Spiral CT Colonography: Reader Agreement and Diagnostic Performance with Two- and Three-dimensional Image-Display Techniques[1]

Elizabeth G. McFarland, MD
James A. Brink, MD
Thomas K. Pilgram, PhD
Jay P. Heiken, MD
Dennis M. Balfe, MD
Daniel A. Hirselj, BS
Leonard Weinstock, MD
Benjamin Littenberg, MD

Index terms:
Colon, CT, 75.12115, 75.12117
Colon neoplasms, 75.311
Computed tomography (CT), image processing, 75.12115, 75.12117
Computed tomography (CT), three-dimensional, 75.12117
Images, analysis, 75.12115, 75.12117

Radiology 2001; 218:375–383

Abbreviations:
2D = two-dimensional
3D = three-dimensional

[1] From the Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S Kingshighway Blvd, St Louis, MO 63110 (E.G.M., T.K.P., J.P.H., D.M.B., D.A.H.); the Department of Diagnostic Radiology, Yale University School of Medicine, New Haven, Conn (J.A.B.); the Department of Internal Medicine, Gastroenterology Division, Washington University School of Medicine, St Louis, Mo (L.W.); and the Primary Care Internal Medicine Group, Fletcher Allen Health Care, University of Vermont, Burlington (B.L.). From the 1998 RSNA scientific assembly. Received March 9, 2000; revision requested April 26; revision received June 2; accepted June 27. Supported by National Cancer Institute Prostate-Lung-Colon-Ovary Screening Trial grant N01-CN-25516 (E.G.M, J.A.B, T.K.P, J.P.H, D.M.B, L.W., B.L.), GE Medical Systems Association of University Radiologists (GERRAF; E.G.M.), Washington University Cancer Center (E.G.M.), and American Cancer Center Institutional Review Grant 36-38-04 (E.G.M.). **Address correspondence to** E.G.M. (e-mail: *mcfarlandb@mir.wustl.edu*).

E.G.M. and J.A.B. are medical advisory board members of Vital Images, which produces the software primarily used for the image analysis in this study.

Author contributions:
Guarantor of integrity of entire study, E.G.M.; study concepts, E.G.M., J.A.B., T.K.P., B.L.; study design, B.L., T.K.P., E.G.M.; definition of intellectual content, E.G.M.; literature research, E.G.M.; clinical studies, J.A.B., J.P.H., D.M.B., E.G.M.; data acquisition, E.G.M., L.W.; data analysis, D.A.H., E.G.M., T.K.P.; statistical analysis, T.K.P., B.L., E.G.M.; manuscript preparation, E.G.M.; manuscript editing and review, E.G.M., J.A.B., T.K.P., B.L., J.P.H.; manuscript final version approval, E.G.M.

PURPOSE: To assess the diagnostic performance and reader agreement with two-dimensional (2D) and three-dimensional (3D) display techniques for detecting colorectal polyps with spiral computed tomographic (CT) colonography.

MATERIALS AND METHODS: A test set of 30 colonic segments was developed from spiral CT colonographic studies (12 with polyps and 18 without). The 12 segments with polyps contained 22 lesions (11 polyps <10 mm, 11 polyps or cancers ≥10 mm), with all findings verified with colonoscopy. Three specific 2D and 3D image-display techniques were tested. Three experienced abdominal radiologists independently analyzed each test case and were retested 6 weeks later.

RESULTS: The results of readings 1 and 2 were similar for all image-display techniques among the readers. Pooled segment results were sensitivity of 89%–92% and specificity of 72%–83%. Pooled polyp size results for sensitivity and positive predictive value were 77%–86% and 74%–86% (all polyps, $n = 22$), 91%–100% and 85%–100% (polyps or cancers >10 mm, $n = 11$), and 61%–73% and 61%–80% (polyps 5–9 mm, $n = 11$), respectively. Overall intraobserver agreement was good for the three display techniques (κ, 0.60-1.00); however, interobserver agreement for 2D multiplanar reformation was lower (κ, 0.53–0.80).

CONCLUSION: Among experienced abdominal radiologists, similar diagnostic performance in polyp detection was found among 2D multiplanar reformation and 3D display techniques, although individual cases showed improved characterization with 3D display techniques. Evaluation of reader agreement demonstrated good intraobserver agreement, with variable interobserver agreement.

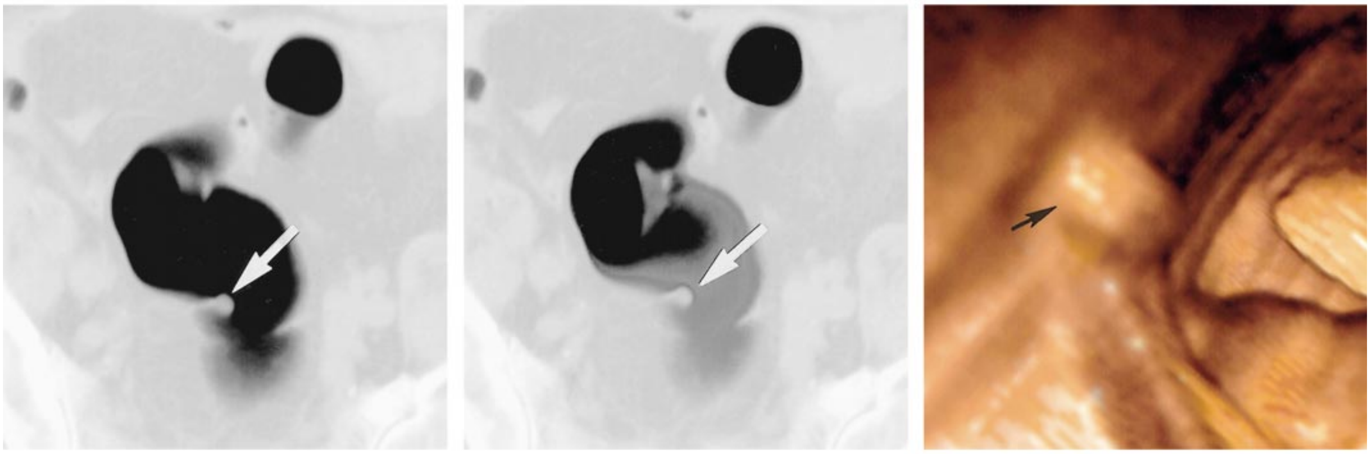
Spiral computed tomographic (CT) colonography is a rapidly evolving technology for the detection of colorectal polyps and permits interactive viewing with two-dimensional (2D) and three-dimensional (3D) image-display techniques (1–5). This imaging method is being widely investigated as a potential noninvasive examination for the detection of colorectal polyps. Recent advances in computer graphics have made possible a wide range of 3D display techniques, including 3D endoscopic navigation with improved interactivity. Preliminary studies (6–16) in which both 2D and 3D techniques were used have shown promising results; however, reader agreement with different image-display techniques has not been well documented.

Our purpose was to perform a multiobserver study of spiral CT colonography to (*a*) compare the diagnostic performance of various 2D and 3D image-display techniques for polyp detection and (*b*) assess intra- and interobserver agreement at CT colonography.

