Published in final edited form as:

Acad Radiol. 2010 August ; 17(8): 948–959. doi:10.1016/j.acra.2010.03.024.

CT Colonography Computer-Aided Polyp Detection: Effect on Radiologist Observers of Polyp Identification by CAD on Both the Supine and Prone Scans

Ronald M. Summers, MD, PhD¹, Jiamin Liu, PhD¹, Bhavya Rehani, MD¹, Phillip Stafford, BA¹, Linda Brown, MD¹, Adeline Louie, MD¹, Duncan S. Barlow, M.D², Donald W. Jensen, M.D², Brooks Cash, MD², J. Richard Choi, MD, ScD³, Perry J. Pickhardt, MD⁴, and Nicholas Petrick, Ph.D⁵

¹ Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892-1182

² Colon Health Initiative, National Naval Medical Center, 8901 Wisconsin Avenue, Bldg 9, Bethesda, MD 20889

³ Walter Reed Army Medical Center, Washington, DC 20307

⁴ Department of Radiology, University of Wisconsin Medical School, E3/311 Clinical Science Center, 600 Highland Ave., Madison, WI 53792

⁵ Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10903, New Hampshire Avenue, WO62-4118, Silver Spring, MD 20993-0002

Abstract

Corresponding Author and Reprint Requests: Ronald M. Summers, M.D., Ph.D., Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bldg. 10, Room 1C368X MSC 1182, BETHESDA MD 20892-1182, Phone: (301) 402-5486, FAX: (301) 451-5721, rms@nih.gov, Web: <http://www.cc.nih.gov/drd/summers.html>.

Presented in part at the 2008 meeting of CARS, Barcelona, Spain.

Author contributions.

guarantor of integrity of entire study - RMS
study concepts - RMS
study design - RMS
literature research - RMS
clinical studies - JRC, PJP
data acquisition - PS, LB, AL, DB, DJ, BR, PS
data analysis - JL, BR, NP
statistical analysis - JL, BR, RMS, NP
manuscript preparation - RMS, JL, BR
manuscript definition of intellectual content - RMS
manuscript editing - RMS, BR, BC, NP
manuscript final version approval - all authors

Potential financial interest.

Author Summers has pending and/or awarded patents for the subject matter described in the manuscript and receives royalty income for a patent license from iCAD. His lab is supported in part by a Cooperative Research and Development Agreement with iCAD. Viatronix supplied the V3D Colon software to NIH free of charge. Author Pickhardt is on the medical advisory boards of Viatronix, Inc. and Medicsight, Inc., a consultant to Covidien and co-founder of VirtuoCTC. Author Choi is on the medical advisory boards of Viatronix, Inc and QI and has received research support from E-Z-EM.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

PURPOSE—To determine whether the display of computer-aided detection (CAD) marks on individual polyps on both the supine and prone scans leads to improved polyp detection by radiologists compared to the display of CAD marks on individual polyps on either the supine or the prone scan, but not both.

METHOD AND MATERIALS—The acquisition of patient data for this study was approved by the Institutional Review Board and was HIPAA-compliant. Subsequently, the use of the data was declared exempt from further IRB review. Four radiologists interpreted 33 CT colonography cases, 21 of which had one adenoma 6 to 9 mm in size, with the assistance of a CAD system in the first reader mode, i.e., the radiologists reviewed only the CAD marks. The radiologists were shown each case twice, with different sets of CAD marks for each of the two readings. In one reading a true positive CAD mark for the same polyp was displayed on both the supine and prone scans (a double-mark reading). In the other reading a true positive CAD mark was displayed either on the supine or prone scan but not both (a single-mark reading). True positive marks were randomized between readings and there was at least a one-month delay between readings to minimize recall bias. Sensitivity and specificity were determined and receiver operating characteristic (ROC) and multiple-reader multiple-case analyses were performed.

RESULTS—The average per polyp sensitivities were 60% [38%, 81%] vs. 71% [52%, 91%] ($p = .03$) for single-mark and double mark readings, respectively. The areas [95% confidence intervals] under the ROC curves were 0.76 [0.62, 0.88] and 0.79 [0.58, 0.96], respectively ($p = \text{NS}$). Specificities were similar for the single-mark compared to the double-mark readings.

CONCLUSION—The display of CAD marks on a polyp on both the supine and prone scans led to more frequent detection of polyps by radiologists without adversely affecting specificity for detecting 6–9 mm adenomas.

Keywords

CT, colon; CT, virtual imaging; Colon cancer; image processing; automated detection; observer performance

CT colonography (CTC) computed-aided detection (CAD) of polyps has advanced considerably over the past decade [1]. Several recent studies have found that radiologists' performance at polyp detection improves significantly with the aid of CAD [2–6]. There is the prospect that in the near future, both CTC and CAD will be widely used for colorectal cancer screening.

The successful clinical implementation of CAD depends on a number of factors including both technical and perceptual factors. The CAD software must locate the polyp and the radiologist must correctly interpret the CAD finding as a polyp. Previous studies have found that radiologists occasionally ignore true positive CAD findings [2–6]. Such behavior undermines the potential benefit of CAD. A better understanding of the causes of such errors could lead to improved radiologist performance.

