



Locoregional CT staging of colon cancer: does a learning curve exist?

Eun Kyoung Hong^{1,2} · Francesca Castagnoli^{1,3} · Nicolò Gennaro^{1,4} · Federica Landolfi^{1,5} · Carlos Perez-Serrano^{1,6} · Ieva Kurilova^{1,7} · Sander Roberti⁸ · Regina Beets-Tan¹

Received: 8 June 2020 / Revised: 13 July 2020 / Accepted: 18 July 2020 / Published online: 30 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose To evaluate the learning curve for locoregional staging of colon cancer in radiologist trainees.

Methods Eighty-eight cases of colon cancer CT were included in this retrospective study. Four senior radiology residents staged the CTs according to TNM classification. Two out of four radiologists received feedback after reading every 20 cases. Radiologic staging was compared with pathologic staging and the learning curve, diagnostic performance, reader confidence and reading time were evaluated and compared between the two groups (feedback vs. no feedback). Generalized estimating equations logistic regression, QICu statistic, ANOVA and *t* test/Mann–Whitney test were utilized.

Results Radiologists demonstrated a significant increase in their performance to distinguish between $\leq T2$ and $\geq T3$ and reached an inflection point at 38 cases, with a significant association with increased number of cases reviewed ($P < 0.001$). Sensitivity ($P < 0.001$), specificity ($P = 0.030$) and NPV ($P = 0.002$) demonstrated significant associations with increased experience. The overall reader's confidence was significantly higher in the group which received feedback ($P < 0.001$). There was no significant improvement in performance nor in reader's confidence for N staging (N0 vs. $\geq N1$) for all readers. Reading time decreased with experience and showed a significant negative association with experience ($P < 0.001$).

Conclusion Diagnostic performance of senior radiology trainees in differentiating between T2 and T3 colon cancer on CTs improved with increased experience. In contrast, evaluation of lymph node involvement did not improve with more experience. Feedback had no significant effect on improvement of diagnostic performances.

Keywords Colon cancer · Staging · Learning curve · Computed tomography

Introduction

Recent randomized controlled trials have proved that accurate presurgical staging of colon cancer is important for selecting patients who will benefit from neoadjuvant

chemotherapy in locally advanced colon cancer [1] and identify patients who are not eligible for laparoscopic surgery in advanced state of colon cancer [2]. These studies have utilized CT to stratify high-risk patients into a differentiated treatment. Therefore, the role of CT in primary staging of colon cancer patients will increasingly extend from mainly accurate distant staging towards locoregional staging of the disease. Although CT has for years been

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00261-020-02672-7>) contains supplementary material, which is available to authorized users.

✉ Eun Kyoung Hong
amyh0803@gmail.com

¹ Department of Radiology, Netherlands Cancer Institute, Amsterdam, Netherlands

² Department of Radiology, Seoul National University Hospital, Seoul, Korea

³ Department of Radiology, University of Brescia, Brescia, Italy

⁴ Department of Radiology, Humanitas Clinical and Research Center IRCCS, Milan, Italy

⁵ Radiology Unit, Sant'Andrea Hospital, Sapienza University of Rome, Rome, Italy

⁶ Department of Radiology, Hospital Clinic of Barcelona, Barcelona, Spain

⁷ GROW School of Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands

⁸ Department of Epidemiology and Biostatistics, Netherlands Cancer Institute, Amsterdam, Netherlands

adopted as the standard imaging workup, most published data have addressed the performance of CT for staging liver and extrahepatic metastases and evidence on locoregional staging accuracies is weaker [3]. One recent meta-analysis has shown that the sensitivity and specificity of CT for detecting tumor over T3 staging was 90 and 69% but demonstrated lower performance for detecting tumor invasion beyond muscular propria layer over 5 mm (T3cd-T4) with sensitivity and specificity of 77 and 70%, respectively. Moreover, the sensitivity and specificity of CT in detecting lymph node involvement was relatively low (71 and 67%, respectively) [4]. Majority of these studies have included readers who had at least over 2 years of experiences [5]. Not much is known about the expertise of the radiologists that is required to accurately stage the local tumor for reliable clinical decision making. Learning curve is a concept of improvement in performance after repetition, introducing the importance of practice experience as an essential tool for accomplishing the expertise [6]. Locoregional CT staging of colon cancer can be especially challenging due to anatomical reasons, such as presence of haustra and its peculiar orientation within abdominal cavity, also geographical distortion caused by peristaltic movements. These require visualization of colon cancer in 3-dimensional projection, utilizing techniques like multiplanar reconstruction (MPR). Therefore, familiarizing oneself to view colon cancer in different perspectives can demand longer period and repetition to achieve a certain level of expertise. Previous studies in other cancer types have demonstrated significant improvement of reader's performance through continuous feedback [7]; however, data for colon cancer staging are lacking. Considering the increasingly important role of CT to select high-risk patients and stratify treatment it is important to know whether a learning curve for locoregional staging exists and whether senior trainees with minimal practice experience would benefit from an interactive feedback.

Materials and methods

Study population

This retrospective study was approved by our institutional review board (IRBd18-119). The requirement for written informed consent was waived. Six hundred patients with pathological colon cancer diagnoses, between January 2010 and December 2016, were included from institutional database search. We consecutively included patients who underwent presurgical CT followed by surgical resection of colon cancer within one-month period. Exclusion criteria included patients with rectal cancers, mucinous cancers, small sized colon tumor lesion which could not be visualized on CT, poor image quality and patients

who received neoadjuvant chemotherapy. The included cohort comprised 88 patients, which were consecutively divided into four batches of cases, creating 20 cases of four batches for the learning curve analysis. The remaining 8 cases were utilized as part of introductory training and was not included in the batches that were given to the readers for learning curve analysis [8]. Table 1 summarized the demographic and tumor characteristics of the final included cohort.