MATERIALS AND METHODS

Spiral CT Acquisition

Spiral CT colonography was performed in 16 patients (eight men, eight women; mean age, 61 years) who were suspected of having polyps at prior flexible sigmoidoscopy or



a. b. c.

Figure 1. Three image-display techniques tested in this reader study. **(a)** Transverse 2D multiplanar reformation, **(b)** transverse 3D multiplanar reformation, and **(c)** endoscopic view of 3D perspective volume-rendered CT images illustrate a 9-mm polyp (arrow) in the rectum. Of note for 2D and 3D multiplanar reformation displays, coronal and sagittal planes were also evaluated (not shown).

barium enema examination and were scheduled for colonoscopy. This study was approved by the institutional review board. All subjects gave informed consent. Patients underwent a standard 24-hour bowel cleansing preparation (phospho-soda, 24-hour Fleet 1 preparation, Fleet Pharmaceuticals, Lynchburg, Va; or polyethylene glycol electrolyte solution, GoLytely, Braintree Laboratories, Braintree, Mass) used for colonoscopy the day before the CT and colonoscopic examinations. Each patient underwent the CT examination before the colonoscopic examination. After air insufflation per rectum, the patients were placed in the supine and prone positions. An anteroposterior CT topogram was acquired to confirm adequate insufflation after each position, and further air was instilled per rectum as needed.

All scans were obtained with a spiral CT scanner (Somatom Plus 4, Siemens Medical Systems, Iselin, NJ; or HiSpeed Advantage, GE Medical Systems, Milwaukee, Wis). Images of the abdomen and pelvis were obtained by using 5-mm collimation, pitch of 1.6, and a reconstruction interval of 2 mm. In one patient, a different technique was used: 3-mm collimation, pitch of 2.0, and reconstruction interval of 1 mm; this patient was included as showing an excellent example of a focal polyp suspended on a fold. The transverse source images were transferred to a 3D workstation (Silicon Graphics 02; Silicon Graphics, Mountain View, Calif) with 1-gigabyte random accessible memory. Software analysis was performed (Vitrea 1.1; Vital Images, Minneapolis, Minn). For most of the data sets, further linear interpolation of the section data was performed and resulted in nearly isotropic sampling. The final in-plane pixel dimensions were 0.5–0.8 mm^2, and the final through-plane pixel dimensions were 0.6–2.0 mm. After the CT examination, the patients proceeded to colonoscopy. The colonoscopic examinations were performed by board-certified gastroenterologists in the gastroenterology division of the department of internal medicine (Washington University School of Medicine, St Louis, Mo), and each lesion found was carefully recorded.

An image library of 30 colonic segments from 16 patient studies was created for this reader study, with all CT findings strictly verified with colonoscopic results (17). The mean length of the segments was 18 cm, with a distribution of 13 rectosigmoid, five descending colon, five transverse colon, and seven ascending colon segments. For each colonic segment, embedded arrows were electronically placed on the images at the beginning and end of each segment to clearly demarcate the area for each reader to evaluate.

A total of 18 negative and 12 positive colonic segments were created; 17 were obtained with the patient in the supine position and three with the patient in the prone position. Cases were selected to give a range of clinically relevant polyp sizes and morphologies. In the 12 positive segments, 22 lesions were present and ranged from intermediate to large in diameter. There were 11 intermediate lesions (range, 5–9 mm; mean, 8 mm), with histologic findings of four hyperplastic lesions, five adenomatous lesions, and two indeterminate lesions, both of which were hyperplastic or adenomatous polyps histologically without being specified definitively. There were 11 polyps or cancers that were larger than 10 mm (range, 12–30 mm; mean, 19 mm), with histologic findings of two hyperplastic lesions, six adenomatous lesions, and three adenocarcinomas. All lesions were polypoid in morphology except one cancer that was a circumferential, advanced wall lesion.

Image Analysis

Three image-display techniques were tested (Fig 1). First, 2D multiplanar reformation permitted simultaneous interaction of transverse, coronal, and sagittal orientations at two fixed window level settings—a high-contrast polyp window width of 1,500 HU and level of −200 HU and a soft-tissue window width of 400 HU and level of 30 HU. Second, in a 3D multiplanar reformation we used a slab 10 mm thick in the transverse, coronal, and sagittal planes, with monochromatic display of the opacity assignment laid into the 2D sections. Third, in the 3D perspective volume-rendered endoscopic view we used a diverging lighting source to discriminate near-field from far-field objects. A nonlinear opacity function, which assigns a percentage of opacity across each attenuation value of the frequency histogram, and a mucosal color map assignment were used, as previously reported (18).

Reader Protocol

Three senior, fellowship-trained abdominal radiologists (J.A.B., J.P.H., D.M.B.)

TABLE 1
Reader Protocol for Three Image-Display Techniques

Colon Segments	Reader 1	Reader 2	Reader 3
1–10	2D MPR	2D MPR and 3D MPR	2D MPR and 3D PVR
11–20	2D MPR and 3D PVR	2D MPR	2D MPR and 3D MPR
21–30	2D MPR and 3D MPR	2D MPR and 3D PVR	2D MPR

Note.—MPR = multiplanar reformation image, PVR = perspective volume-rendered image.

with a mean of 14 years (range, 8–18 years) of faculty-level body CT experience participated in this study. Each reader underwent training in spiral CT colonography. The training protocol consisted of a teaching file of five patient data sets that had a range of different sizes and shapes of polypoid lesions. The readers used the teaching file to become familiar with the use of the three image-display techniques. Training involved testing and retesting with these teaching files over three to four sessions per reader. Trainees (J.A.B., J.P.H., D.M.B.) were observed and coached by an experienced reader (E.G.M.) until they demonstrated facility with thorough systematic review patterns for accurate polyp detection.

To make a feasible reading protocol and to decrease the recall bias of reading the same case multiple times, a reader study design was tested (Table 1). All three readers each read the images of the 30 colonic segments one time, with one of the three protocols randomly and uniquely assigned to each reader: (*a*) 2D multiplanar reformation images alone (n = 10); (*b*) 2D multiplanar reformation, followed by 3D multiplanar reformation images (n = 10); (*c*) 2D multiplanar reformation, followed by 3D perspective volume-rendered images (n = 10). For each case, the reader went through the designated segment with 2D multiplanar reformation and evaluated each focal finding with simultaneous interaction of the three views in the polyp window and soft-tissue window. The x, y, and z coordinates of each focal finding were recorded, and the readers measured the greatest long-axis dimension of the focal finding. A five-point rated response was used to evaluate the presence of a polyp or cancer for each focal finding: 1, definitely not present; 2, probably not present; 3, possibly present; 4, probably present; 5, definitely present. Each of the transverse, sagittal, or coronal views was then also scrolled through in its entirety for any additional intraluminal colorectal findings. An overall rated response for the presence of polyp or cancer was also assigned for the entire segment.