One important aspect of the radiologist's use of CAD for CTC is whether individual polyps can be found on both the supine and prone examinations [7–9]. Radiologists are trained to search for the polyp on both examinations to increase their confidence that the finding represents a true polyp and not residual fecal matter. For example, residual fecal matter changes position but a non-pedunculated colonic polyp should not.

Current CAD systems analyze the supine and prone CTC scans independently. The results of the two independent CAD analyses are then combined by the radiologist or CAD evaluator after considering the size, location and morphology of the CAD marks, so that a polyp is considered to be "detected" if it is marked by CAD on either or both scans. Detection of the

polyp by CAD on both supine and prone examinations is typically not a requirement during development and validation of the CAD software and has not been an explicit priority of CAD developers although polyps are likely to be detected on both scans if they are sufficiently conspicuous to the CAD algorithm. Polyps may be missed by CAD on one of the two scans for a variety of reasons, including image noise, inadequate colonic distention, poorly-tagged fluid or stool, and changes in polyp shape (such as a flatter appearance).

The purpose of this observer performance study was to determine, in a well-controlled randomized experiment, whether radiologists detected polyps more frequently when they were shown CAD marks on both the supine and prone scans rather than just on one scan. Our study focused on the more challenging 6 to 9 mm polyps for which CAD has been shown to provide a benefit in observer studies [2,5].

MATERIALS AND METHODS

The acquisition of patient data for this study was approved by the Institutional Review Board and was HIPAA-compliant. Subsequently, the use of the data was declared exempt from further IRB review. Viatronix V3D Colon software (Stony Brook, NY) was used for part of this project and was supplied free of charge to the investigators. Authors who were not Viatronix board members had full control of the data.

Patient population

Patients were a subset of those enrolled in previous CTC and CAD trials of consecutive screening patients [10,11]. Inclusion and exclusion criteria are shown in Figure 1 and will be described in more detail below. Patients were selected from a previously defined training set [11]. The ratio of abnormal to normal patients was chosen to be approximately 2:1. All abnormal patients from the training set that met the criteria to be described were included in the study (Figure 1). There were 20 men and 13 women ranging in age from 47 to 76 years (mean 59.4 years).

Bowel Preparation

Patients underwent a 24-hour colonic preparation that consisted of oral administration of 90 ml sodium phosphate (Fleet 1 preparation, Fleet Pharmaceuticals), 10 mg bisacodyl, 500 ml of barium (2.1% by weight; Scan C, Lafayette Pharmaceuticals) and 120 ml of diatrizoate meglumine and diatrizoate sodium (Gastrografin, Bracco Diagnostics) given in divided doses.

CT Scanning

The colon was distended with patient-controlled insufflation of room air. CT scanning occurred during one breathhold in each of the prone and supine positions using a four-channel or eight-channel CT scanner (General Electric LightSpeed or LightSpeed Ultra). CT scanning parameters included 1.25 – 2.5 mm section collimation, 15 mm per second table speed, 1 mm reconstruction interval, 100 mAs and 120 kVp.

Optical Colonoscopy

Patients underwent same-day optical colonoscopy by one of 17 colonoscopists. The colonoscopies were performed using segmental unblinding, wherein CTC results were revealed to the colonoscopists during the examination to create an enhanced reference standard. Polyp sizes were determined at optical colonoscopy using a calibrated guidewire.

Polyp Identification

Ground truth was established manually by the following method. Each polyp seen by optical colonoscopy was located on the prone and supine CTC images using Viatronix V3D Colon software (Stony Brook, NY). To match a polyp on optical colonoscopy and CTC, the polyp had to be located within the same or adjacent colonic segment and measure within 50% by size. Using a graphical user interface, a voxel within each polyp was marked manually to enable later review and then the polyp's borders were traced on each CT slice containing the polyp. The markers and tracings were placed by one of four trained research assistants under the supervision of a board-certified radiologist with experience with over 400 proven positive research CTC examinations (blinded).

CAD System

The CAD system has been previously described [11–14]. It identifies the colon lumen and wall, electronically subtracts the contrast-enhanced colonic fluid, calculates the colonic surface features, segments the potential polyps to determine their 3-dimensional boundaries, and classifies the potential polyps as true or false detections according to a set threshold. It outputs the locations of the polyp candidates in the CTC images. If any voxel within a polyp candidate matched those within a traced reference standard polyp, then the polyp candidate was labeled a true positive; otherwise, it was labeled a false positive.

For display to radiologists, the polyp candidates determined by CAD were loaded into Viatronix V3D Colon software. The radiologists could select each candidate from a list to see the candidate which was colored blue in the three-dimensional endoluminal display and indicated by a rectangle in the two-dimensional transverse, sagittal and coronal displays. The Viatronix software has electronic fluid subtraction capability, which was used by the radiologists according to their individual preferences.