Computed tomography

All patients were kept on nil per os for 2–4 h before the CT scan and had oral intake of 20 cc of omnipaque 300 mg/ml (General Electronics Healthcare, Milwaukee, WI, USA) diluted in 1 l of water, 1 h before the examination. Bowel preparation was not performed before CT. One of three different 16–64 slice CT scanners (Gemini TF (Philips, Amsterdam, Netherlands), Sensation (Siemens, Erlangen, Germany) and Aquillion (Toshiba, Tokyo, Japan)) were used to perform CT scan. Patients underwent preoperative abdomen CT with iodine-based intravenous contrast (3 ml/s, total amount of 90–130 ml, followed by bolus injection of 30 cc normal saline) in portal-venous phase at 70 s delay. Images from all CT scanners were reconstructed at 1–2 mm slice thickness.

Table 1 Patient demographic and tumor characteristics

Characteristics (<i>N</i> =88)	Number	Percentage (%)
Age (median)	72	
Range	34–93	
Tumor side		
Cecum	31	35.2
Ascending colon	8	9.1
Hepatic flexure	6	6.8
Transverse colon	2	2.3
Splenic flexure	3	3.4
Descending colon	7	8.0
Sigmoid colon	31	35.2
T stage		
T1	6	6.8
T2	14	15.9
T3	49	55.7
T4	19	21.6
N stage		
N0	37	42.0
N1	22	25.0
N2	29	33.0

Reader interpretation

Four senior radiology residents, trained in different centers in Europe participated in the study and all four readers had exposures to abdomino-pelvic CT during their training and had specific interest in abdominal imaging. However, all of them had almost no previous experience in precise staging of colon cancer on CT. All readers were provided with imaging literature regarding colon cancer staging and recent developments in colon cancer treatment. Thereafter the four readers received a 90-min interactive introductory lecture on colon cancer CT staging provided by an experienced board-certified radiologist (E.H., with over 6 years of experience in abdominal CT imaging), under the supervision of a senior faculty member (R.B., with over 20 years of experience in abdominal CT imaging). This lecture covered anatomy of colon, T and N staging of colon cancer using CT, with images demonstrated each staging category. Subsequently, they reviewed 8 exemplary colon cancer CTs that were not included in the study on a separate workstation. Readers used multiplanar reconstruction software, which allowed images to be viewed from three-dimensional views (Fig. 1). This exercise aided the readers to familiarize themselves with interpreting images and staging colon cancer on CT and using the local PACS system and software (Carestream Health, Rochester, NY, USA).

The readers were blinded to all clinical and histological information, except the location of tumor to ensure the consistency and accurate time measurement of the image interpretation. The four readers reviewed images independently on separate workstations, interpreted images

in sequence and recorded their findings in online survey tool that was created for the analysis. The following items were scored; (1) T staging of tumor and reader's confidence (using 5-scale, with 5 as the most confident and 1 as the least confident), (2) N staging of tumor and reader's confidence, (3) Time required for interpretation of each cases in seconds. T staging was divided into 3 categories: T1-2 tumor was defined as mass in colon lumen limited to the bowel wall with clear pericolic fat on CT or luminal projection of mass without any bowel wall distortion. T3 tumor was defined as a tumor which demonstrated smooth or nodular extension beyond the normal contour of the bowel wall. T4 tumors were defined as a tumor which extends into adjacent peritoneum or grown into other adjacent tissues or organ (Fig. 2). For N staging, a lymph node was considered as malignant when a lymph node was over 9 mm in short axis. Additionally, cluster of more than 3 lymph nodes along the locoregional vascular pedicle was considered as malignant [9–11].

The four readers were randomly divided into two groups. One group of 2 readers received interactive feedback by an experienced radiologist after each batch of 20 CTs. They were able to review radiological findings and pathological information and discuss their findings with experienced radiologist. The feedback was given for each case and the readers were allowed to ask any questions regarding image and pathological data interpretation. In another group, readers were only provided with pathological data after the interpretation of each batch and left with own will to compare their findings.

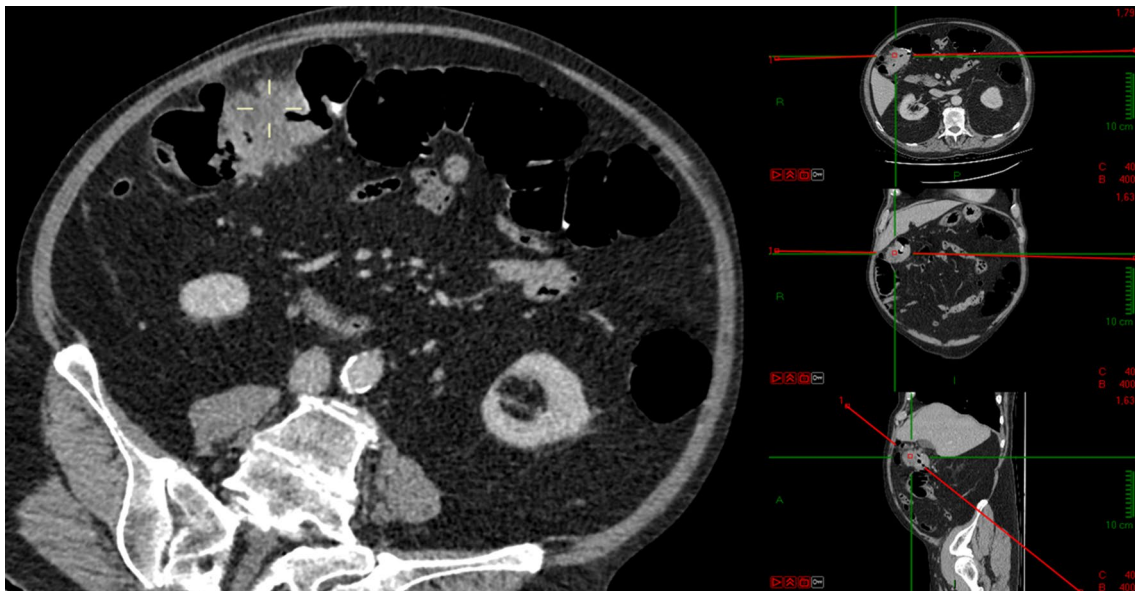


Fig. 1 Example of radiologic stage T4N1 colon cancer is depicted on designated local PACS system. Readers were advised to use multiplanar reconstruction software to analyze in three-dimensional views

