CT Colonography: Role of a second reader CAD paradigm in the initial training of radiologists

Emanuele Neri a,*, Lorenzo Faggioni a, Daniele Regge b, Paola Vaglia a, Francesca Turini a, Francesca Cerri a, Eugenia Picano a, Sabina Giusti a, Carlo Bartolozzi a

a Diagnostic and Interventional Radiology, University of Pisa, Italy
b Institute for Cancer Research and Treatment, Candiolo, Italy

A R T I C L E   I N F O

Article history:
Received 15 January 2010
Received in revised form 19 July 2010
Accepted 19 July 2010

Keywords:
CT colonography
Virtual colonoscopy
Computed tomography
Colorectal polyps
Computer aided diagnosis

A B S T R A C T

Purpose: To evaluate the influence of CAD for the evaluation of CT colonography (CTC) datasets by inexperienced readers during the attendance of a dedicated hands-on training course.

Method and materials: Twenty-seven radiologists inexperienced in CTC (11 with no CTC training at all, 16 having previously reviewed no more than 10 CTC cases overall) attended a hands-on training course based on direct teaching on fifteen workstations (four Advantage Windows 4.4 with Colon VCAR software, GE; six CADCOLON, Im3D; five ColonScreen (Toshiba/Voxar) with ColonCAD™ API, Medicsight). During the course, readers were instructed to analyze 26 CTC cases including 38 colonic lesions obtained through low-dose MDCT acquisitions, consisting of 12 polyps sized less than 6 mm, 9 polyps sized between 6 and 10 mm, 12 polyps sized between 11 mm and 30 mm, and 5 colonic masses sized >3 cm. CTC images were reviewed by each reader both in 2D and 3D mode, respectively by direct evaluation of native axial images and MPR reconstructions, and virtual endoscopy or dissected views. Each reader had 15 min time for assessing each dataset without CAD, after which results were compared with those provided by CAD software. Global rater sensitivity for each lesion size before and after CAD usage was compared by means of two-tailed Student’s t test, while sensitivity of each single reader before and after CAD usage was assessed with the McNemar test.

Results: For lesions sized <6 mm, global rater sensitivity was 0.1852 ± 0.1656 (mean ± SD) before CAD-assisted reading and 0.2345 ± 0.1761 after CAD (p = 0.0018). For lesions sized 6–9 mm, sensitivity was 0.2870 ± 0.1016 before CAD-assisted reading and 0.3117 ± 0.1099 after CAD (p = 0.0027). For lesions sized 10–30 mm, sensitivity was 0.5308 ± 0.2120 before CAD-assisted reading and 0.5637 ± 0.2133 after CAD (p = 0.0086), while for lesions sized >30 mm, sensitivity before CAD-assisted reading was 0.3556 ± 0.3105 and did not change after CAD usage (p = 1). Sensitivity of each single rater did not significantly differ before and after CAD for any lesion size category (McNemar test, p > 0.05). Specificity was not significantly different before and after CAD for any lesion size (>96% for all size categories).

Conclusion: CAD usage led to increased overall sensitivity of inexperienced readers for all polyps sizes, except for lesions >30 mm, but sensitivity of individual raters was not significantly higher compared with CAD-unassisted reading.

1. Introduction

As computed tomography colonography (CTC) continues to evolve and improve, its use is shifting from highly specialized academic centers to community hospitals and nonacademic radiology practices [1–3]. Thus, many radiologists are experiencing pressure from clinical colleagues to offer CTC as part of the routine services provided in their practices, and hands-on courses at reading workstations are increasingly being encouraged as the most suitable method for training and certify CTC readers [4].

Manual reading of CTC images is increasingly prone to errors due to the high number of images to be analyzed, which may lead to reader’s fatigue [5]. Moreover, image interpretation is subjected to reader’s bias (as high incidence of perceptual errors and fatigue) and no systematic method has been designed so far for lesion visualization (2D or 3D) [6].