To prepare the set of true and false positive detections in sufficient number and characteristics for this study, the CAD system's classifier (false positive reducer) was deactivated, effectively increasing both the sensitivity and false positive rates. The CAD scores reported by the classifier, however, were recorded for later use in selecting false positive candidates, to be described below.

Overview of study design

Each CTC scan (supine or prone) shown to the radiologist observers was prepared such that there were 4 computer-aided detections per scan for a total of eight detections per patient. The purpose of fixing the number of detections at four was to standardize the protocol and eliminate the potential variability due to differing numbers of false positives per patient.

There was at most one true positive CAD mark shown for each scan. Two sets of computer-aided detections were prepared for each patient: a "single-mark" set and a "double-mark" set. For the single-mark data set and readings, there was one true positive CAD mark on either the supine or the prone scan but not both, even though the polyp was detectable by CAD on both scans (the CAD mark on the other scan was not revealed to the radiologists). For the double-mark data set and readings, there was one true positive CAD mark of the same polyp on each of the supine and prone scans (i.e., the true positive CAD marks on both scans were revealed to the radiologists).

Selection of abnormal patients and detections—The CTC cases were selected according to the following rules (Figure 1). All abnormal patients had to have at least one adenomatous polyp in the 6 to 9 mm size range. Their supine and prone CTC scans had to have at least four false positive computer-aided detections each (8 false positives per patient).

Adenomas were selected according to the following rules. Each adenoma had to be retrospectively identifiable (by trained research assistants supervised by the board-certified radiologist) on both the supine and prone CTC scans. In addition, each adenoma had to have been detected by the CAD software on both the supine and prone scans. If a patient had more than one adenomatous polyp that met these criteria, then one of their adenomas was chosen randomly; other adenoma(s) were not included in the eight detections shown to the radiologists (only 1 patient met that criterion). Because of these rules, the CAD system had an effective sensitivity of 100% for these patients. There were six 6 mm, three 7 mm, ten 8 mm, and two 9 mm adenomas.

Selection of false positive detections

Single-mark data sets had seven false positives per data set (four on one scan and three on the other) and double-mark data sets had six false positive marks per data set (three on each scan). False positives on the ileocecal valve were discarded prior to selection of the false positives used in this study. To make the observers' task more challenging, we chose false positives matched in size with the target 6 to 9 mm lesions and thought by CAD to be more polyp-like. To do so, an effective diameter was computed from the segmented volume of the false positive detection assuming a hemispherical model. False positives with an effective diameter between 6 and 9 mm, inclusive, were considered candidate false positives for this experiment. The candidate false positives were ranked by their CAD score (from the classifier stage of the CAD software; a relative measure of how "polyp-like" a particular detection appears to the CAD software) and the top ranked detections were added to the detection list until there were four detections for that particular scan. If there were ties amongst the CAD scores, false positive candidates were selected randomly and added to the detection list. One hyperplastic polyp was included in the list of false positives inadvertently since it was not found during preparation of the ground truth.

Selection of normal cases

Normal cases were selected from those patients having no optical colonoscopy confirmed lesions. From these cases, those patients having at least four false positives on each of the supine and prone scans were identified and 12 patients were randomly chosen from this subset. False positives were selected according to the rules in the preceding section.

To make the radiologists' interpretive task more challenging, one of the false positives in nine of the 12 normal cases were matched with a co-located false positive on the alternate scan. For this matching, co-location was defined as being within ± 5 cm along the length of the colon centerline. The CAD score of each co-located pair of false positives were summed and then the pair having the highest combined CAD score was included within the list of false positives shown to the radiologists for the case.

Randomization

Each reader read each case twice but the reading order (single-mark or double-mark reading) was randomized. There was an interval of at least one month between each of the two reading sessions for a particular patient by a given reader to minimize recall bias. For the single-mark reading, the single true positive CAD mark was randomly assigned to either the supine or prone scan; for a particular patient, the assigned scan for the true-positive CAD mark (supine or prone) was the same for all four readers. The position of the true positive detection in the list of four detections presented to the reader for a particular scan was also randomized. For readers one and two, 5 normal and 2 polyp cases were re-read because one or more detections were outside the colon; this was fixed before readers three and four began interpreting. For readers three and four in the second read only, the polyp cases were read as a block due to an error in data recording.

Reader experience

Reader 1 had read over 500 CTC cases and reader 2 had read over 100 CTC cases prior to this study. Readers 3 and 4 had each read over 2000 clinical CTC cases. In the results and discussion, Readers 1 and 2 are considered the less-experienced set of readers. All four readers had prior experience with the Viatronix interpretation software. The software provides for a primary three-dimensional image interpretation with two-dimensional images for problem-solving.

Training

To familiarize the radiologists with the data recording software (written in Visual Basic, Microsoft), the radiologists reviewed five cases that were not used as part of the study.