For the latter two protocols, the additional 3D multiplanar reformation and 3D perspective volume-rendered display techniques were also evaluated directly after the 2D multiplanar reformation, first by rescoring the focal findings seen in 2D multiplanar reformation with the rated response. The entire segment was then evaluated for any additional focal findings. For 3D thick-slab multiplanar reformation, this involved scrolling through the three planes. For 3D perspective volume-rendered, the readers interactively went through the 3D endoscopic view by using the 3D endoscopic autonavigation, with forward and backward viewing.

Six weeks after completion of reading one, each reader reevaluated all 30 colonic segments with the same display technique for each case but with a new randomly assigned order. The time required to evaluate each case was recorded for each reader for both the first and second reading sessions. After completion of reading two, readers were interviewed regarding the amount of recall of the cases.

The image quality of the colonic segments was graded for each segment by each reader using four-point subjective grading scales. Parameters evaluated were luminal distention (excellent, good, fair, poor), fluid degradation (none, mild, moderate, severe), tortuosity (none, minimal, moderate, marked), and overall depiction (excellent, good, fair, poor).

Statistical Analysis

Diagnostic performance.—The individual and pooled results of CT compared with colonoscopy were analyzed according to both segment (n = 30) and polyp (n = 22) across the three different image-display techniques for reading one and reading two. For analysis according to segment, the sensitivity and specificity of the three techniques, with 95% CIs, were calculated after binary collapse of the five-point rated response data: 1–2 were negative and 3–5 were positive. Results were calculated for individual readers and pooled for all three readers. For analysis according to polyp, sensitivity and positive predictive value with 95% CIs were calculated for all polyps, polyps or cancers at least 10 mm in size, and polyps less than 10 mm in size. Statistical significance at $P < .05$ was determined when the mean value of a parameter was not contained within the 95% CIs of another parameter. In addition, receiver operating characteristic analyses were performed on the five-point rated response data, and the areas under the curves plus or minus the SEM were calculated.

Reader agreement.—Intraobserver agreement for the three image-display techniques was determined with both the κ statistic and the Spearman rank order correlation. The κ statistic was used to evaluate agreement of the categorical divisions of negative (rated responses of 1–2) and positive (rated responses of 3–5) responses, with the κ scores ranging from −1 (perfect disagreement) to +1 (perfect agreement). Spearman rank order correlation was used to evaluate agreement of the actual five-point rated response data with the calculated values of the Spearman rho coefficient, which ranged from −1 (perfect disagreement) to +1 (perfect agreement). Diagnostic performance was analyzed according to colonic segment for individual and pooled results. Interobserver agreement for the 2D multiplanar reformation was also determined with these two statistical tests.

Timed reader sessions.—For each image-display technique, pooled reading times were determined. The relationship between reading times for each image-display technique and image quality scores was examined with analysis of variance. The image quality score was the categorical independent variable, and reading time was the continuous dependent variable.

RESULTS

Diagnostic Performance among the Three Image-Display Techniques

Colonic segment.—Table 2 shows the diagnostic performance according to colonic segment (n = 30; 12 positive and 18 negative segments) for the individual and pooled results across the three image-display techniques. Overall, there were no statistically significant differences among readers or image-display techniques. There was a slight improvement in specificity at reading two, which was

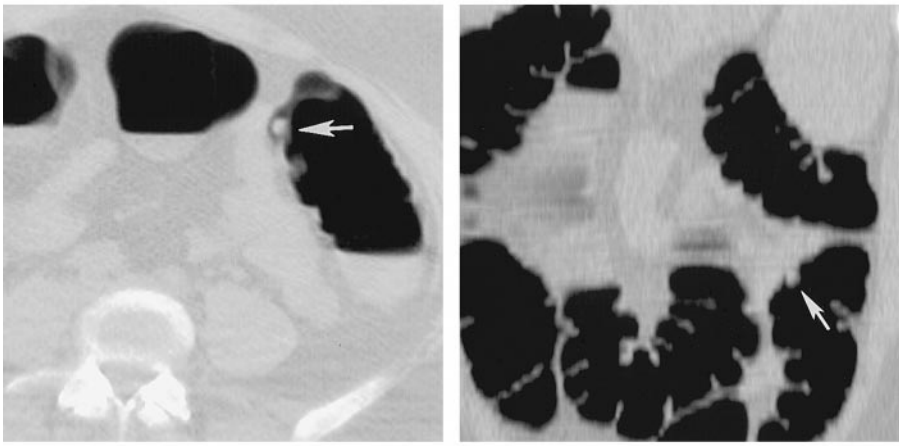
TABLE 2
Diagnostic Performance in Each Segment

Readers	Reading 1						Reading 2					
	TP	FN	TN	FP	Sensitivity*	Specificity*	TP	FN	TN	FP	Sensitivity*	Specificity*
Individual												
2D MPR												
1	12	0	13	5	100 (70, 100)	72 (46, 89)	11	1	13	5	92 (60, 100)	72 (46, 89)
2	10	2	16	2	83 (51, 97)	89 (64, 98)	10	2	15	3	83 (51, 97)	83 (58, 96)
3	11	1	13	5	92 (60, 100)	72 (46, 89)	11	1	15	3	92 (60, 100)	83 (58, 96)
3D MPR												
1	5	0	5	0	100 (46, 100)	100 (46, 100)	5	0	5	0	100 (46, 100)	100 (46, 100)
2	2	1	5	2	67 (13, 98)	71 (30, 95)	2	1	5	2	67 (13, 98)	71 (30, 95)
3	4	0	4	2	100 (40, 100)	67 (24, 94)	4	0	5	1	100 (40, 100)	83 (36, 99)
3D PVR												
1	4	0	5	1	100 (40, 100)	83 (36, 94)	4	0	5	1	100 (40, 100)	83 (36, 94)
2	5	0	3	2	100 (46, 100)	60 (17, 93)	5	0	5	0	100 (46, 100)	100 (46, 100)
3	2	1	5	2	67 (13, 98)	71 (30, 95)	2	1	5	2	67 (13, 98)	71 (30, 95)
Pooled												
2D MPR	33	3	42	12	92 (76, 98)	78 (64, 88)	32	4	43	11	89 (73, 96)	80 (66, 89)
3D MPR	11	1	14	4	92 (60, 100)	78 (52, 93)	11	1	15	3	92 (60, 100)	83 (52, 93)
3D PVR	11	1	13	5	92 (60, 100)	72 (46, 89)	11	1	15	3	92 (60, 100)	83 (52, 93)

Note.—Data are the number of segments (n = 30). A total of 30 colonic segments were evaluated; 12 contained polyps and 18 did not. For 2D MPR, each reader evaluated all 30 segments (90 pooled segments). For 3D displays, each reader evaluated 10 different segments (30 pooled segments). FN = false-negative, FP = false-positive, MPR = multiplanar reformation image, PVR = perspective volume-rendered image, TN = true-negative, TP = true-positive.
* Data are percentages. Data in parentheses are the 95% CI.

not statistically significant, and no change in sensitivity. Although there were slight improvements in reader performance during reading two compared with during reading one, no statistically significant differences were present. Sensitivities among the individual readers for 2D multiplanar reformation were 83%–100%. Pooled sensitivities among all three image-display techniques were 89%–92%. Specificities among the individual readers for 2D multiplanar reformation were 72%–89%. Pooled specificities among all three image-display techniques were 72%–83%. The receiver operating characteristic analyses demonstrated that the area under the curve plus or minus the SEM for 2D multiplanar reformation ranged from 0.91 ± .06 to 0.98 ± .02 across the three readers. Pooled areas under the curve were highest for 3D perspective volume-rendered technique (0.96 ± 0.04); however, no statistically significant differences were present.