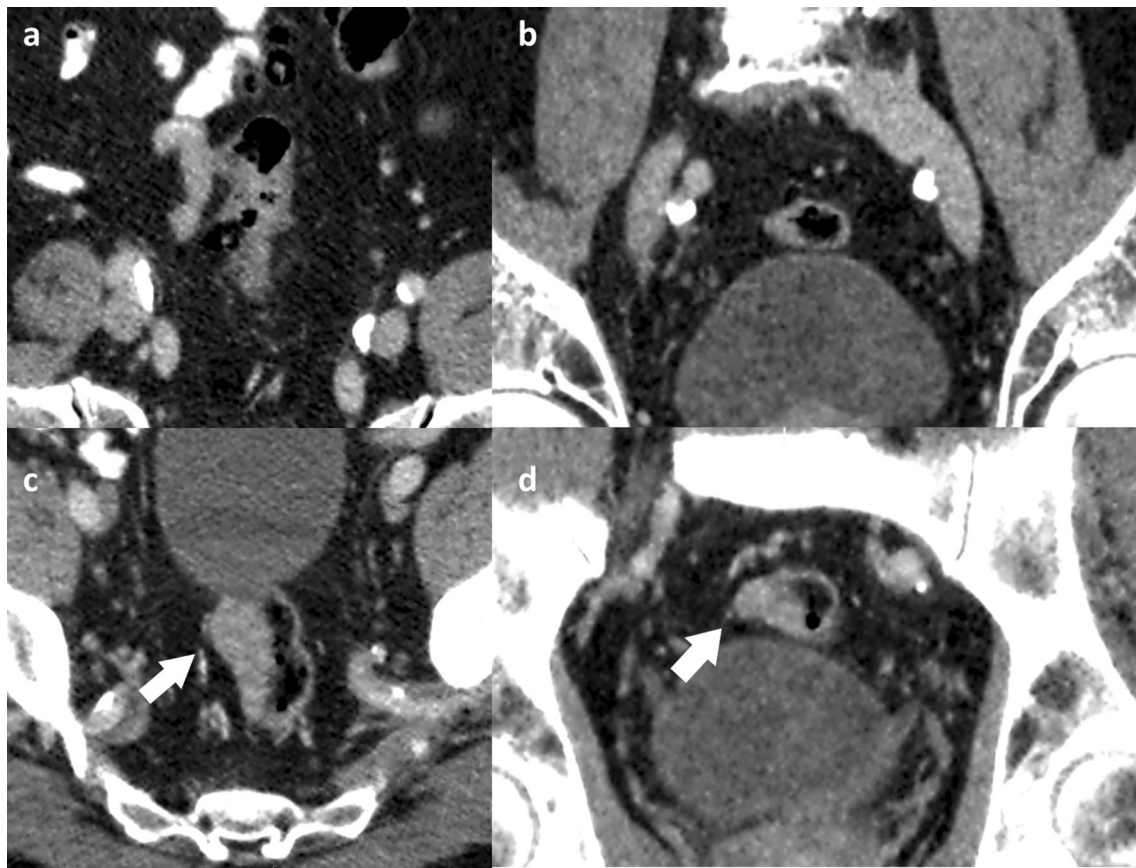


Fig. 2 Axial (a) and coronal (b) reformatted CT images demonstrating pathologic T2 stage sigmoid colon cancer, depicting bowel wall thickening with clear pericolic fat. Axial (c) and coronal (d) reformat-

ted CT images demonstrating pathologic T3 stage sigmoid colon cancer, with nodular extension beyond the normal contour of the bowel wall (white arrow)

2.4 Reference standard

Specimens were processed according to the local institutional protocol that is based on guidelines by College of American Pathologists and Royal College of Pathologists. Reporting of the histopathologic information followed synoptic protocol according to Pathologisch Anatomisch Landelijk Geautomatiseerd Archief [12].

Statistics

The readers' accuracies, sensitivities, specificities, positive and negative predictive values (PPV and NPV) in differentiating between $\leq T2$ and $\geq T3$ tumors (criteria used to select patients who are eligible for neoadjuvant chemotherapy [1]) and assessing lymph node involvement were evaluated. Learning curves for each reader were constructed by smoothing the scatter plots of correct assessment 0/1 against the number of evaluated tumors using loess regression. Generalized estimating equations logistic regression models were used to model the probability of a correct

assessment, with each reader and the number of evaluated tumors as predictors. The model also incorporated an interaction effect between the number of evaluated tumors and feedback group. The effect of number of evaluated tumors was assumed to plateau, with the same point of inflection for readers with and without feedback. The correlation between multiple evaluations of the same tumor was modeled with an exchangeable correlation structure, i.e., the same assumed correlation between all pairs of evaluations of the same tumor. The optimal point of inflection was determined by determining the integer point (2–79) associated with the smallest QICu statistic.

Averages (median for ordinal outcomes and mean for the continuous outcome) and standard deviations were assessed for T- and N-staging confidence (ordinal 1–5), and for reading time (continuous). Smoothed curves of confidence and reading time against number of evaluated tumors were constructed using loess regression. For all three outcomes, linear generalized estimating equations models with predictors reader ID and the number of evaluated tumors. The models further incorporated an interaction effect between the

number of evaluated tumors and feedback group, and an exchangeable correlation structure between multiple evaluations of the same tumor.

