value	DR FR CAD	*P* Value
Diameter	0.7 (0.5, 0.9)	.01	0.6 (0.46,0.9)	.008
Colon				
Proximal colon (reference)	. . .	. . .	1.0	
Distal colon			3.5 (0.8, 16)	.1
Lesion morphologic characteristics		.001		.03
Sessile (reference)	1.0	. . .	1.0	. . .
Pedunculated	0.2 (0.03, 1.1)	.06	0.95 (0.1, 7)	.96
Flat	9.1 (1.2, 188)	.03	14 (1.5,129)	.02
Polyps seen on both scans (reference)	1.0		1.0	
Polyps seen on only one scan	0.12 (0.01,0.88)	.04	0.08 (0.01,0.6)	.01
Relation to folds				
Not related (reference)	1.0	. . .	. . .	. . .
Related	0.6 (0.2, 2.8)	.5	. . .	. . .
Segment distention				
Optimal (reference)	1.0	. . .	1.0	. . .
Moderate or poor	0.22 (0.05, 1.06)	.06	0.5 (0.1,2.5)	.4
No. of CAD prompts	. . .	. . .	1.0 (0.1, 2.5)	.7

Note.—Numbers in parentheses are 95% CIs. Data are the odds of a polyp being missed either for one-unit increase in the explanatory variable (for variables on a continuous scale) or for each category relative to the odds of baseline category (for categorical explanatory variables).

polyps visible on both prone and supine scans (OR, 0.08; 95% CI: 0.01, 0.6). Conversely, the odds increased for flat polyps (OR, 14; 95% CI: 1.5, 129). A similar pattern was observed for the unassisted read (Table 4).

False-Positive Detections

With unassisted interpretation, there were 23 false-positive detections of 6 mm or larger in 21 patients (13.5% of cases); with DR FR CAD, there were 31 false-positive detections of 6 mm or larger in 22 studies (14% of cases). Of the 22 studies with false-positive detections, three had 10 or fewer, eight had 20 or more, and 11 had between 10 and 20 CAD prompts. False-positive detections of the unassisted read were mainly bulbous folds; the most frequent source of false-positive detections with the DR FR CAD read were fecal or fluid residues, followed by normal anatomy (Table E4 [online]).

Discussion

According to our study, per-patient sensitivity of unassisted interpretation and that of DR FR CAD interpretation were similar, and all 10 masses were recognized by using both modalities. With DR FR CAD, specificity was also reduced insignificantly, with no relationship between the increasing number of CAD prompts and the readers' specificity. This finding is consistent with findings in a previous study (34) in which experts used second-reader CAD in a low-prevalence population. FR CAD was less time-consuming than the unassisted read. As expected, the additional phase after CAD decreased time efficiency. Although this review added less than 1 minute for assessing CAD prompts, it represented a 22% increase in reviewing time. Unsurprisingly, the time to review the images with CAD increased with the number of CAD prompts. Unassisted reviewing time also depended on the number of CAD prompts. This factor may simply indicate that more complex images generate more CAD prompts but also that they require more time to be analyzed. Overall, reviewing time with DR FR CAD was similar to that of the unassisted read. However, this study also shows that DR FR CAD had better time efficiency than unassisted interpretation for studies generating 20 or fewer CAD prompts. Therefore, a review of a list of fewer than 20 CAD locations per patient is needed to improve overall time efficiency. Obviously, CAD threshold values should be appropriately adjusted for this purpose to maintain the same case-based CAD sensitivity.

A substantial number of polyps were unreported by most of the readers at both reading modes; more than 70% of them were correctly prompted by CAD. Interestingly, we found that the same polyps that were difficult to correctly characterize by unassisted readers were also hard to interpret by using CAD. Taylor et al (16) found that large polyps are at greater risk of being missed with CAD. Therefore, our result was opposite to their results. A possible explanation for this difference might be that we used the term *missed polyps* for those polyps unreported by most of the study readers. We believe that the detection of polyps by only one of the study readers was a result of interradiologist variation. It is therefore likely that when study radiologists had a consensus result of missing a polyp, this polyp would likely be incorrectly interpreted by most other general radiologists. Alternative explanations may include different readers' experiences and/or different CAD reading modes. We also observed that a lesion was more likely to be correctly identified if it had been prompted by CAD on both prone and supine scans rather than on only one acquisition. Thus, schemes such as "double-matching" CAD prompts (35) could be useful to make the best use of CAD-assisted reading. Not surprisingly, non–CAD-detected polyps were often missed at CAD-assisted reading. However, these polyps were also at greater risk of being missed without CAD. Thus, the fact that a small percentage (<10%) of polyps would be unmarked by CAD does not appear to influence reader sensitivity, as most of these would be dismissed anyway.