In the attempt at overcoming the limitations of software-unassisted reading of CTC datasets, several works have addressed the issue of evaluating the diagnostic performance of CAD (Com-
puter Aided Diagnosis) systems for CTC [7–12]. Most of the studies focus on the impact of CAD as second reader in experienced readers, however it is of interest to evaluate the impact of CAD systems on the diagnostic performance of readers without dedicated CTC experience, as could be those working in nonacademic centers, involved in reading large amount of CTC generated by a screening program. In this direction, the aim of our work was to evaluate the influence of CAD on the diagnostic performance of radiologists with no previous CTC training who have attended a residential hands-on CTC course.

2. Materials and methods

2.1. Training of radiologists and case selection

Twenty-seven radiologists with no previous CTC experience attended a three-day hands-on training course held at our Department, based on direct teaching on workstations. All of them had received no prior CTC training; 16 reported to have occasionally reviewed (not reported) a CTC exam in a clinical setting (on average 4 cases; minimum 2 and maximum 10) without the supervision of an expert; anyway, for the purpose of the study, at the beginning of the course all readers declared to the leading author (EN) that they have not received training in CTC before, and they had no experience in CTC reading.

The course was held over three consecutive days with 5 h of preliminary training to interpretation of CTC images and usage of workstations and CAD software. During the preliminary training the novices were simply trained on the use of the software tools, i.e. prone and supine images synchronization and scrolling, fly-through with endoscopic simulation, generation of a virtual dissected view, usage of measurement and reporting tools, and starting a CAD session in a second reader paradigm. All novices were trained on how to use the software with a set of three negative and

---

Table 1
Mean per-polyp sensitivity and false positive rates of CAD systems.

<table>
<thead>
<tr>
<th>CAD</th>
<th>CAD #1</th>
<th>CAD #2</th>
<th>CAD #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean sensitivity (&lt;6 mm)</td>
<td>0.8532 ± 0.2610</td>
<td>0.8124 ± 0.1917</td>
<td>0.8693 ± 0.3452</td>
</tr>
<tr>
<td>Mean FP rate (&lt;6 mm)</td>
<td>0.1852 ± 0.4833</td>
<td>0.1683 ± 0.3754</td>
<td>0.2154 ± 0.3255</td>
</tr>
<tr>
<td>Mean sensitivity (6–9 mm)</td>
<td>0.8734 ± 0.1622</td>
<td>0.8593 ± 0.2403</td>
<td>0.8195 ± 0.2221</td>
</tr>
<tr>
<td>Mean FP rate (6–9 mm)</td>
<td>0.2504 ± 0.3389</td>
<td>0.3088 ± 0.3674</td>
<td>0.2455 ± 0.3771</td>
</tr>
<tr>
<td>Mean sensitivity (10–30 mm)</td>
<td>0.9290 ± 0.3703</td>
<td>0.9088 ± 0.3831</td>
<td>0.8878 ± 0.3068</td>
</tr>
<tr>
<td>Mean FP rate (10–30 mm)</td>
<td>0.3135 ± 0.2754</td>
<td>0.2710 ± 0.3276</td>
<td>0.2643 ± 0.3411</td>
</tr>
<tr>
<td>Mean sensitivity (&gt;30 mm)</td>
<td>0.9843 ± 0.2938</td>
<td>0.9567 ± 0.4245</td>
<td>0.9901 ± 0.4955</td>
</tr>
<tr>
<td>Mean FP rate (&gt;30 mm)</td>
<td>0.1852 ± 0.3958</td>
<td>0.1755 ± 0.2398</td>
<td>0.2319 ± 0.3403</td>
</tr>
</tbody>
</table>