Different combinations of sequential batches are utilized (e.g., batch 1 vs. 2, 3 and 4, batch 1, 2 vs. 3, 4 and batch 1, 2, 3 vs. 4) for the analysis and comparison of diagnostic performance and reader improvement according to the location of primary tumor. Reader improvement was evaluated using ANOVA and error variance was permitted to eliminate the hypothesis of homogeneity according to different groups (e.g., groups of batches, feedback vs. no feedback). When comparing groups to identify differences in the change of diagnostic performance with increased experience, diagnostic performance according to increased number of reviewed cases are considered independently within the groups.

Reader confidence and reading time was compared using *t* test (for variable with normal distribution) and Mann–Whitney test (for variables that are not normally distributed) between groups of readers who received feedback and who did not.

All statistical analyses were performed using R version 3.6.1 and MedCalc version 19.1.3. *P* values <0.05 were considered statistically significant.

Results

T staging ($\leq T2$ vs. $\geq T3$)

The learning curve for accuracy of staging between $\leq T2$ and $\geq T3$ colon cancer is demonstrated in Fig. 3. All the readers have significantly improved and showed statistically significant association with number of cases reviewed ($P < 0.001$) in the 1st 40 cases in both groups (with and without feedback). The four readers have reached plateau in their accuracy after 38 cases, which is presented as an inflection point in Fig. 3. There was no significant difference in the slopes between the groups who received feedback and who did not. Sensitivity of T staging also demonstrated significant association with experience ($P < 0.001$) and displayed inflection point at 32 cases. Specificity and NPV also showed significant associations with increased number of reviewed cases (0.030 and 0.002, respectively), whereas PPV did not show a significant association with increased experience. The presence of feedback revealed a significant correlation with NPV ($P = 0.003$), but did not show a significant association with accuracy, sensitivity, specificity and PPV ($P = 0.093$ – 0.780).

The diagnostic performance of all four readers in distinguishing between $\leq T2$ and $\geq T3$ colon cancer is depicted in Table 2 and Supplementary Table S1. At the baseline, there was no significant difference in accuracy, sensitivity, specificity, PPV and NPV between readers and groups of feedback vs. no feedback ($P = 0.268$ – 0.826). The improvement

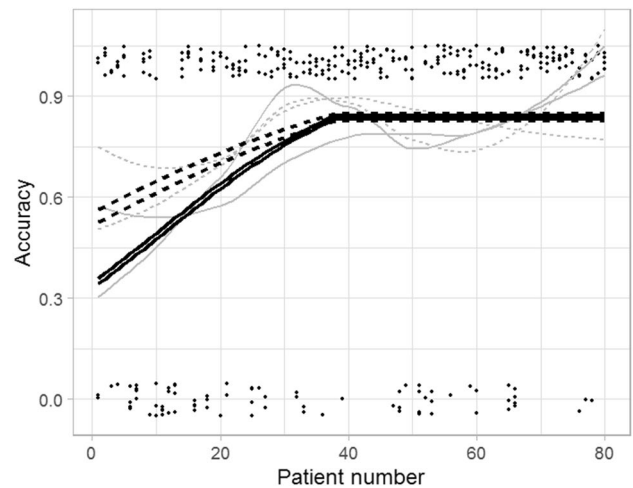


Fig. 3 Learning curves for T staging ($\leq T2$ vs. $\geq T3$) of colon cancer with increasing number of reviewed cases. Accuracy of T staging in readers who received feedback (dashed line) and did not receive feedback (solid line) are presented. Accuracy was plotted and fitted using generalized estimating equations logistic regression models

of accuracy was significant between batch 1 vs. 2–4 ($P < 0.001$), batches 1, 2 vs. 3, 4 ($P = 0.001$) and batch 1, 2, 3 vs. 4 ($P = 0.003$). Sensitivity and NPV improved with increased number of examined cases with statistical significance; batch 1 vs. 2–4 ($P_s < 0.001$), batches 1, 2 vs. 3, 4 ($P_s = 0.001$) and batch 1, 2, 3 vs. 4 ($P = 0.042$ for sensitivity and $P = 0.005$ for NPV). Specificity showed statistical significance between batch 1, 2, 3 vs. 4 ($P = 0.014$) and PPV did not reveal a significant difference in any combination of batch groups. There was no statistically significant difference in the improvement of diagnostic performance between the groups who received feedback and who did not ($P_s > 0.05$) (Supplementary Table S1).

Accuracy of locoregional colon cancer according to the location of primary tumor is demonstrated in Table 3. There were statistically significant differences in accuracy, sensitivity, specificity and PPV according to the groups of primary colon tumor location ($P_s < 0.05$). T staging for tumors located in hepatic/splenic flexures showed the highest accuracy and PPV ($P = 0.001$ and < 0.001 , respectively) and tumors located in cecum/ascending colon tumor revealed higher sensitivity and specificity in ($P < 0.001$ and 0.004 , respectively). NPV for T staging was not significantly different according to the location of primary colon tumor. Importantly, there was a significant difference in the improvement of accuracy of colon tumor T staging according to the location of primary tumor ($P = 0.002$). Accuracy was improved between batches 1 vs. 2,3,4 and batches 1, 2 vs. 2, 3 for T staging of tumors in transverse/descending colon, sigmoid colon and hepatic/splenic flexures ($P_s < 0.05$), whereas there was no statistically significant improvement of accuracy was

Table 2 Diagnostic performance of all readers in distinguishing between $\leq T2$ and $\geq T3$ colon cancer and lymph node involvement with increasing number of evaluated cases