There were several limitations to this study. First, this was a retrospective study, and reading behavior might

not be the same as in a clinical setting (36). Although this was the largest study to date (to our knowledge), investigating an FR CAD scenario, a larger number of cases and readers would give even more robust data. More observers would be advantageous to provide a greater generalizability of the study results to a general reader population. However, studies with three readers are not uncommon in imaging literature (37) and are considered to be adequate for reporting diagnostic performance in the hands of experienced readers where interreader variability is expected to be low (37–39). The study readers had considerable experience in CT colonography; thus, our results may not be transferable to less experienced radiologists. Our data set was polyp enriched. This factor leads to a diagnostic performance different than that potentially achieved if a screening cohort (with few polyps) had been used. Despite that the sensitivity for 6 mm or larger lesions in our series was in agreement with the sensitivity in a previous large multicenter series, the reading time was much shorter, leading to potential underestimation of the time gain achievable with the DR FR CAD mode (40). All masses present in our data set were correctly identified by the CAD system, preventing us from assessing the post-CAD reader sensitivity in detecting eventual masses missed by the system. However, experienced readers are already well known to have a very high sensitivity for masses. On the other hand, it is unlikely that the lack of masses missed by CAD affected the validity of the assessment of specificity or reading time with this modality that may be considered as the main outcomes of our study. Our results reflect one particular CAD system; therefore, they may not be generalized to other systems. A further limitation was that, although we considered a number of variables as possible predictors of detection failure, it is possible that other variables that we did not assess were related to the outcome. Finally, while we cannot exclude recall bias, the reading paradigm used for each case was randomized between the two reading sessions for each reader, such that any recall bias would affect both paradigms equally.

In summary, a DR where CAD is the FR and the radiologist is the second reader has diagnostic performance that is similar to that of unassisted reading. Moreover, this reading paradigm is less time-consuming, particularly when an acceptable number of CAD marks are generated. Therefore, it may be an attractive reading strategy in a screening setting where the prevalence of disease is expected to be low and cost-effectiveness is an issue.

Disclosures of Conflicts of Interest: G.I. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: paid for consultancy by im3d Medical Imaging Lab, Torino, Italy. Other relationships: none to disclose. **L.C.** No relevant conflicts of interest to disclose. **C.S.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: received payment for accommodation and travel expenses for a scientific meeting organized by Given Imaging. Other relationships: none to disclose. **N.S.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: paid $1500 for participate at an advisory board meeting for Colorectal Cancer Screening in January 2011 as an expert on colorectal cancer blood screening assay for Roche Diagnostics Ltd. Other relationships: none to disclose. **A.L.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: institution received grants from GE Healthcare and Promefarm; author received payment for development of educational presentations from Bracco, Im3d, Bayer, and GE Healthcare. Other relationships: none to disclose. **F.I.** No relevant conflicts of interest to disclose. **D.C.** No relevant conflicts of interest to disclose. **E.N.** No relevant conflicts of interest to disclose. **F.C.** No relevant conflicts of interest to disclose. **C.H.** No relevant conflicts of interest to disclose. **D.R.** Financial activities related to the present article: institution received a grant from im3D. Financial activities not related to the present article: was paid consultancy fees by im3D. Other relationships: none to disclose.

References

1. Levin B, Lieberman DA, McFarland B, et al. 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 2008;134(5):1570–1595.
2. Klabunde CN, Lanier D, Nadel MR, McLeod C, Yuan G, Vernon SW. Colorectal cancer screening by primary care physicians: recommendations and practices, 2006-2007. Am J Prev Med 2009;37(1):8–16.
3. Heresbach D, Djabbari M, Riou F, et al. Accuracy of computed tomographic colonography in a nationwide multicentre trial, and its relation to radiologist expertise. Gut 2011;60(5):658–665.
4. Macari M, Milano A, Lavelle M, Berman P, Megibow AJ. Comparison of time-efficient CT colonography with two- and three-dimensional colonic evaluation for detecting colorectal polyps. AJR Am J Roentgenol 2000;174(6):1543–1549.
5. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Boer R, Wilschut J, Habbema JD. At what costs will screening with CT colonography be competitive? a cost-effectiveness approach. Int J Cancer 2009;124(5):1161–1168.
6. Mani A, Napel S, Paik DS, et al. Computed tomography colonography: feasibility of computer-aided polyp detection in a "first reader" paradigm. J Comput Assist Tomogr 2004;28(3):318–326.
7. Lawrence EM, Pickhardt PJ, Kim DH, Robbins JB. Colorectal polyps: stand-alone performance of computer-aided detection in a large asymptomatic screening population. Radiology 2010;256(3):791–798.
8. Summers RM, Handwerker LR, Pickhardt PJ, et al. Performance of a previously validated CT colonography computer-aided detection system in a new patient population. AJR Am J Roentgenol 2008;191(1):168–174.
9. Summers RM, Yao J, Pickhardt PJ, et al. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. Gastroenterology 2005;129(6):1832–1844.
10. Mang T, Peloschek P, Plank C, et al. Effect of computer-aided detection as a second reader in multidetector-row CT colonography. Eur Radiol 2007;17(10):2598–2607.
11. Robinson C, Halligan S, Iinuma G, et al. CT colonography: computer-assisted detection of colorectal cancer. Br J Radiol 2011;84(1001):435–440.
12. Regge D, Laudi C, Galatola G, et al. Diagnostic accuracy of computed tomographic colonography for the detection of advanced neoplasia in individuals at increased risk of colorectal cancer. JAMA 2009;301(23):2453–2461.
13. Regge D, Dellamonica P, Correale L, et al. The CAD-IMPACT study: computer-aided detection as second reader in a prospective,

multicenter setting [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2009; 401.

14. Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 2003;349(23):2191–2200.
15. Gluecker TM, Fletcher JG, Welch TJ, et al. Characterization of lesions missed on interpretation of CT colonography using a 2D search method. AJR Am J Roentgenol 2004;182(4):881–889.
16. Taylor SA, Robinson C, Boone D, Honeyfield L, Halligan S. Polyp characteristics correctly annotated by computer-aided detection software but ignored by reporting radiologists during CT colonography. Radiology 2009; 253(3):715–723.
17. Bert A, Dmitriev I, Agliozzo S, et al. An automatic method for colon segmentation in CT colonography. Comput Med Imaging Graph 2009;33(4):325–331.
18. Campanella D, Morra L, Delsanto S, et al. Comparison of three different iodine-based bowel regimens for CT colonography. Eur Radiol 2010;20(2):348–358.
19. Delsanto S, Morra L, Agliozzo S, et al. Computer-aided detection of polyps in virtual colonoscopy with sameday faecal tagging. In: Giger ML, Karssemeijer N, eds. Proceedings of SPIE: medical imaging 2008—computer-aided diagnosis. Vol 6915. Bellingham, Wash: SPIE–The International Society for Optical Engineering, 2008.
20. Regge D, Agliozzo S, Correale L, et al. Comparing the performance of a CT colonography computer-aided diagnosis (CAD) system to that of the experienced reader in a prospective multicenter setting [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2008; 279.
21. Zalis ME, Barish MA, Choi JR, et al. CT colonography reporting and data system: a consensus proposal. Radiology 2005;236(1):3–9.
22. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health. Draft Guidance for Industry and FDA Staff Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions. Rockville, Md: U.S. Department of Health, October 21, 2009.
23. Schilham AM, van Ginneken B, Loog M. A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. Med Image Anal 2006;10(2):247–258.
24. Petrick N, Sahiner B, Chan HP, Helvie MA, Paquerault S, Hadjiiski LM. Breast cancer detection: evaluation of a mass-detection algorithm for computer-aided diagnosis — experience in 263 patients. Radiology 2002;224 (1):217–224.
25. Efron B, Tibshirani R. An introduction to the bootstrap. London, England: Chapman & Hall, 1993.
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845.
27. Hosmer D, Lemeshow S. Applied logistic regression. New York, NY: Wiley, 1989.
28. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2009.
29. Lesnoff M, Lancelot R. aod: Analysis of overdispersed data R package version 1.1-33. http://cran.r-project.org.
30. Imai K, King G, Lau O, Toward A. Common framework for statistical analysis and development. J Comput Graph Stat 2008;17(4):892–913.
31. Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York, NY: Springer, 2002.
32. Sing T, Sander O, Beerenwinkel N, et al. ROCR: Visualizing the performance of scoring classifiers R package version 1.0-3. http://cran.r-project.org.
33. Robin X, Turck N, Hainard A, et al. Display and analyze ROC curves: tools for visualizing, smoothing and comparing receiver operating characteristic (ROC curves)—(partial) area under the curve (AUC) can be compared with statistical tests based on U-statistics or bootstrap: confidence intervals can be computed for (p)AUC or ROC curves R package 1.2.1. http://cran.r-project.
34. Taylor SA, Greenhalgh R, Ilangovan R, et al. CT colonography and computer-aided detection: effect of false-positive results on reader specificity and reading efficiency in a low-prevalence screening population. Radiology 2008;247(1):133–140.
35. Summers RM, Liu J, Rehani B, et al. CT colonography computer-aided polyp detection: effect on radiologist observers of polyp identification by CAD on both the supine and prone scans. Acad Radiol 2010; 17(8):948–959.
36. Gur D, Bandos AI, Cohen CS, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 2008; 249(1):47–53.
37. Shiraishi J, Pesce LL, Metz CE, Doi K. Experimental design and data analysis in receiver operating characteristic studies: lessons learned from reports in *Radiology* from 1997 to 2006. Radiology 2009; 253(3):822–830.
38. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol 2004;182(4):867–869.
39. Obuchowski NA. New methodological tools for multiple-reader ROC studies. Radiology 2007;243(1):10–12.
40. Johnson CD, Chen M-H, Toledano AY, et al. Accuracy of CT colonography for detection of large adenomas and cancers. N Engl J Med 2008;359(12):1207–1217.