Polyp.—Table 3 shows the diagnostic performance according to lesion (n = 22; 11 less than and 11 greater than 10 mm) for the individual and pooled results across the three image-display techniques. Similar to the colonic segment results, there were no statistically significant differences among readers or among image-display techniques for reading one and reading two results. Pooled sensitivity and positive predictive values were 77%–86% and 74%–86% for all polyps (n = 22), 91%–100% and 85%–100% for polyps or cancers 10 mm or larger (n = 11), and 61%–73% and 61%–80% for 5–9-mm polyps (n = 11), respectively. Figures 2–5 illustrate representative cases of varying reader performance.



a. b.

Figure 2. **(a)** Transverse and **(b)** coronal 2D multiplanar reformation CT images show a 7-mm polyp (arrow) in the descending colon. The polyp was correctly diagnosed as such by all three readers with the 2D multiplanar reformation images alone.

There were 15 false-negative findings among the three readers at the initial 2D multiplanar reformation evaluation for reading one, which encompassed eight lesions—three lesions missed by all three readers, one lesion missed by two readers, and four lesions missed by one reader. For these eight false-negative findings, the mean size was 10 mm—six polyps were less than 10 mm and two were greater than 10 mm (range, 8–18 mm); seven polyps were present in the sigmoid colon and one in the ascending colon near the hepatic flexure.

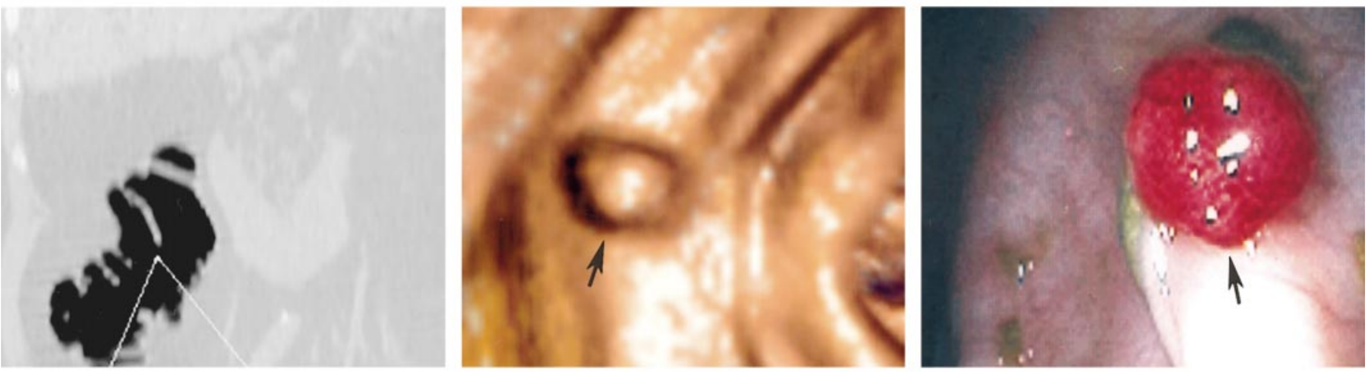
For reading two, there were 11 false-negative findings among the three readers at the initial 2D multiplanar reformation evaluation, which encompassed five lesions, with no new lesions missed compared with findings at reading one. The

TABLE 3
Diagnostic Performance in Each Polyp

Readers	Reading 1 TP	Reading 1 FN	Reading 1 FP	Reading 1 Sensitivity*	Reading 1 Positive Predictive Value*	Reading 2 TP	Reading 2 FN	Reading 2 FP	Reading 2 Sensitivity*	Reading 2 Positive Predictive Value*
Individual										
2D MPR										
1	19	3	7	86 (64, 96)	73 (52, 88)	19	3	5	86 (64, 96)	79 (57, 92)
2	16	6	4	73 (50, 88)	80 (56, 93)	17	5	4	77 (54, 91)	81 (57, 94)
3	16	6	7	73 (50, 88)	70 (47, 86)	19	3	5	86 (64, 96)	79 (57, 92)
3D MPR										
1	7	0	0	100 (56, 100)	100 (56, 100)	7	0	0	100 (56, 100)	100 (60, 100)
2	5	4	2	56 (23, 85)	71 (30, 95)	5	4	2	56 (23, 85)	71 (30, 95)
3	6	0	1	100 (52, 100)	86 (42, 99)	6	0	2	100 (52, 100)	75 (36, 96)
3D PVR										
1	6	0	1	100 (52, 100)	86 (42, 99)	6	0	2	100 (46, 100)	75 (36, 96)
2	6	1	2	86 (42, 99)	75 (36, 96)	7	0	1	100 (56, 100)	88 (47, 99)
3	5	4	3	71 (30, 95)	62 (26, 90)	6	3	2	67 (31, 91)	75 (36, 96)
Pooled										
All polyps										
2D MPR	51	15	18	77 (65, 86)	74 (62, 83)	55	11	14	83 (72, 91)	80 (68, 88)
3D MPR	18	4	3	82 (59, 94)	86 (63, 96)	18	4	4	82 (59, 94)	82 (59, 94)
3D PVR	17	5	6	77 (54, 91)	74 (51, 89)	19	3	5	86 (64, 96)	79 (57, 92)
Polyps 10 mm or larger										
2D MPR	31	2	5	94 (78, 99)	86 (70, 95)	31	2	3	94 (78, 99)	91 (75, 98)
3D MPR	11	0	1	100 (68, 100)	92 (60, 100)	11	0	0	100 (68, 100)	100 (68, 100)
3D PVR	10	1	1	91 (57, 100)	91 (57, 100)	11	0	2	100 (68, 100)	85 (54, 97)
Polyps smaller than 10 mm										
2D MPR	20	13	13	61 (42, 77)	61 (42, 77)	23	10	10	70 (51, 84)	70 (51, 84)
3D MPR	7	4	2	64 (32, 88)	78 (40, 96)	7	4	4	64 (32, 88)	64 (32, 88)
3D PVR	8	3	5	73 (39, 93)	62 (32, 85)	8	3	2	73 (39, 93)	80 (44, 96)

Note.—Data are the number of polyps (n = 22). A total of 22 colorectal lesions were evaluated; 11 were 10 mm or larger, and 11 were smaller than 10 mm. For 2D MPR, each reader evaluated all polyps (22 polyps, 66 pooled polyps). For 3D displays, each reader evaluated 10 different cases (22 pooled polyps). FN = false-negative, FP = false-positive, MPR = multiplanar reformation image, PVR = perspective volume-rendered image, TP = true-positive.