Batch no	Accuracy	Sensitivity	Specificity	PPV	NPV
T staging ($\leq T2$ vs. $\geq T3$)					
1	53.8 (43/80)	52.8 (38/72)	62.5 (5/8)	92.7 (38/41)	12.8 (5/39)
2	81.3 (65/80)	94.6 (53/56)	50.0 (12/24)	81.5 (53/65)	80.0 (12/15)
3	78.6 (63/80)	86.7 (52/60)	55.0 (11/20)	82.2 (52/61)	57.9 (11/19)
4	87.5 (70/80)	88.3 (53/60)	85.0 (17/30)	94.6 (53/56)	70.8 (17/24)
All	75.3 (241/320)	79.0 (196/248)	62.5 (45/72)	87.9 (196/223)	46.4 (45/97)
N staging (N0 vs. $\geq N1$)					
1	56.3 (45/80)	72.9 (35/48)	31.3 (10/32)	61.4 (35/57)	43.5 (10/23)
2	76.3 (61/80)	82.8 (53/64)	50.0 (8/16)	86.9 (53/61)	42.1 (8/19)
3	45.0 (36/80)	63.9 (23/36)	29.5 (13/44)	42.6 (23/54)	50.0 (13/26)
4	66.3 (53/80)	77.1 (37/48)	50.0 (16/32)	69.8 (37/53)	59.3 (16/27)
All	60.9 (195/320)	75.5 (148/196)	37.9 (47/124)	65.8 (148/225)	49.5 (47/95)

Note numbers are in percentages

PPV positive predictive value, NPV negative predictive value

Table 3 Improvement of diagnostic performance of all readers in distinguishing between $\leq T2$ and $\geq T3$ colon cancer according to the location of primary colon tumor

Tumor location	Batch no	Accuracy	Sensitivity	Specificity	PPV	NPV
Cecum and ascending colon	1 (N=12)	75.0 (9/12)	100.0 (8/8)	25.0 (1/4)	72.7 (8/11)	100.0 (1/1)
	2 (N=48)	81.3 (39/48)	96.9 (31/32)	50.0 (8/8)	79.5 (31/39)	88.9 (8/9)
	3 (N=32)	90.6 (29/32)	100.0 (28/28)	25.0 (1/4)	90.3 (28/31)	100.0 (1/1)
	4 (N=40)	80.0 (32/40)	70.6 (29/36)	75.0 (3/4)	96.7 (29/30)	30.0 (3/10)
	All (N=132)	82.6 (109/132)	92.3 (96/104)	46.4 (13/28)	86.5 (96/111)	61.9 (13/21)
Transverse and descending colon	1 (N=8)	0.0 (0/8)	0.0 (0/8)	N.A	N.A	0.0 (0/8)
	2 (N=8)	50.0 (4/8)	100.0 (4/4)	0.0 (0/4)	50.0 (4/8)	N.A
	3 (N=20)	75.00 (15/20)	83.3 (10/12)	62.5 (5/8)	76.9 (10/13)	71.4 (5/7)
	4 (N=4)	75.0 (3/4)	N.A	75.0 (3/4)	0.0 (0/1)	100.0 (3/3)
	All (N=40)	55.0 (22/40)	58.3 (14/24)	50.0 (8/16)	63.6 (14/22)	44.4 (8/18)
Sigmoid colon	1 (N=48)	56.3 (27/48)	52.3 (23/44)	100.0 (4/4)	100.0 (23/23)	16.0 (4/25)
	2 (N=20)	90.0 (18/20)	87.5 (14/16)	100.0 (4/4)	100.0 (14/14)	66.7 (4/6)
	3 (N=20)	55.0 (11/20)	50.0 (6/12)	62.5 (5/8)	66.7 (6/9)	45.5 (5/11)
	4 (N=24)	95.8 (12/24)	100.0 (12/12)	91.7 (11/12)	92.3 (12/13)	100.0 (11/11)
	All (N=112)	70.5 (79/112)	65.5 (55/84)	85.7 (24/28)	93.2 (55/59)	45.3 (24/53)
Hepatic and splenic flexures	1 (N=12)	58.3 (7/12)	58.3 (7/12)	N.A	100.0 (7/7)	0.0 (0/5)
	2 (N=4)	100.0 (4/4)	100.0 (4/4)	N.A	100.0 (4/4)	N.A
	3 (N=8)	100.0 (8/8)	100.0 (8/8)	N.A	100.0 (8/8)	N.A
	4 (N=12)	100.0 (12/12)	100.0 (12/12)	N.A	100.0 (12/12)	N.A
	All (N=36)	86.1 (31/36)	86.1 (31/36)	N.A	100.0 (31/31)	0.0 (0/5)

Note numbers are in percentages

N.A. not applicable, PPV positive predictive value, NPV negative predictive value

noted for T staging of tumors located in cecum/ascending colon with increased reader experience ($P_s > 0.05$).

Reader's confidence for distinguishing between $\leq T2$ and $\geq T3$ colon cancer showed statistically significant association with feedback as the number of reviewed cases increased ($P=0.002$) (Fig. 4). Reader's confidence showed increment between batches 1,2 vs. 3,4 ($P=0.006$) and batches 1, 2, 3

vs. 4 ($P=0.020$) in groups of readers who received feedback but did not show any increment in groups of readers who did not receive feedback ($P > 0.05$). There was a significant difference in overall reader confidence between the two groups ($P < 0.001$). Comparison of reader confidence according to the batch number showed statistically higher reader confidence in readers who received feedback for batches 1, 3

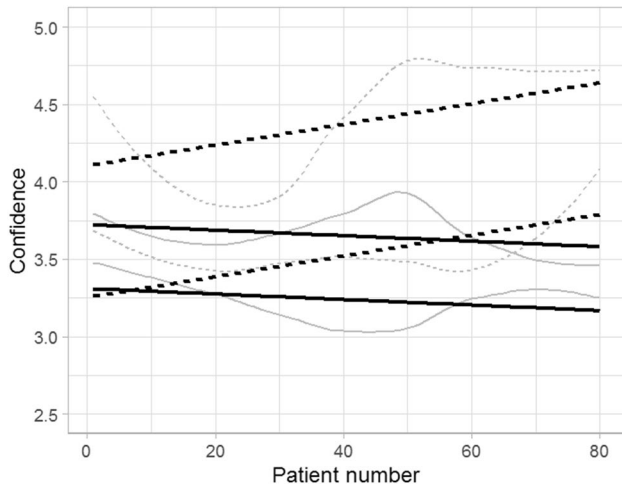


Fig. 4 Reader confidence with increasing number of reviewed cases for T staging of colon cancer in scale of 1–5 in readers with feedback (dashed line) and without feedback (solid line)