Instructions given to the radiologist observers

The radiologist observers were given a sheet of instructions to follow. Radiologists were told that polyps might be marked by CAD on one or both supine and prone scans. They were to determine which CAD marks were on polyps. The radiologists were not told that sometimes the true positive CAD mark would be withheld from them on one of the two CTC scans. The radiologists were told to only review the CAD marks (“first-read” paradigm) and not to do a complete interpretation of the scan. However, to fully evaluate a CAD mark on one scan, they were to try to find a corresponding co-located abnormality on the alternate scan using axial, multiplanar reformatted, or interactive three-dimensional endoluminal fly-through images. They could also use the translucent display and make size measurements on either 2-D or 3-D images. They were told that all detections presented in this study were determined by the CAD system to have diameters between six and 9.5 mm. For each polyp, they were to record whether the polyp was visible on both scans. They were also to record their level of confidence on a scale of 0–100 that the finding was a polyp (0=definitely not a polyp; 100=definitely a polyp). For each detection that the radiologist considered not to be a polyp (confidence=0), they were to provide a reason for rejecting the detection. After reviewing supine and prone scans and reaching a decision, radiologists entered their findings for each of the eight detections per patient into a Microsoft Access database using the data recording software. Multiple reading sessions were required by each of the radiologists to complete each of the readings. Total reading times were recorded for each reading of each patient dataset (supine and prone).

A radiologist false positive diagnosis occurred when the radiologist gave a confidence score ≥ 1 for a CAD false positive detection. A radiologist true positive diagnosis occurred when the radiologist gave a confidence score ≥ 1 for a CAD true positive detection. A radiologist false negative diagnosis occurred when the radiologist gave a confidence score of 0 for a CAD true positive detection. A radiologist true negative diagnosis occurred when the radiologist gave a confidence score of 0 for a CAD false positive detection.

Statistical Analysis

The major outcome measures were the per polyp sensitivities and the area under the curves in the receiver operating characteristic (ROC) analyses for the single-mark versus double-mark experiments. Sensitivity was defined as the fraction of true positive polyps or patients assigned a confidence score ≥ 1 . A secondary outcome measure was the sensitivity for detecting individual polyps on both the supine and prone scans compared to detecting the polyps on either the supine or the prone scan but not both scans.

An ROC analysis was done using the confidence scores. If the radiologist diagnosed more than one “polyp” in a patient (even though there was at most one polyp marked per patient), the maximum confidence score for the patient was used for the computation of ROC for that patient and radiologist. If the radiologist gave different confidence scores for the same polyp on the

prone and supine double-mark scans, the maximum confidence score for the patient was used for the computation of ROC for that patient and radiologist.

A multi-reader multi-case (MRMC) bootstrap technique was used to determine confidence limits and statistical significance when reporting per-patient average reader sensitivity, specificity and the area under the curve (AUC) results [5,15]. This technique provides a nonparametric performance estimate that accounts for both reader and case variability, two major sources of error in radiological reader studies. Individual reader per-patient sensitivity and specificity were calculated with the bootstrap method assuming the reader as a fixed effect, thereby resampling only cases. Specificity was defined as the fraction of normal patients assigned a confidence score of 0.

We reported significant differences in performance for each radiologist and the average reader when comparing the single-mark versus double-mark experiments. A P-value of less than .05 indicated a significant difference. For the average reader, two-sided P values were estimated using the percentage of bootstrap experiments with a difference in performance below zero multiplied by two. Reported per-patient confidence limits were the 2.5% and 97.5% limits obtained from the bootstrap histogram. All performance estimates were based on 2000 bootstrap sample sets.

Three-factor ANOVA with modality and readers as fixed effects and cases as random effects which included both one and two way interactions was used to determine confidence limits and statistical significance when reporting per-polyp average reader sensitivity results (Matlab 7.9.0.529 (R2009b)). The Fisher exact test was used to compare per-polyp sensitivities for an individual reader [16]. Cochran's Q test was used to compare per polyp sensitivities across the readers for either the single or double mark readings [17].

P less than 0.05 defined statistical significance. The radiologists' reasons given for not calling a polyp (radiologist false negative interpretations) were tabulated.

RESULTS

The sensitivities per polyp were greater in the double-mark readings compared to the single-mark readings for three of four radiologists (Table 1). The 11% increase in per polyp sensitivity for the average reader was statistically significant ($p=.03$). The differences for individual readers were not statistically significant. The per polyp sensitivities of the readers for single mark reads were not statistically significantly different but those for double mark reads were statistically significantly different ($p=.002$). For polyps found by an individual reader on both single- and double-mark readings, the reader's confidence increased more often than it decreased for three of the four readers. An example of a polyp found by readers on the double-mark but not on the single-mark reading is shown in Figure 2.

The sensitivities per patient were greater in the double-mark readings compared to the single-mark readings for three of four radiologists (Table 2). Specificities were greater for two of the four radiologists in double-mark reading but the differences for individual readers and for the average reader were not statistically significant (Table 2).

When a single-mark reading was shown, the radiologists found the polyp on both scans (including the one without the CAD true positive mark) in an average of 42% of patients (Table 3). When a double-mark reading was shown, the radiologists found the polyp on both scans in an average of 63% of patients ($p=0.002$) (Table 3). For individual readers, the increase in sensitivity was statistically significant for one of the two less experienced radiologists and trended towards significance for the other. Figure 3 shows a polyp that was found more frequently by radiologists in both scans when CAD marked it on both scans.