* Data are percentages. Data in parentheses are the 95% CI.



a. b. c.

Figure 3. CT images depict an 8-mm polyp in the ascending colon near the hepatic flexure, which was detected only with the 3D perspective volume-rendered technique. **(a)** Coronal 2D multiplanar reformation image obtained in a region of multiple nodular folds, with field of view shown for the 3D technique, demonstrates that no lesion could be discriminated. **(b)** Endoscopic view of 3D perspective volume-rendered image depicts the polyp (arrow). **(c)** Corresponding colonoscopic image depicts the polyp (arrow).

additional 3D images, obtained in 10 of 15 false-negative cases at reading one and eight of 11 false-negative cases at reading two, made a difference in two of these false-negative cases. In one case, an 8-mm polyp in the ascending colon was missed with 2D multiplanar reformation by all readers. However, the one reader who further evaluated the case with 3D perspective volume-rendered images detected it only during reading two but gave it a low response score of 2 (false-negative finding) (Fig 3). The second case was a 12-mm polyp in the sigmoid colon, which was given a low score with 2D multiplanar reformation (false-negative finding) but was correctly classified as a true-positive polyp with 3D perspective volume-rendered images by the same reader for both reading one and reading two (Fig 4). No polyps were inappropriately downgraded from a true-positive finding with 2D multiplanar reformation to a false-negative finding with the 3D views by any reader for reading one or reading two.

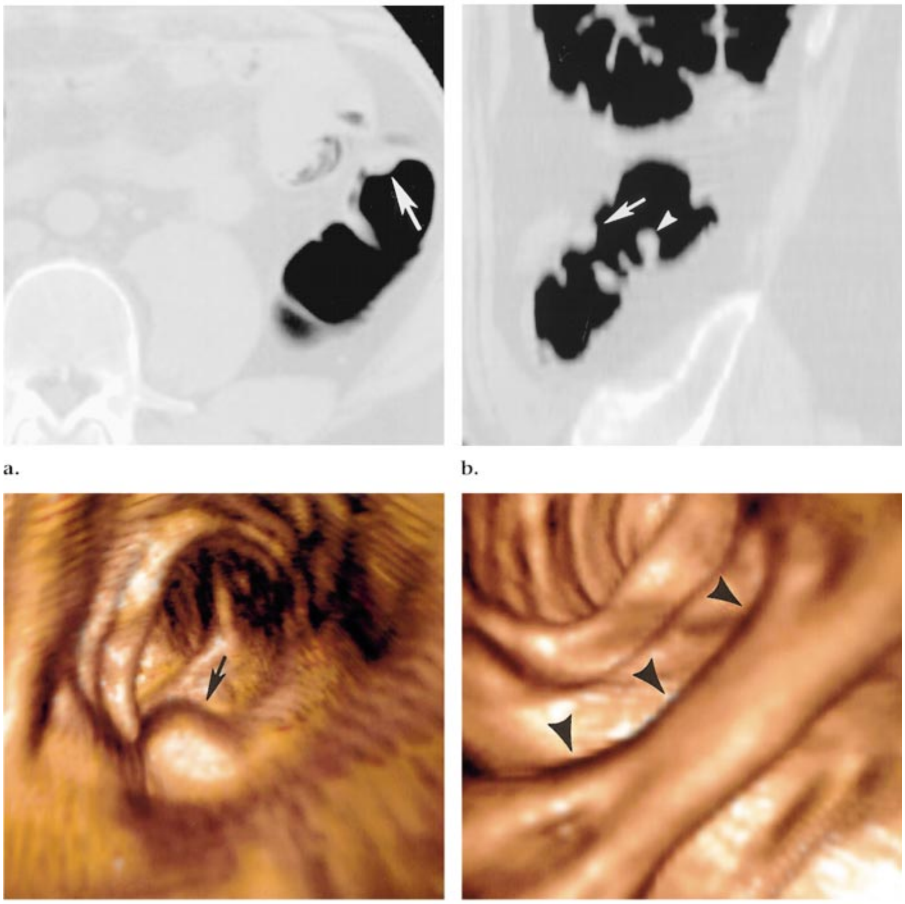
There were 18 false-positive cases with the 2D multiplanar reformation evaluations among the three readers for reading one, which encompassed 14 findings—one in all three readers, two in two readers, and 11 in one reader. Of these 14 false-positive findings, the mean size estimated by the readers was 7 mm (range, 3–13 mm), with five false-positive findings 10 mm or greater. These false-positive findings were present in the rectum ($n = 1$) (Fig 5), sigmoid ($n = 3$), ascending colon ($n = 6$), transverse colon ($n = 3$), and descending colon ($n = 1$).

For reading two, there were 14 false-positive findings among the three readers, which included nine findings, with two new false-positive findings compared with findings at reading one. The additional 3D images, obtained in 10 of 18 false-positive cases at reading one and 11 of 14 false-positive cases at reading two, were used to correctly reclassify lesions as true-negative findings in four cases (one at reading one and three at reading two). These reclassified true-negative cases included a 12-mm polyp in the rectosigmoid, a 7-mm polyp in the sigmoid, a 5-mm polyp in the descending colon, and a 7-mm polyp in the transverse colon, with three of these cases evaluated by the same reader. No polyps were inappropriately upgraded from a true-negative finding with 2D multiplanar reformation to a false-positive finding with the 3D views by any reader at reading one.

At reading two, two readers each incorrectly reclassified a lesion to a false-positive finding on the basis of the additional 3D views. One false-positive finding was an 8-mm lesion in the sigmoid colon where thickened folds were present. The other false-positive finding was a 5-mm rectal lesion and was found only with the 3D perspective volume-rendered view.

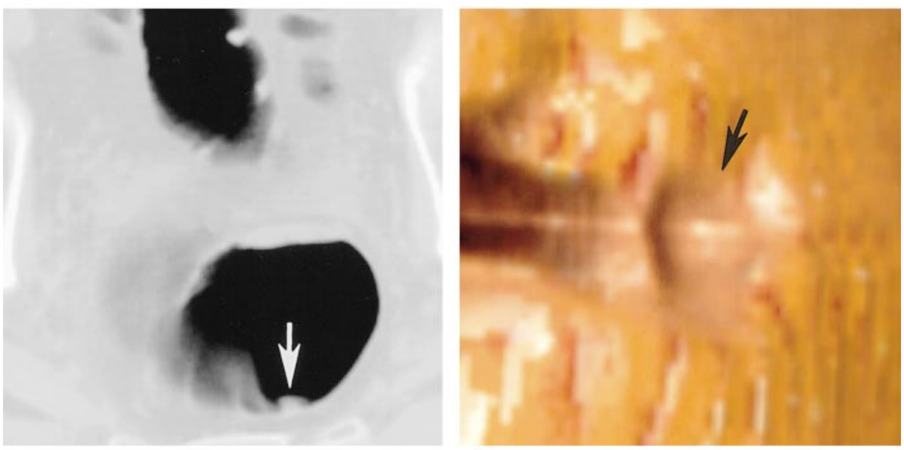
Reader Agreement

Intraobserver agreement.—Table 4 shows the intraobserver agreement for the three display techniques for each reader, as analyzed according to colonic segment for both reading one and reading two. No statistically significant differences were seen among readers or image-display techniques. Individual κ results, based on categorical divisions of positive or negative segments, were 0.73–0.93 for 2D



a. b. c. d.