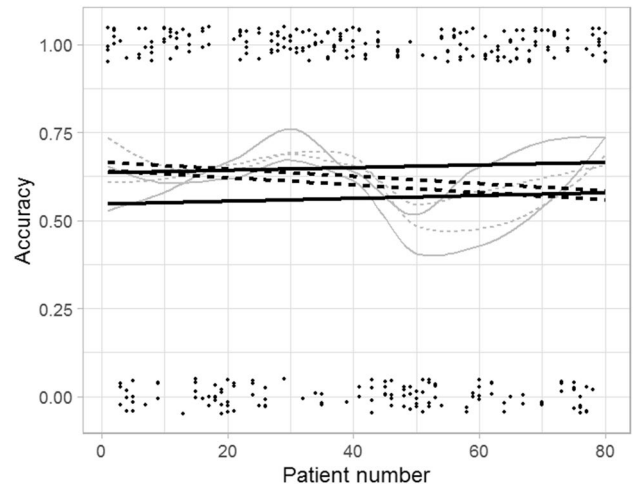


Fig. 5 Learning curve for N staging (N0 vs. \geq N1) of colon cancer with increasing number of reviewed cases. Accuracy of N staging in readers who received feedback (dashed line) and did not receive feedback (solid line) are presented. Accuracy was plotted and fitted using generalized estimating equations logistic regression models

Table 4 Comparison of reader confidence in distinguishing between \leq T2 and \geq T3 colon cancer and lymph node involvement

Batch no	All readers	With feedback	Without feedback	P
T staging (\leqT2 vs. \geqT3)				
1	3.68 ± 0.67	3.83 ± 0.71	3.53 ± 0.60	0.045
2	3.55 ± 0.79	3.70 ± 0.91	3.40 ± 0.63	0.091
3	3.77 ± 0.87	4.05 ± 0.96	3.50 ± 0.68	0.004
4	3.79 ± 0.87	4.23 ± 0.80	3.35 ± 0.70	< 0.001
All	3.70 ± 0.81	3.95 ± 0.87	3.44 ± 0.65	< 0.001
N staging (N0 vs. \geqN1)				
1	3.74 ± 0.76	3.83 ± 0.75	3.65 ± 0.77	0.305
2	3.78 ± 0.76	4.00 ± 0.78	3.56 ± 0.68	0.010
3	3.74 ± 0.87	3.95 ± 1.04	3.53 ± 0.60	0.028
4	3.61 ± 1.05	3.90 ± 1.19	3.33 ± 0.80	0.013
All	3.72 ± 0.87	3.95 ± 0.95	3.52 ± 0.72	< 0.001

Note Numbers are mean ± standard deviation, using 5-score scale. P values are derived from the comparison of groups with vs. without feedback

and 4 ($P=0.045$, 0.004 and <0.001 , respectively) (Table 4), implying that the readers in the feedback group had higher confidence initially and feedback enhanced the increment of their confidence, compared to the readers who did not receive feedback.

Evaluation of lymph node involvement

Analysis of learning curve for accuracy of lymph node involvement of colon cancer revealed no significant improvement with increased number of reviewed cases ($P=0.850$) (Table 2, Fig. 5). Also, Sensitivity, specificity, PPV and NPV did not reveal any improvement with

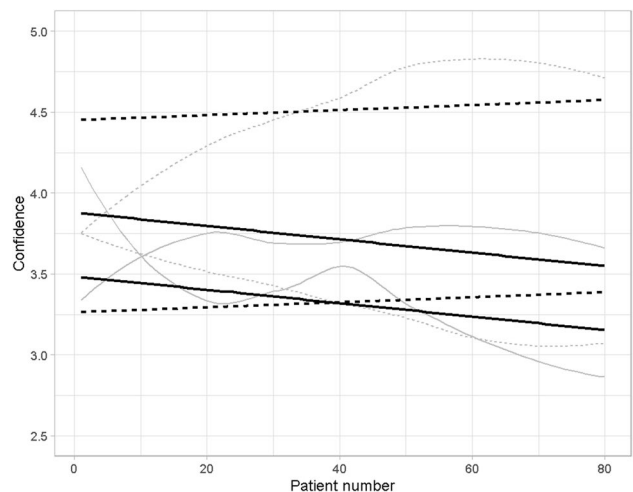


Fig. 6 Reader confidence with increasing number of reviewed cases for N staging of colon cancer in scale of 1–5 in readers with feedback (dashed line) and without feedback (solid line)

increased reader experience ($P > 0.05$). There was no statistically significant difference in diagnostic performance of identifying lymph node involvement between groups with and without feedback ($P > 0.05$).

There was no significant increase in reader confidence of N staging in both groups of reader who received feedback and who did not ($P_s < 0.05$) and there was no significant association with feedback as the number of reviewed cases increased ($P=0.070$) (Fig. 6). Reader confidence for identifying lymph node involvement was higher in groups

of readers who received feedback in batches 2, 3, 4 and all examinations combined ($P_s < 0.05$) (Table 4).

Reading time

There is a statistically significant decrease in reading time for all readers as the number of reviewed cases increased (between batch 1 vs. 2, 3, 4, batch 1, 2, vs. 3, 4 and batch 1, 2, 3 vs. 4, all $P < 0.001$) and reading time showed a significant negative correlation with number of reviewed cases ($P < 0.001$) (Fig. 7). Between groups with and without feedback, readers who received feedback had significantly shorter reading time in batch 1, 3, 4 and all cases combined ($P_s < 0.05$) (Table 5).