The reasons for radiologist false negatives are shown in Table 4. The most common reason for a false negative diagnosis was misinterpreting a polyp as stool. An example of a CAD true-positive missed by radiologists on both single-mark and double-mark presentations is shown in Figure 4.

There were in total 66 false positive radiologist findings. An example of a CAD-induced false-positive radiologist diagnosis is shown in Figure 5. A relatively small fraction (3.5%, 66/1860) of the CAD false positives induced a false-positive radiologist diagnosis. False-positive radiologist diagnoses were 29% (3.6% versus 2.8%) more common in the non-matched false positives compared to the matched false positives (for which CAD false positive marks in the same part of the colon were intentionally included to see if they would confuse the radiologists), but about half of the non-matched radiologist false positives were on two findings (Table 5).

Of these 66 false positives, the radiologists stated that 34 had a matching CAD finding on both scans. Of these 34, 16 were due to a single colonic finding that looked very polyp-like and was thought to represent an inverted diverticulum; all four radiologists thought this finding was a polyp. Nine false positive findings by all four radiologists were on a hyperplastic polyp. Of the remaining 9 false positives, 2 were thought in retrospect to be on stool, 5 were on folds in the same part of the colon and in 2 cases the radiologist matched a true positive on one scan to a false positive on the other scan.

Of the 66 false positives, the radiologists stated that the remaining 32 did not have a matching CAD finding on both scans. Of these 32, 5 were on a single finding that in retrospect looked like a polyp and was also found prospectively during the initial clinical trial; all four radiologists thought this finding was a polyp. 13 false positives were on folds, 10 of which were from Reader 2. Five were on tagged fluid or stool, 3 were on gas bubbles, 3 were on a hyperplastic polyp and 3 were on a single colonic finding that looked very polyp-like but not found prospectively during the initial clinical trial.

The AUCs for the average reader for the single-mark and double-mark readings were 0.76 [0.62, 0.88] and 0.79 [0.58, 0.96] ($p=NS$). There was no statistically significant improvement in AUC with double-mark detection for any reader.

DISCUSSION

In this study, readers found medium-sized adenomas 11% more often on average when the polyp was marked by CAD on both the supine and prone scans rather than on only one scan. This sensitivity increase is large and potentially clinically highly relevant, particularly if the findings translate to the concurrent or second reader paradigms. Therefore, this study motivates further research on this topic.

The sensitivities per polyp in our study varied greatly amongst the readers and the differences were statistically significantly different for the double-mark readings. Large variations amongst readers or low sensitivities for some readers using CAD for colonography have been reported in the literature [2,4,5,18]. Higher sensitivities (93–97%), however, have been reported by others [6]. Factors in the design of our study that may affect the sensitivity include the use of the first reader mode, relatively limited training, selection of more difficult medium-sized polyps, a high prevalence of polyp-like false positive CAD marks and the “laboratory effect” [19]. It is interesting to note that in a similar study design evaluating detection of lung nodules with a CAD system having hypothetical ideal performance of 100% sensitivity, radiologists also missed lesions [20].

The AUC's and specificities did not significantly change when double-mark CAD was used. The lack of statistical significance in the AUC's may in part be due to CAD marks on a

hyperplastic polyp, an inverted diverticulum, and a polyp-like finding that were considered false positives for the purposes of this study. The lack of a significant change in specificity suggests that any potential downside to double-mark CAD presentation is small.

The AUC benefit of CAD in Ref. [5] (0.03) is the same as the benefit of double-mark CAD; neither were significant. From the viewpoint of AUC, the double-mark CAD doubles the benefit (i.e., the diagnostic efficacy) of CAD. Therefore, there may be an additional benefit of double-mark CAD, but to prove this, a larger sample size will be required. Using the tables in Ref. [21], we estimate that the number of readers would need to be about 10 to show statistical significance in a future study using a similar number of cases.

CT colonography examinations are almost universally done using at least two scans, most frequently supine and prone scans, because of the proven benefit of scanning in two positions [7–9]. When both supine and prone scanning is performed, poorly distended segments in one position can become well-distended and residual fluid and stool can shift, enabling visualization of polyps undetectable in the other position. However, the effect on radiologist readings of the use of CAD that detects the polyp on only one versus both scans has not previously been investigated. Different CT colonography CAD systems differ in their ability to detect polyps on both scans versus one or no scans. A CAD system that has high per polyp sensitivity for detecting polyps on either the supine or prone scan may have either high or low per polyp sensitivity for detecting polyps on both the supine and prone scans. As typically reported in the literature, however, this distinction is typically neither emphasized nor reported. Instead, the performance of the CAD is usually determined by calculating how many polyps are detected on either the prone or the supine scan. Detection on both scans is usually not required. The withholding of CAD marks from the radiologists in this experiment was designed to simulate this distinction and investigate the importance of CAD marking polyps on both scans. Our results indicate that this distinction matters for clinical interpretation of CTC images and that there is a benefit to having CAD cue the radiologist to a potential polyp's location on both scans.