Figure 4. CT images obtained with the patient prone depict a 12-mm polyp in the proximal sigmoid next to a nodular fold and illustrate improved characterization of findings with 3D perspective volume-rendered compared with 2D multiplanar reformation images. **(a)** Transverse 2D multiplanar reformation image of polyp (arrow) with flat morphology along the colon wall. **(b)** Coronal 2D multiplanar reformation image shows similar appearance of the polyp (arrow) and fold (arrowhead). **(c)** Endoscopic view of 3D perspective volume-rendered image of the polyp (arrow) clearly depicts focal polypoid morphology. **(d)** Endoscopic view of 3D perspective volume-rendered image shows the adjacent fold (arrowheads) stretching across the colon wall.



a. b.

Figure 5. **(a)** Transverse 2D multiplanar reformation CT image and **(b)** endoscopic view of 3D perspective volume-rendered CT image depict a 5-mm rectal hemorrhoid (arrow) that was a false-positive finding by all readers.

TABLE 4
Intraobserver Agreement for Three Image-Display Techniques

Image-Display Technique	κ Value*	r Value†
2D MPR		
Reader 1	0.80 (0.11)	0.89
Reader 2	0.93 (0.07)	0.91
Reader 3	0.73 (0.12)	0.83
3D MPR		
Reader 1	1.00	0.97
Reader 2	1.00	0.98
Reader 3	0.80 (0.19)	0.82
3D PVR		
Reader 1	0.60 (0.25)	0.82
Reader 2	0.60 (0.23)	0.86
Reader 3	1.00	0.96

Note.—For 2D MPR, each reader reevaluated all 30 segments. For the additional 3D displays, each reader reevaluated the same 10 segments from reading one, but in a new randomized order. MPR = multiplanar reformation image, PVR = perspective volume-rendered image.

* Data in parentheses are the SEM.

† Determined with the Spearman rank order correlation.

TABLE 5
Interobserver Agreement among Readers at 2D Multiplanar Reformation

Readers	κ Value*	r Value†
Reading 1		
1 and 2	0.68 (0.13)	0.86
1 and 3	0.53 (0.16)	0.64
2 and 3	0.61 (0.14)	0.74
Reading 2		
1 and 2	0.80 (0.11)	0.81
1 and 3	0.73 (0.12)	0.83
2 and 3	0.66 (0.14)	0.76

Note.—Agreement was analyzed for each colonic segment ($n = 30$).

* Data in parentheses are the SEM.

† Determined with the Spearman rank order correlation.

multiplanar reformation, 0.80–1.00 for 3D multiplanar reformation, and 0.60–1.00 for 3D perspective volume-rendered images. Of note, since κ is used to remove the number of agreements expected by chance and calculate the remaining proportion of agreement, the worst case of κ of 0.60 actually represents a total agreement in eight of 10 cases. Spearman rank order correlation coefficients, based on the individual rated responses, were 0.83–0.91 for 2D multiplanar reformation, 0.82–0.98 for 3D multiplanar reformation, and 0.82–0.96 for 3D perspective volume-rendered images.

When interviewed at the end of reading two regarding recall of the cases, one reader had recall of some findings in four cases; however, he could not recall how he rated the findings or the other colonic findings. The other two readers stated they had no or minimal recall.

Interobserver agreement.—Table 5 shows the interobserver agreement for 2D multiplanar reformation, as analyzed according to colonic segment for reading one and reading two. The 2D multiplanar reformation was used to determine interobserver agreement, since this display technique was evaluated in all 30 segments by all readers, unlike the additional 3D display techniques that were subdivided among the readers. Overall, there was a trend for improved interobserver agreement at reading two, with κ scores of 0.53–0.68 at reading one and increased κ scores of 0.66–0.80 at reading two. The findings were similar with the Spearman rank order correlation coefficients, which were 0.64–0.86 at reading one and which increased to 0.76–0.83 at reading two.

Timed Reader Sessions and Image Quality Evaluations

All three readers improved their timed sessions of the colonic segments from reading one to reading two for each image-display technique. The pooled 2D multiplanar reformation timed session was 9.2 minutes for reading one (individual ranges, 8.3–10.7 minutes) and decreased to 7.0 minutes for reading two (individual ranges, 6.2–7.8 minutes). The incremental timed sessions for the 3D display techniques following 2D multiplanar reformation also showed overall decreases. The 3D multiplanar reformation times were the shortest and had the least variation, which is likely because of the similarity to the 2D multiplanar reformation views that directly preceded them. The pooled 3D multiplanar reformation timed session was 7.1 minutes for reading one (individual ranges, 6.8–7.5 minutes) and 5.6 minutes for reading two (individual ranges, 5.4–5.9 minutes). The endoscopic 3D perspective volume-rendered timed sessions were the longest and showed the most improvement in time efficiency owing to the more demanding endoscopic navigational demands. The pooled 3D perspective volume-rendered timed session was 15.4 minutes for reading one (individual ranges, 10.8–18.4 minutes), which decreased to 11.1 minutes for reading two (individual ranges, 10.2–12.4 minutes).

The pooled reader evaluations of the image quality of the colonic segments with 2D multiplanar reformation during reading one demonstrated the following distributions for the four parameters. For luminal distention, 58% of segments were excellent, 31% were good, 11% were fair, and no segments were poor. For the degree of fluid degradation, 53% of segments had none, 29% had mild degradation, 18% had moderate degradation, and no segments had severe degradation. For tortuosity, 20% of segments were evaluated to have none, 48% were minimal, 24% were moderate, and 8% were marked. For overall depiction, 41% of the segments were excellent, 42% were good, 17% were fair, and no segments were poor.

Results of the analysis of variance with use of the four image quality parameters as the independent variables and the timed reader sessions as the dependent variables demonstrated that the image quality parameters had the strongest effect on reading times for the endoscopic 3D perspective volume-rendered view. Specifically, increased reading times for 3D perspective volume-rendered images were associated with decreased overall depiction ($P = .005$). There were trends of increased reader sessions as tortuosity increased with 3D perspective volume-rendered ($P = .08$) and 3D multiplanar reformation images ($P = .06$). There was no association between any image quality parameter and reading times for 2D multiplanar reformation images.