Discussion

This study evaluated the learning curve for T ($\leq T2$ vs. $\geq T3$) and N staging (N0 vs. $\geq N1$) of colon cancer on primary CT of four radiology senior trainees, from various countries and different radiology training programs. All four readers,

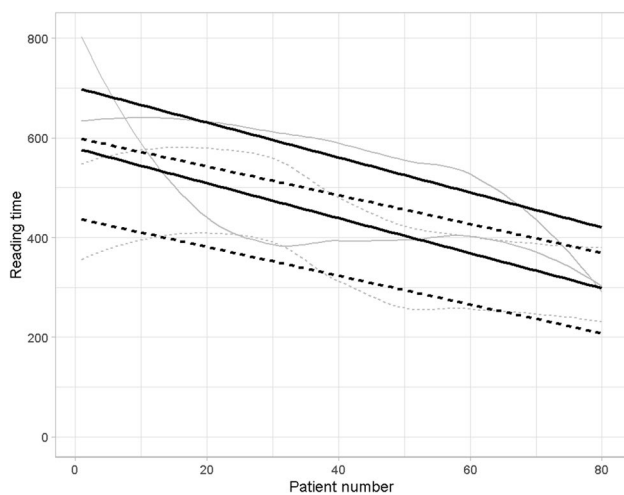


Fig. 7 Reading time with increasing number of reviewed cases in seconds in readers with feedback (dashed line) and without feedback (solid line)

Table 5 Comparison of reading time with increased number of reviewed cases

Batch no	All readers	With feedback	Without feedback	<i>P</i>
1	537.60 ± 187.89	476.70 ± 166.72	598.50 ± 190.01	0.003
2	489.21 ± 184.45	468.93 ± 173.51	509.50 ± 194.84	0.328
3	406.89 ± 179.87	333.03 ± 123.46	480.75 ± 197.75	<0.001
4	368.13 ± 137.21	333.00 ± 127.66	403.25 ± 138.96	0.021
All	450.46 ± 185.20	402.91 ± 163.90	498.00 ± 193.40	<0.001

Note Numbers are mean ± standard deviation, in seconds. *P* values are derived from the comparison of groups with vs. without feedback

regardless of feedback, which was given after every 20 cases, demonstrated a significant improvement in their performance of T staging and reached a plateau after 38 cases. Although there was a higher accuracy for tumors located in right hemicolon and hepatic, splenic flexures, there was no significant difference in performance between readers who received feedback and who did not. Interestingly, drastic improvement of accuracy was found in initial 38 cases in all readers, with or without feedback and the improvement slowed down after this point. However, diagnostic performances of the readers continued to improve after this point, and reached level of comparable diagnostic performance of experienced readers that were reported [4] after reviewing all the cases in the study. Such finding suggests that further experience after the inflection point is helpful in improvement of diagnostic accuracy of colon cancer and further study with larger number of readers and cases can be helpful to identify the learning curve of radiologists after the inflection point. However, reader's confidence was significantly different between the two groups. The group that received feedback revealed a significant increase in confidence, with increased number of cases. For all readers—regardless of whether or not feedback was given, the diagnostic performance and reader's confidence for N staging however did not improve with increased number of cases. The time that is required to read the images decreased for all readers with increased number of cases. Shorter reading times were achieved for readers who received feedback as compared to those who did not.

With proven feasibility of neoadjuvant chemotherapy and immunotherapy in colon cancer patients in clinical trials [1, 13], there is a need for a dedicated training program for radiologists to reach a competent level of locoregional CT staging. Especially, when the decision for neoadjuvant chemo or immunotherapy will be based on CT imaging findings.

Previous studies used a cut off for high-risk colon cancer as $\geq T3c$ [14, 15], whereas our study applied the cut off between $\leq T2$ and $\geq T3$, following published clinical trials which considered patients with T3 disease as high-risk, regardless of the degree of extension beyond muscular propria [1, 15]. In our own experience, precise measurement of tumor extension beyond bowel wall is not reproducible

due to the limited soft tissue contrast of CT, though further studies are needed to confirm our hypothesis.

We found that readers performed better in T staging of tumors located in cecum, ascending colon (right hemicolon), compared with that located in sigmoid colon (tumors in these two locations were enough in numbers for statistical analysis in this study). This can be explained by the anatomical orientation of the right colon, situated vertically in craniocaudal direction of the body, revealing the perpendicular dimension of the tumor in axial imaging, which most of the radiologists are the most familiar with. The sigmoid colon can be diverse in length and alignment according to individuals, which makes it harder to evaluate tumor stage and requires multiplanar reconstruction tools for better staging. Interestingly, tumors in hepatic and splenic flexures are staged more accurately than in transverse and descending colon but because of the small number of cases it's too early to draw any conclusions. Further research with large number of tumors in various locations within colon should be done to validate the findings of this study.

We showed no significant improvement in the two groups for N staging. Also reader's confidence did not increase with more experience, regardless of the presence of feedback. This is in concordance with previous meta-analysis that revealed high variability of CT performances for nodal staging and sensitivities and specificities varied according to the criteria chosen [4]. In this study, we have utilized a simplified version criteria that is used in the FOxTROT trial based on the size of lymph nodes to identify malignant nodes [9, 11] and found there is a high variability among readers and no improvement with more experience.

A previous report showed no effect of feedback to unexperienced readers in detecting tumor in prostate cancer [7]. The authors suggest that radiologists, as a trained visual learners, are capable of improving themselves in self-directed manner and with experience [7]. Although there was no significance difference in the diagnostic performance of readers between the two groups of readers with and without feedback, the readers who received feedback gained significantly more confidence with experience than those who did not receive feedback. This could be explained by the improved reassurance that readers perceive if he/she experiences positive feedback [1]. The findings of the current study suggest radiologists to be trained with exposure of at least 38 cases of colon cancer to ensure the quality of the report. Unfortunately, for N staging training did not provide a significant increase in performance nor in reader's confidence.