The reasons for the improvement in radiologist performance are unclear. Possibilities include a perceptual benefit, since the radiologist has in essence two chances to identify the polyp, and a reduction in fatigue since the effort required to find the polyp on the alternate view is reduced. The marking of a potential abnormality on multiple images of the same body part may have reinforcing benefits through repetition that facilitates the radiologist's visual perception of the abnormality and improves diagnostic confidence.

The coordination of radiologic information from multiple scans of a body part has been found advantageous for mammography CAD where multiple images of each breast are commonly obtained from different angles [22–25]. CTC CAD systems currently do not analyze the supine and prone examinations in a coordinated way although there has been some early work in this area [26–30].

Stool and normal mucosa were the most common reasons given for ignoring CAD true positive marks. Half of the radiologist false positives were attributable to a small number of CAD marks including marks on either plausible polyps (including an inverted diverticulum) or a nonadenomatous polyp. A substantial number of false positives due to a thickened haustral fold were reported by a single reader; focused training may address this issue.

The fraction of radiologist false positives induced by CAD marks that were intentionally matched was similar to the fraction that were not matched. This finding suggests that just because CAD marks are on an object in the corresponding place on two scans, the radiologist correctly ignores the CAD marks if the object does not look the same, i.e., marking things in the same general location in the colon does not seem to confuse the radiologists. This is

advantageous since CAD developers do not need to check for such concordances that might occur by chance.

We studied medium-sized (6–9 mm) polyps because they are more difficult to detect than large (>9 mm) polyps both with and without CAD (15). Nevertheless, the detection of medium-sized polyps is clinically important as they may require immediate polypectomy or surveillance according to C-RADS category 2 or 3 criteria [31].

We used a first-reader paradigm in a well-controlled randomized experiment to focus the readers on interpreting and matching the computer-aided detections and to avoid the potentially confounding effects of a complete colon inspection. We are not advocating that the first reader paradigm be used routinely for CT colonography interpretation.

The limitations of the study include a relatively small dataset, a small number of cases that had to be re-read, the display of a fixed number of CAD marks for each scan and the use of the first reader mode. We showed a fixed number of CAD marks for each scan to reduce any effect on the results due to variations in the number of CAD marks.

In conclusion, double-mark display of CAD findings led to higher average per polyp sensitivity without a statistically significant adverse effect on specificity for detecting 6–9 mm adenomas.

Acknowledgments

We thank Andrew Dwyer, MD, for critical review of the manuscript. We thank Dr. William Schindler for supplying CT colonography data. Brandon Gallas and Frank W. Samuelson are thanked for assistance with the statistical methodology. Viatronix supplied the V3D Colon software free of charge. This work was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center and the National Institute of Biomedical Imaging and Bioengineering (NP). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government and no official endorsement of any equipment or product of any company mentioned in the publication should be inferred.

References

1. Yoshida H, Dachman AH. Computer-aided diagnosis for CT colonography. *Semin Ultrasound CT MR* 2004;25:419–431. [PubMed: 15559125]
2. Halligan S, Altman DG, Mallett S, et al. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 2006;131:1690–1699. [PubMed: 17087934]
3. Baker ME, Bogoni L, Obuchowski NA, et al. Computer-aided detection of colorectal polyps: can it improve sensitivity of less-experienced readers? Preliminary findings. *Radiology* 2007;245:140–149. [PubMed: 17885187]
4. Taylor SA, Charman SC, Lefere P, et al. CT Colonography (CTC): Investigation of the Optimum Reader Paradigm Using Computer Aided Detection Software. *Radiology* 2008;246:463–471. [PubMed: 18094263]
5. Petrick N, Haider M, Summers RM, et al. CT Colonography and Computer-aided Detection as a Second Reader: Observer Performance Study. *Radiology* 2008;246:148–156. [PubMed: 18096536]
6. 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;17:2598–2607. [PubMed: 17351780]
7. Chen SC, Lu DS, Hecht JR, Kadell BM. CT colonography: value of scanning in both the supine and prone positions. *AJR Am J Roentgenol* 1999;172:595–599. [PubMed: 10063842]
8. Fletcher JG, Johnson CD, MacCarty RL, Welch TJ, Reed JE, Ahlquist DA. CT Colonography in 180 patients: The benefit of prone imaging. *Gastroenterology* 1999;116:G1770.
9. Yong AA, Harris JE, Shorvon PJ. The value of prone imaging in CT pneumocolon. *Clin Radiol* 2000;55:959–963. [PubMed: 11124076]