---

Fig. 1. 6-mm sessile polyp located in the ascending colon as seen on native axial CTC image (a), identified by CAD software (b: polyp marked in red); in (c) the polyp is visualized in virtual endoscopy mode. The CAD-proposed candidate polyp was rejected by 7 readers and accepted by 11 readers; CAD confirmed diagnosis of 9 readers.
Fig. 2. Axial CTC images show a 12-mm sessile polyp located on a cecal folder (a), identified by CAD software (b) (green arrows); the polyp is marked red on virtual endoscopy view (c). The CAD-proposed candidate polyp was rejected by 8 readers and accepted by 13 readers; CAD confirmed diagnosis of 6 readers.

three positive cases (with lesions of mixed size), in which diagnosis was given prior to training. The use of positive cases was aimed at training the novices on the use of measurement and reporting tools, as well as the recognition of the CAD markers.

Subsequently, 26 cases (excluding those already reviewed in the preliminary training) with 38 colonic lesions obtained through low-dose MDCT acquisitions were presented, consisting of 12 polyps sized less than 6 mm, 9 polyps sized between 6 and 10 mm, 12 polyps sized between 11 mm and 30 mm, and 5 colonic masses (>3 cm size).

Institutional review board approval was waived since all cases were retrospectively selected among previous CTC studies with a colonoscopic correlation, performed in our Department. All patients gave written informed consent for CTC.

2.2. Data acquisition

CTC datasets were obtained in 19 cases after a cathartic preparation based on oral introduction of 4 × 37.9 g of polyethelenglycole (PEG) diluted in 3–4 L of water 24 h before CTC. In seven cases a fluid tagging preparation was employed, consisting of 12 polyps sized less than 6 mm, 9 polyps sized between 6 and 10 mm, 12 polyps sized between 11 mm and 30 mm, and 5 colonic masses (>3 cm size).

Institutional review board approval was waived since all cases were retrospectively selected among previous CTC studies with a colonoscopic correlation, performed in our Department. All patients gave written informed consent for CTC.

2.2. Data acquisition

CTC datasets were obtained in 19 cases after a cathartic preparation based on oral introduction of 4 × 37.9 g of polyethelenglycole (PEG) diluted in 3–4 L of water 24 h before CTC. In seven cases a fluid tagging preparation was employed, consisting of 6 × 30 mL diatrizoate dimeglumine (Gastrografin\textsuperscript{TM}, Bracco Diagnostics, Milan, Italy) administered orally two days before CTC, diluted into 100 mL of water after every meal (six doses for a total of 180 mL of diatrizoate dimeglumine diluted into 600 mL of water). In both cases, a residue-free diet was prescribed starting 3 days before CTC.

All data were acquired on a 64-row MDCT scanner (LightSpeed VCT, GE Medical Systems, Milwaukee, WI) in the supine and prone position with the following parameters: tube voltage 120 kV, tube current 50–80 mA depending on patient size, rotation time 500 ms, beam pitch 1.375:1, detector configuration 64 × 0.625 mm, reconstructed slice thickness 1.25 mm, reconstruction increment 1 mm, standard reconstruction kernel.

2.3. Image reading

All CTC datasets were selected and reviewed by an experienced reader (EN) according to the following criteria: radiological findings had been confirmed by means of optical colonoscopy, and data should also have optimal colonic cleansing, distention, and tagging, and to be polyp-enriched.

CT images were exported in DICOM format to 15 workstations equipped with CAD software for evaluation of CTC datasets working in a second reader paradigm (four Advantage Windows 4.4 with Colon VCAR software, GE Medical Systems [CAD #1]; six CADCOLON, Im3D, Turin, Italy [CAD #2]; five ColonScreen (Toshiba/Voxar) with ColonCAD\textsuperscript{TM} API, Medicsight PLC, London, UK [CAD #3]).

All cases were pre-processed by each CAD. Mean per-polyp sensitivity and false positive rate was calculated for each CAD (Table 1).