DISCUSSION

With the rapid advancement of image-display techniques for spiral CT colonography, well-controlled evaluations of reader performance are needed to optimize reader protocols for larger prospective series. In this study, an image library of clinically relevant colorectal lesions in 30 colonic segments was constructed to uniquely evaluate both diagnostic performance and reader agreement among 2D multiplanar reformation and specific 3D image-display techniques. Overall similar results in diagnostic performance were seen among the 2D multiplanar reformation and the 3D display techniques for reading one and reading two. Pooled results according to polyp size category for sensitivity and positive predictive value among all three display techniques were 77%–86% and 74%–86% (all polyps, $n = 22$), 91%–100% and 85%–100% (polyps or cancers ≥10 mm, $n = 11$), and 61%–73% and 61%–80% (polyps, 5–9 mm; $n = 11$), respectively. Although no statistically significant differences were seen in diagnostic performance with the additional 3D thick-slab multiplanar reformation or 3D endoscopic perspective volume-rendered images, a few individual cases showed improved detection and characterization with the 3D views. Evaluation of reader agreement among the three experienced abdominal readers demonstrated that overall intraobserver agreement between reading one and reading two was good, with a broader range of interobserver variability. Learning effects were most pronounced for the less familiar 3D display techniques with respect to both time required for navigation and correct recognition of 3D morphologies.

Results of prospective studies (6–16) of spiral CT colonography have shown a wide range of diagnostic performance, with differences present in influential variables such as patient cohort, case mix of lesions, types of image-display techniques, and reader protocol. Overall, the sensitivity for detection of polyps 10 mm and larger was 50%–90% (15,16). The number of polyps evaluated in this size category ranged from six to 22 and were reported in patient cohorts of 44–100 (6–16). Authors of these studies have reported different uses of 2D and 3D display techniques, including evaluation of transverse 2D sections alone, 2D multiplanar reformation combined with various 3D endoscopic display techniques, and digital 3D movies.

The results of two recent studies demonstrate the overall range of reader results to date. Fenlon et al (16) examined 100 patients by using a combined evaluation of 2D transverse images and 3D endoscopic antegrade and retrograde digital movies. This study reported sensitivities per polyp of 91% for polyps 10 mm or larger (n = 22) and 82% for 6–9-mm polyps (n = 40). Positive predictive values ranged from 75% for polyps 5 mm or smaller to 91% for polyps 10 mm or larger. These results contrast with those of another recent report by Rex et al (15), in which 46 patients were examined by using combined 2D multiplanar reformation and 3D display techniques. The sensitivities per polyp were 50% for polyps 1 cm or larger (n = 14) and 43% for 6–9-mm polyps (n = 14), with a specificity of 89% for patients with polyps 1 cm or larger.

Although diagnostic performance in an image library of selected colonic segments may not generalize to prospective evaluation of complete colons, comparisons can be made with careful attention to the CT reader protocol. Our reader protocol of the various 2D and 3D image-display techniques differed from those in prior investigations in several ways. Although we discriminated between 2D and 3D performance, our evaluation of the 2D display technique did not include the transverse 2D images alone. Rather, our 2D evaluation involved the interactive viewing of transverse, sagittal, and coronal 2D multiplanar reformation images as the initial evaluation. Our rationale was that, given the limitation in the number of display techniques that could be feasibly tested within a reader study, the 2D multiplanar reformation as the primary evaluation was a ubiquitously available simple display technique, which was more concordant with the increasing capabilities of thin-section volumetric CT data acquisition.

Our pooled results of sensitivity for 2D multiplanar reformation alone, stratified to polyp size, were 94% for polyps 10 mm or larger (n = 11) and 61%–70% for 5–9-mm polyps. Results of prior investigations (7,15) have shown poorer performances with transverse 2D sections alone than with combinations of 2D multiplanar reformation and 3D display techniques. In one series (7) of 70 patients, sensitivities for polyps 10 mm or larger (n = 15) ranged from 47% with transverse 2D images to 67%–73% with 3D evaluations. In another series (15) of 46 patients, sensitivities for polyps 10 mm or larger (n = 14) ranged from 29% with transverse 2D images to 50% with 3D renderings. The lack of statistically significant differences between performance with 2D and 3D images in our study may be attributable in part to the advantages of simultaneous viewing in multiple planes with 2D multiplanar reformation, compared with viewing 2D transverse images alone.

Our use of the 3D display techniques also differed from that of prior study designs. Our 3D display techniques were used both to further characterize any focal finding on 2D multiplanar reformation images and to detect any additional findings by means of evaluating of the entire segment with the 3D display techniques. Evaluation of the interactive, continuous antegrade, and retrograde depiction with 3D perspective volume-rendered images of the entire colonic segment has not been well studied to date. Our study represents a preliminary evaluation of a rapidly evolving technology within computer graphics and differs from the prior use of key frame, noninteractive 3D digital movies for endoscopic depiction.

Although no statistically different results between our 2D multiplanar reformation and 3D display techniques were present, individual cases showed examples of improved characterization and detection with the 3D views. In one case (Fig 3), an 8-mm polyp was seen behind a haustral fold only on 3D perspective volume-rendered images at reading two; however, it was not correctly classified as a polyp. Although this demonstrates the potential of improved depiction of surface area with 3D imaging to detect additional lesions, lack of reader familiarity with 3D morphologies and distraction with 3D navigation hindered performance.

The greatest value of the additional 3D display techniques was in improved characterization of false-positive findings on 2D multiplanar reformation images (Fig 4). Prior study (7,10) results have also documented the utility of 3D renderings to help characterize nodular folds from polyps. False-positive findings are a major concern in the evaluation of new screening tests. The potential of the 3D views to decrease the false-positive rate of nodular folds may have important implications, if CT colonography is to be used as a screening algorithm in selected cohorts.

Although the development and evaluation of the 3D image-display techniques are still evolving, the time currently required to read and interpret these images is variable. Our timed reader sessions of 3D perspective volume-rendered images had the greatest range in times and showed the greatest improvement in efficiency with reading two. Dachman et al (10) reported timed reading sessions for entire colons with the use of 2D transverse sections combined with 3D endoscopic views in 44 patients with two readers. The mean time spent on interpretation was 28.5 minutes (range, 14–65 minutes), with a small increase in reader times with greater use of the 3D endoscopic view. In that study, 3D was used selectively to problem solve focal findings at 2D review, unlike the additional continuous navigation throughout the entire segment that we report. Further evaluations of the differences in diagnostic performance and timed sessions with the advancing image-display techniques will be needed.

Reader agreement at spiral CT colonography has not been well documented. In our investigation, the diagnostic performance of three readers was evaluated after an initial training period. Our three readers were new to spiral CT colonography, and their training sessions represented their primary experience before beginning the formal reader protocol. After reading one of the protocol, reading two was performed 6 weeks later, with rerandomization of the order of the cases. Interobserver agreement with 2D multiplanar reformation, as analyzed according to colonic segment, demonstrated a moderate to good range of κ scores. As expected, intraobserver agreement between reading one and reading two was better than interobserver agreement. Each reader demonstrated good agreement for all three image-display techniques, with the largest range seen with 3D perspective volume-rendered images.