10. 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;349:2191–2200. [PubMed: 14657426]
11. Summers RM, Yao J, Pickhardt PJ, et al. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology* 2005;129:1832–1844. [PubMed: 16344052]
12. Summers RM, Jerebko AK, Franaszek M, Malley JD, Johnson CD. Colonic polyps: complementary role of computer-aided detection in CT colonography. *Radiology* 2002;225:391–399. [PubMed: 12409571]
13. Jerebko AK, Malley JD, Franaszek M, Summers RM. Support vector machines committee classification method for computer-aided polyp detection in CT colonography. *Acad Radiol* 2005;12:479–486. [PubMed: 15831422]
14. Li J, Huang A, Yao J, et al. Optimizing computer-aided colonic polyp detection for CT colonography by evolving the Pareto front. *Med Phys* 2009;36:201–212. [PubMed: 19235388]
15. Dorfman DD, Berbaum KS, Lenth RV. Multireader, multicase receiver operating characteristic methodology: a bootstrap analysis. *Acad Radiol* 1995;2:626–633. [PubMed: 9419615]
16. Langsrud, O. Fisher's Exact Test.
17. Siegel, S.; Castellan, NJ, Jr. Nonparametric statistics for the behavioral sciences. 2. New York: McGraw-Hill; 1988.
18. 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;28:318–326. [PubMed: 15100534]
19. 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;249:47–53. [PubMed: 18682584]
20. Shiraishi J, Abe H, Engelmann R, Doi K. Effect of high sensitivity in a computerized scheme for detecting extremely subtle solitary pulmonary nodules in chest radiographs: observer performance study. *Acad Radiol* 2003;10:1302–1311. [PubMed: 14626305]
21. Obuchowski NA. Sample size tables for receiver operating characteristic studies. *AJR Am J Roentgenol* 2000;175:603–608. [PubMed: 10954438]
22. Zheng B, Leader JK, Abrams GS, et al. Multiview-based computer-aided detection scheme for breast masses. *Med Phys* 2006;33:3135–3143. [PubMed: 17022205]
23. Van Engeland S, Timp S, Karssemeijer N. Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. *Med Phys* 2006;33:3203–3212. [PubMed: 17022213]
24. Sahiner B, Chan HP, Hadjiiski LM, et al. Joint two-view information for computerized detection of microcalcifications on mammograms. *Med Phys* 2006;33:2574–2585. [PubMed: 16898462]
25. Gupta S, Markey MK. Correspondence in texture features between two mammographic views. *Med Phys* 2005;32:1598–1606. [PubMed: 16013719]
26. Acar B, Napel S, Paik DS, Li P, Yee J, Beaulieu CF. Registration of supine and prone CT colonography data: Method and evaluation. *Radiology* 2001;221:332–332.
27. Nappi J, Okamura A, Frimmel H, Dachman A, Yoshida H. Region-based supine-prone correspondence for the reduction of false-positive CAD polyp candidates in CT colonography. *Acad Radiol* 2005;12:695–707. [PubMed: 15935968]
28. Huang, A.; Roy, D.; Franaszek, M.; Summers, RM. Teniae coli guided navigation and registration for virtual colonoscopy. *Proceedings of the IEEE Visualization Conference*; 2005. p. 279-285.
29. Wang, S.; Van Uitert, RL.; Summers, RM. Automated matching of supine and prone colonic polyps based on PCA and SVMs. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*; 2008.
30. Huang, A.; Summers, RM.; Roy, D. Synchronous navigation for CT colonography. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*; 2006.
31. Zalis ME, Barish MA, Choi JR, et al. CT colonography reporting and data system: a consensus proposal. *Radiology* 2005;236:3–9. [PubMed: 15987959]

Figure 1.
Patient flowchart.

Figure 2.

Example of benefit of double-mark CAD presentation for polyp detection. Eight mm adenoma in transverse colon of 69 year-old woman. (A,C) Supine and (B,D) prone three-dimensional endoluminal CTC images. (A,B) Single-mark and (C,D) double-mark CAD presentations. Blue CAD marks shown on supine single-mark and supine and prone double-mark images. In single-mark CAD presentation, three radiologists missed polyp. In double-mark CAD presentation, three of four radiologists detected polyp.

Figure 3.

Polyp found more frequently by radiologists on both scans when CAD marked it on both scans. Seven mm adenoma in splenic flexure colon of 50 year-old man. (A,C) Supine and (B,D) prone three-dimensional endoluminal CTC images. (A,B) Single-mark and (C,D) double-mark CAD presentations. Blue CAD marks shown on supine single-mark and supine and prone double-mark images. All four radiologists detected polyp on single-mark and double-mark CAD presentations. Four radiologists stated that the polyp was visible on both scans in the double-mark situation but only one radiologist stated that the polyp was visible on both scans in the single-mark situation.

Figure 4.

False negative polyp example. Eight mm adenoma in rectum of 57 year-old man. (A,C) Supine and (B,D) prone three-dimensional endoluminal CTC images. (A,B) Single-mark and (C,D) double-mark CAD presentations. Blue CAD marks shown on prone single-mark and supine and prone double-mark images. In single-mark CAD presentation, two radiologists missed polyp (reasons for rejection: “normal mucosa”, “stool”). In double-mark CAD presentation, three radiologists missed polyp (reasons for rejection: “normal mucosa”, “stool”, “hemorrhoid”).