Each workstation was set to run with its own standard factory CAD settings and was assigned to one or two readers: in this latter case, the second reader assessed CTC images after the first one and
in blind to him. All raters worked in a blind fashion. All trainees attending the course had general experience in abdominal imaging. CTC images were reviewed by readers in 2D or 3D mode, or in combination, on the basis of their subjective preferences, using direct evaluation of native axial images and/or MPR reconstructions and/or virtual endoscopy and/or dissected views (Figs. 1–3).

Each reader was instructed to perform a systematic per-polyp analysis of each CTC dataset by classifying any detected polyp according to its size: small polyps (<6 mm), medium polyps (6–9 mm), large polyps (10–30 mm), and colonic masses (>30 mm). Each reader had 15 min time for assessing each dataset, after which results were compared with those provided by CAD software, which was thus used as second reader. In the 5 min allocated for the CAD sessions, CAD marks were analyzed by scrolling through the dataset. Raters operated in a blind fashion in both CAD-unassisted and CAD-assisted reading sessions. A diagram summarizing the procedural workflow of the study is provided in Fig. 4.

During the CAD-unassisted 15-min reading session, readers recorded their results on a pre-printed form, on which they reported the slice number of the lesion, its maximum diameter, its segmental location, and its morphology. All suspected lesions had to be marked. In the CAD-assisted session, readers had to verify the correspondence between marked lesions and CAD-marked polyp candidates, but also reviewed CAD marks which had not eventually been marked at CAD-unassisted reading. The final decision was made by combining the polyp candidates selected in the CAD-unassisted and CAD-assisted reading settings.

As stated above, lesions were classified according to the CRADS criteria and the correspondence between CAD-unassisted and CAD-assisted marks was based on CT slice number and maximum diameter (1 mm mismatch between candidate measurement was tolerated). Anyway, all CAD software was designed to mark polyp candidates with a colored label, so that correspondence between them and those marked in the CAD-unassisted reading could easily be identified.

After the CAD-unassisted and CAD-assisted reading sessions, feedback was provided for each case; a tutor showed the same cases read by novices in a plenary session, and addressed the lesions con-
Table 2

<table>
<thead>
<tr>
<th>Lesion size</th>
<th>% sens NC</th>
<th>% sens C</th>
<th>p</th>
<th>% spec NC</th>
<th>% spec C</th>
<th>p</th>
<th>FP rate NC</th>
<th>FP rate C</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;6 mm</td>
<td>18.52 ± 16.56</td>
<td>23.45 ± 17.61</td>
<td>0.0018</td>
<td>99.71 ± 1.03</td>
<td>99.29 ± 1.86</td>
<td>0.0028 ± 0.0103</td>
<td>0.0071 ± 0.0186</td>
<td>0.1846</td>
<td></td>
</tr>
<tr>
<td>6–9 mm</td>
<td>28.70 ± 10.16</td>
<td>31.17 ± 10.99</td>
<td>0.0027</td>
<td>98.57 ± 3.23</td>
<td>98.57 ± 2.92</td>
<td>0.0142 ± 0.0323</td>
<td>0.0142 ± 0.0323</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10–30 mm</td>
<td>53.08 ± 21.20</td>
<td>56.37 ± 21.33</td>
<td>0.0086</td>
<td>97.19 ± 2.88</td>
<td>96.93 ± 2.92</td>
<td>0.0281 ± 0.0288</td>
<td>0.0307 ± 0.0292</td>
<td>0.1612</td>
<td></td>
</tr>
<tr>
<td>&gt;30 mm</td>
<td>35.56 ± 31.05</td>
<td>35.56 ± 31.05</td>
<td>1</td>
<td>99.44 ± 1.20</td>
<td>99.44 ± 1.20</td>
<td>0.0056 ± 0.0120</td>
<td>0.0056 ± 0.0120</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

firmed by endoscopy. During the 3-day course, with 20 min per case (26 cases), a total of 9 h of training was counted. Direct training on workstations was interleaved with short lectures and discussion on CTC practical issues, such as patient preparation, insufflation, and CTC indications. This format allowed, on one hand, to complete the training program with essential information for the clinical implementation of the technique, and on the other hand, to reduce readers’ fatigue after assessing multiple cases (four cases were read in each session, interleaved with 1–2 lectures of 15–20 min each).