Only two of the four published prospective series (7,10) have reported results of two independent readers, with no analysis of intra- or interobserver variability. Hara et al (7) reported on 70 patients in 1997, with 115 polyps evaluated independently by two radiologists. Sensitivity for the two readers for the 3D evaluation was 67% and 73% for polyps 10 mm or larger (n = 15), 56% and 69% for 5–9-mm polyps (n = 21), and 25% and 27% for polyps smaller than 5 mm (n = 79), respectively. Dachman et al (10) reported on 44 patients with 22 polyps independently evaluated by two radiologists. In this investigation (10), transverse 2D source images were used in combination with 3D shaded-surface display for problem solving; the results of sensitivity were nearly identical for the two readers, with sensitivities of 83% for polyps 8 mm or larger (n = 6), 33% for 5–8-mm polyps (n = 3), and 0% and 15%, respectively, for polyps smaller than 5 mm (n = 13). Further evaluation of reader agreement among less experienced readers in prospective series will be important in future studies.

There were several important factors that may have influenced the results of this study. The selection bias involved in forming an image library of colonic segments precludes generalization to prospective series. However, such a study design permits efficient testing of diagnostic performance and learning effects for different image-display techniques. There is a more positive case mix of polyps than would be encountered in general screening. Authors of prior reports (19,20) have estimated that 10%–20% of patients who undergo routine screening have polyps 1 cm or greater. In this study, nine (30%) of 30 segments had lesions of at least 1 cm. This spectrum bias of an increased proportion of positive cases may alter diagnostic performance (21).

The use of colonic segments rather than complete colons also influences reader evaluations; however, our design facilitated full evaluation with the 3D display techniques. In our investigation, a range of colonic segments was used, with varying image quality. In a prior investigation by Chen et al (13), luminal distention was quantified with prone and supine spiral CT colonographic acquisitions in 23 patients. In this study (13), the combined scores of luminal distention for the prone and supine colonic segments were collapsed in 4%, poor in 1%, good in 17%, and optimal in 77%. Our results of luminal distention in the 30 colonic segments were proportionally similar, with no segments with poor distention, 11% with fair distention, 31% with good distention, and 58% with excellent distention.

Although our evaluation of interobserver agreement among experienced abdominal radiologists cannot be generalized to more novice readers, this evaluation represents the first step in early technology assessment and allows design of future studies to involve less specialized readers. Recall bias can influence evaluation of intraobserver agreement. However, in this study we believe no substantial recall was present.

Finally, it is difficult to assess reader performance with 3D endoscopic depiction, given the rapidly evolving hardware and software advances in 3D imaging. The 3D perspective volume-rendered technique is undergoing constant improvement, with advances in autonavigation, extended field of view depictions, and improved cross-registration between 2D and 3D. Thus, these results represent the diagnostic performance of 3D imaging at a given point in evolution of the technology, in addition to the early experience of reader familiarity with 3D imagery.

In summary, the results of this investigation demonstrate that spiral CT colonography is accurate for detection and characterization of colorectal polyps in an image library of colonic segments, with good intraobserver agreement and variable interobserver agreement. Individual cases showed added value of improved characterization and detection with the 3D image displays, with learning effects noted. Training in the systematic evaluation of the entire colon for accurate recognition of 2D and 3D morphologies will be essential in the development of new readers for spiral CT colonography. Future evaluations of spiral CT colonography need to integrate new advances, such as improved image acquisition and processing, into well-designed reader studies with more generalizable study cohorts and less experienced readers.

References

1. Johnson CD, Hara AK, Reed JE. Computed tomographic colonography (virtual colonoscopy): a new method for detecting colorectal neoplasms. Endoscopy 1997; 29:454–461.
2. Kay CL, Evangelou HA. A review of the technical and clinical aspects of virtual endoscopy. Endoscopy 1996; 28:768–775.
3. Rubin GD, Beaulieu CF, Argiro V, et al. Perspective volume rendering of CT and MR images: applications for endoscopic imaging. Radiology 1996; 199:321–330.
4. Royster AP, Gupta AK, Fenlon HM, Ferrucci JT. Virtual colonoscopy: current status and future implications. Acad Radiol 1998; 5:282–288.
5. McFarland EG, Brink JA. Spiral CT colonography (virtual colonoscopy): challenge between advancing technology and generalizability. AJR Am J Roentgenol 1999; 173:549–559.
6. Hara AK, Johnson CD, Reed JE, Ehman RL, Ilstrup DM. Colorectal polyp detection with CT colography: two- versus three-dimensional techniques—work in progress. Radiology 1996; 200:49–54.
7. Hara AK, Johnson CD, Reed JE, et al. Detection of colorectal polys with CT colography: initial assessment of sensitivity and specificity. Radiology 1997; 205:59–65.
8. Royster AP, Fenlon HM, Clarke PD, Nunes DP, Ferrucci JT. CT colonoscopy of colorectal neoplasms: two-dimensional and three-dimensional virtual-reality techniques with colonoscopic correlation. AJR Am J Roentgenol 1997; 169:1237–1242.
9. Harvey CJ, Amin Z, Hare CMB, et al. Helical CT pneumocolon to assess colonic tumors: radiologic-pathologic correlation. AJR Am J Roentgenol 1998; 170:1439–1443.
10. Dachman AH, Kuniyoshi JK, Boyle CM, et al. CT colonography with three-dimensional problem solving for detection of colonic polyps. AJR Am J Roentgenol 1998; 171:989–995.
11. Beaulieu CF, Jeffrey RB Jr, Karadi C, Paik DS, Napel S. Display modes for CT colonography. II. Blinded comparison of axial CT and virtual endoscopic and panoramic endoscopic volume-rendered studies. Radiology 1999; 212:203–212.
12. Fenlon HM, McAneny DB, Nunes DP, Clarke PD, Ferrucci JT. Occlusive colon carcinoma: virtual colonoscopy in the preoperative evaluation of the proximal colon. Radiology 1999; 210:423–428.
13. Chen SC, Lu DSK, Hecht JR, Kadell BM. CT colonography: value of scanning in both the supine and prone positions. AJR Am J Roentgenol 1999; 172:595–599.
14. Yee J, Hung RK, Akerkar GA, Wall SD. The usefulness of glucagon hydrochloride for colonic distention in CT colonography. AJR Am J Roentgenol 1999; 173:169–172.
15. Rex DK, Vining D, Kopecky KK. An initial experience with screening for colon polyps using spiral CT with and without CT colography (virtual colonoscopy). Gastrointest Endosc 1999; 50:309–313.
16. Fenlon HM, Nunes DP, Schroy PC III, et al. A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. N Engl J Med 1999; 341:1496–1503.
17. McFarland EG, Brink JA, Heiken JP, et al. Spiral CT colography (virtual colonoscopy): multiobserver study of different 2D and 3D image display techniques compared to colonoscopy. Proc SPIE 1999; 3660:106–108.
18. McFarland EG, Brink JA, Loh J, et al. Visualization of colorectal polyps with spiral CT colography: evaluation of processing parameters with perspective volume rendering. Radiology 1997; 205:701–707.
19. Rex DK. Colonoscopy: a review of its yield for cancers and adenomas by indication. Am J Gastroenterol 1995; 90:353–365.
20. Lieberman DA. Cost-effectiveness model for colon cancer screening. Gastroenterology 1995; 109:1781–1790.
21. Ransohoff DF, Feinstein AR. Problems of spectrum bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299:926–930.