Figure 5. False positive example. Supine three-dimensional endoluminal CTC images (A) without and (B) with CAD false positive mark (blue) on a thickened haustral fold in transverse of 51 year old man. One of four radiologists incorrectly diagnosed the CAD mark to be on a polyp.

Table 1

Sensitivity per polyp for single-mark and double-mark CAD true positive presentation

Reader	Sensitivity (n=21)	
	Single-mark	Double-mark
1	8 38%	12 57%
2	15 71%	19 91%
3	15 71%	19 91%
4	12 57%	10 48%
Average	60% [51%, 68%]	71% [63%, 80%]

Numbers are polyps (%). Patients had at most one polyp. Note that per patient and per polyp sensitivities are not necessarily identical because radiologists could miss the true polyp and instead inappropriately mark a false positive as a polyp, leading to a TP patient and FN polyp. The differences between single-mark and double-mark reads for individual readers (Fisher exact test) were not statistically significant. The differences between single-mark and double-mark reads for the average reader ($p = .03$, three-factor ANOVA) was statistically significant. 95% confidence intervals are given for the sensitivities for the average reader. There was no statistically significant difference amongst readers for the single mark reads but there was a statistically significant difference amongst readers for the double mark reads ($p = .002$, Cochran's Q).

Table 2

Sensitivity and specificity per patient for single-mark and double-mark CAD true positive presentation

Reader	Sensitivity (n=21)		Specificity (n=12)	
	Single-mark	Double-mark	Single-mark	Double-mark
1	9 43%	13 62%	11 92%	8 67%
2	17 81%	19 90%	9 75%	10 83%
3	17 81%	19 90%	10 83%	12 100%
4	13 62%	10 48%	11 92%	10 83%
Average	56/84 67% [44%, 86%]	61/84 73% [48%, 94%]	41/48 85% [65%, 100%]	40/48 83% [63%, 98%]

Numbers are patients (%). Patients had at most one polyp. Note that per patient and per polyp sensitivities are not necessarily identical because radiologists could miss the true polyp and instead inappropriately mark a false positive as a polyp, leading to a true positive patient and false negative polyp. The differences between single-mark and double-mark reads for individual readers and for the average reader were not statistically significant using bootstrap analysis. 95% confidence intervals for the average reader are from the MRMC analysis. The data for specificity include three normal cases that were identical on two readings; each reading was arbitrarily assigned to either the single-mark or double-mark groups. For the other nine normal cases, a pair of false positives was either matched (double-mark) or unmatched on the supine and prone scans based on location and CAD score to mimic the situation for the abnormal cases.

Table 3

Sensitivity for polyp detection by radiologists on both supine and prone scans for single-mark and double-mark CAD true positive presentation

Reader	Single-mark (n=21)	Double-mark (n=21)	P
1	5 24%	12 57%	0.06
2	5 24%	14 67%	0.01
3	14 67%	18 86%	0.3
4	11 52%	9 43%	0.8
Average	42% [33%, 51%]	63% [54%, 72%]	0.002

Numbers are polyps (%) detected by radiologists on both supine and prone scans.

P-values are from Fisher exact test for the individual readers and three-factor ANOVA for the average reader. 95% confidence intervals are given for the sensitivities for the average reader.

Table 4

Reasons for False Negative Polyps

Reason Given by Radiologist for Not Calling Polyp	Single-mark Cases (n=34)	Double-mark Cases (n=62)
Stool	15 (44.1%)	20 (32.3%)
Normal Mucosa	8 (23.5%)	14 (22.6%)
Fluid	6 (17.6%)	7 (11.3%)
Fold	3 (8.8%)	10 (16.1%)
Not Viewable	2 (5.9%)	5 (8.1%)
Scan Artifact	0 (0%)	2 (3.2%)
Rectal Tube	0 (0%)	1 (1.6%)
Air Bubble	0 (0%)	1 (1.6%)
Other	0 (0%)	1 (1.6%)

Data are numbers of false negative polyps (%) according to the reasons the radiologists gave for not calling the CAD finding a polyp. These are CAD marks on polyps that were mischaracterized by the radiologists. "Not viewable" means CAD mark was not visible on colonic surface when radiologist clicked on that entry in list of CAD marks.

Table 5

False Positive Radiologist Findings According to Whether the CAD Marks were Intentionally Spatially Co-located on Supine and Prone Scans

Reader	Non-Matched FPs	Matched FPs
1	12	1
2	28	0
3	14	0
4	10	1
Average	64/1788* 3.6%	2/72 2.8%

Data are numbers (%) of false positive radiologist diagnoses according to whether the false positive was intentionally matched by the experimenters with a false positive in the same approximate location of the colon on the other scan. Multiple counts for some of the same false positives for readers and readings were included. The 72 matched false positives consist of a pair of false positives in each of 9 normal patients interpreted by each of four readers.

* Includes 16 false positives (4 for each reader) on the same object, believed to be an inverted diverticulum that was very polyp-like in appearance; and 12 false positives on the same 6 mm hyperplastic polyp.