In the 5-h training session, readers were trained to starting a CAD session and informed about the pros and cons of the CAD; however, a dedicated lecture on CAD principles and performance was given by one expert in this field (DR).

2.4. Statistical analysis

For each rater and lesion size category, sensitivity and specificity before and after CAD usage were calculated and expressed as mean ± standard deviation. As all CTC datasets contained at least one lesion, for each lesion size category true negative cases were computed as the sum of cases from the remaining categories minus false positive cases belonging to the category under examination.

In contrast with other works [9,13], in which reader response was coded in terms of a probability scale or score expressing the rater’s subjective degree of confidence about the likelihood of a finding to be a lesion, we preferred to record each rater’s response as a dichotomous variable (0 = no lesion; 1 = lesion detected), so to resemble as closely as possible the routine clinical scenario, in which the reporting radiologist is required to decide between lesion presence or absence. Given that every rater operated blinded and independently from each other, overall per-polyp sensitivity and specificity of raters as a group before and after CAD-assisted reading were compared by means of two-tailed paired Student’s t test.

In order to assess the potential impact of CAD on the performance of every single rater, sensitivity before and after CAD was also compared for each reader and lesion size by means of two-tailed McNemar test.

Statistical analysis was carried out by using software (MedCalc version 10.3, www.medcalc.be; GraphPad Prism 5.0b, www.graphpad.com). A p-value less than 0.05 was considered statistically significant.

3. Results

Overall values of sensitivity, specificity, and false positive (FP) rate in the reader group ranked by lesion size before and after CAD-assisted reading, respectively, are reported in Table 2.

Overall reader sensitivity was significantly higher after CAD-assisted reading than before for all lesion sizes, except for lesions larger than 30 mm.

Neither overall specificity nor FP rate of readers was significantly different before and after CAD for any lesion size.

Diagnostic performance of each reader in terms of sensitivity and specificity for each lesion size category is illustrated in Graphs 1 and 2, respectively. Sensitivity of each single rater did not significantly differ before and after CAD for any lesion size category (p > 0.05).

4. Discussion

CAD is a useful tool for assisting the radiologist in the diagnosis of several pathological conditions, especially in an oncological setting. Its potential applications have extensively been investigated.
A second reader paradigm. The lack of an increase in sensitivity when CAD is used by inexperienced investigators in comparison to experienced readers on a rater population basis. In particular, CAD images can significantly improve the lesion detection rate of inexperienced readers in the interpretation of diagnostic images.

Our results show that usage of CAD for lesion detection on CTC images can significantly improve the lesion detection rate of inexperienced readers on a rater population basis. In particular, CAD usage leads to a significant increase in the detection rate of lesions sized less than 30 mm, while no difference exists between detection rates of colonic masses (>30 mm). These data are in line with those from other authors [13,11], who report a significant increase in sensitivity when CAD is used by inexperienced investigators in a second reader paradigm. The lack of an increase for the detection of colonic masses can be explained by the fact that CAD software is not suited for detecting such large lesions, which are supposed to be easily recognizable by human readers alone. Despite this, it is interesting that reader sensitivity tends to increase with increasing lesion size up to 30 mm, but falls in the case of colonic masses: this finding could be explained by the fact that inexperienced readers may actually miss such large lesions due to their basic inability to read CTC images and interpret colonic anatomy, which exposes them to the risk of gross image interpretation errors. On the other hand, the progressive increase in sensitivity with increasing lesion size up to 10 mm may be explained by the fact that inexperienced readers tend to overlook smaller lesions owing to perceptual errors, as suggested by Slater et al. [5]. The high variability in lesion sensitivity may confirm this hypothesis, suggesting that lesion detection in inexperienced readers may be substantially hampered by lack of skill in applying systematic diagnostic criteria for lesion seeking, although all readers have recently received an intensive dedicated training course.

Despite the statistically significant improvement in sensitivity brought by CAD, sensitivity value as intended in absolute terms is poor both before and after CAD usage. Other papers [5,9] confirm that CAD is of limited value when primary diagnosis is made by a reader without a previous substantial training in CTC. In particular, in the work by Slater et al. [5] sensitivity ranged from 7.1% to 28.6%, 16.7% to 41.7%, and 16.7% to 83.3% for polyps sized 1–5 mm, 6–9 mm, and >10 mm, respectively. On the other hand, Halligan et al. [9] have reported that on average, even with CAD, detection rate was 51% for polyps above 10 mm and 38.2% for polyps sized between 6 and 9 mm.

It is noteworthy that, although sensitivity calculated on the whole rater population before and after CAD increases in a statistically significant manner for lesions below 30 mm, sensitivity of individual raters does not show a significant difference before and after CAD for any of the raters involved in the study. However, we believe that, from an operative standpoint, individual reader sensitivity is more important than sensitivity of the whole rater population, as it expresses the effective ability of the CAD to improve a reader’s diagnostic performance under real conditions.

Specificity tends to decrease, albeit not significantly and by a very little extent, after CAD application with lesions sized less than 6 mm and between 10 and 30 mm, while it remains unchanged for lesions sized between 6 and 9 mm, and above 30 mm. Potential CAD-related loss of specificity is a known phenomenon that occurs also with experienced readers [12,14], due to the fact that CAD algorithms are primarily designed to detect lesions as alterations of the normal bowel wall by means of automated geometric criteria, while lesion characterization is an issue in which the radiologist’s experience and judgment are more directly involved and, at least as of currently, much more difficult to be reproduced through dedicated algorithms.

Our findings raise the question of the effectiveness of CAD for improving diagnostic performance of inexperienced readers and focus the attention on the importance of adequate CTC training with the build-up of specialized skills. Hands-on courses are constantly offered by Academic Centers and Scientific Societies, as the most effective way to educate radiologists to the interpretation of CTC datasets; however, there is an emerging consensus that one of the most important factors affecting the diagnostic performance of CTC is dedicated training and experience of the radiologist who interprets the examination [15–17]. In particular, it has been suggested that competence in CTC image interpretation cannot be assumed even after directed training via a database of 50 cases [15,17].

Our study has a limitation in that, because of the teaching aim of the CTC hands-on course, we used CAD software by three different manufacturers for evaluation of CTC datasets instead of a single CAD application for all users; this may theoretically entail some
variability in the performance of CAD tools, with a potential impact on the readers' diagnostic gain after CAD usage. However, all CAD plugins operated with their own standard settings in order to maximize diagnostic performance under the widest possible range of conditions, and the fact that sensitivity for lesion detection did not improve for any of the readers corroborates the hypothesis that this finding is independent from the features of individual CAD packages.

Another limitation is the relatively low number of cases (26 patients with a total of 38 colonic lesions) that might have prevented us from finding a statistically significant difference in the diagnostic performance of readers with and without CAD assistance. However, no more cases could be included due to the limited time available in a 3-day course for case revision, also considering that additional time was spent for lectures and critical case review.

In conclusion, our data show that usage of CAD software for CTC image evaluation by inexperienced readers, despite increasing diagnostic sensitivity on a rater population basis, does not lead to significantly higher sensitivity of individual raters. This finding may indicate that CAD tools are unable to compensate for lack of reader experience, thus stressing the importance of adequate training (presumably greater than that achievable through a CTC hands-on course) in order for radiologists to gain a satisfactory degree of diagnostic accuracy for the assessment of CTC datasets in a clinical setting.

Conflict of interest

None.

References