There are several limitations to this study. First, this is a retrospective study. Second, only senior radiology residents participated in this study. Learning curve of the general radiologists who would be more likely to perform locoregional CT staging in clinical setting has not been investigated. However we believe that the findings of our study

have provided a good direction towards how many CTs young radiologists should be exposed to to reach an acceptable accuracy for therapeutic decision making. Prospective studies involving radiologists with diverse expertise should be performed in the future. Third, we only evaluated colon cancer staging in binary fashion as discriminating tumor between T2 and T3 stage for T staging and N0 and N1 stage for N staging. This approach allowed us to identify the learning curve of the most robust criteria that will result in different treatment approaches taking in consideration the most current and upcoming changes of the treatment guidelines of colon cancer [1]. Fourth, the feedback was given regularly after reviewing each batch, consisting of 20 cases, a different scenario than in the clinical setting where trainees would more likely be given feedback after each case. However, during the course of the study, the readers in the group with feedback were allowed to ask further information from expert radiologists at any time when needed. Fifth, our cohort included rather large number of right-sided colon cancers. The suspected reason for this finding is that the center involved in the study is specialized for treating MSI tumors, which tend to have more right-sided colon cancers. Further study with cohort of colon cancers with evenly distributed location of tumors should be done.

In conclusion, diagnostic performance of CT T staging of colon cancer improved significantly in the first 38 cases in senior residents radiologists and continued to improve at a slower rate after this point, whereas CT N staging did not show a significant improvement. Colon cancers located the in right hemicolon was easier for radiologists to accurately stage than that in sigmoid colon and feedback for staging tumors did not lead to an improvement in performance suggesting that experience itself would be sufficient for radiologists to become accurate.

References

1. Foxtrot Collaborative G (2012) Feasibility of preoperative chemotherapy for locally advanced, operable colon cancer: the pilot phase of a randomised controlled trial. *Lancet Oncol* 13 (11):1152–1160. [https://doi.org/10.1016/S1470-2045\(12\)70348-0](https://doi.org/10.1016/S1470-2045(12)70348-0).
2. Group CS (2000) COLOR: a randomized clinical trial comparing laparoscopic and open resection for colon cancer. *Dig Surg* 17 (6):617–622. <https://doi.org/10.1159/000051971>.
3. Labianca R, Nordlinger B, Beretta GD, Brouquet A, Cervantes A, Group EGW (2010) Primary colon cancer: ESMO Clinical Practice Guidelines for diagnosis, adjuvant treatment and follow-up. *Ann Oncol* 21 Suppl 5:v70–77. <https://doi.org/10.1093/annonc/mdq168>.
4. Nerad E, Lahaye MJ, Maas M, Nelemans P, Bakers FC, Beets GL, Beets-Tan RG (2016) Diagnostic Accuracy of CT for Local Staging of Colon Cancer: A Systematic Review and Meta-Analysis. *AJR Am J Roentgenol* 207 (5):984–995. <https://doi.org/10.2214/AJR.15.15785>.

5. Dighe S, Blake H, Koh MD, Swift I, Arnaout A, Temple L, Barbachano Y, Brown G (2010) Accuracy of multidetector computed tomography in identifying poor prognostic factors in colonic cancer. *Br J Surg* 97 (9):1407–1415. <https://doi.org/10.1002/bjs.7096>.
6. Saba L, Guerriero S, Sulis R, Pilloni M, Ajossa S, Melis G, Malarini G (2011) Learning curve in the detection of ovarian and deep endometriosis by using Magnetic Resonance: comparison with surgical results. *Eur J Radiol* 79 (2):237–244. <https://doi.org/10.1016/j.ejrad.2010.01.019>.
7. Rosenkrantz AB, Ayoola A, Hoffman D, Khasgiwala A, Prabhu V, Smereka P, Somberg M, Taneja SS (2017) The Learning Curve in Prostate MRI Interpretation: Self-Directed Learning Versus Continual Reader Feedback. *AJR Am J Roentgenol* 208 (3):W92–W100. <https://doi.org/10.2214/AJR.16.16876>.
8. Leeuwenburgh MM, Wiarda BM, Bipat S, Nio CY, Bollen TL, Kardux JJ, Jensch S, Bossuyt PM, Boormeester MA, Stoker J (2012) Acute appendicitis on abdominal MR images: training readers to improve diagnostic accuracy. *Radiology* 264 (2):455–463. <https://doi.org/10.1148/radiol.12111896>.
9. Horvat N, Raj A, Liu S, Matkowskyj KA, Knezevic A, Capanu M, Shia J, Pickhardt PJ, Gollub MJ (2019) CT Colonography in Preoperative Staging of Colon Cancer: Evaluation of FOxTROT Inclusion Criteria for Neoadjuvant Therapy. *AJR Am J Roentgenol* 212 (1):94–102. <https://doi.org/10.2214/AJR.18.19928>.
10. Maupoey Ibanez J, Pamies Guilabert J, Frasson M, Bosca Robledo A, Giner Segura F, Garcia-Granero Ximenez E (2019) Accuracy of CT colonography in the preoperative staging of colon cancer: a prospective study of 217 patients. *Colorectal Dis* 21 (10):1151–1163. <https://doi.org/10.1111/codi.14724>.
11. Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo-Semedo L, Fenlon HM, Gollub MJ, Gourtsoyianni S, Hallsigan S, Hoeffel C, Kim SH, Laghi A, Maier A, Rafaelsen SR, Stoker J, Taylor SA, Torkzad MR, Blomqvist L (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 28 (4):1465–1475. <https://doi.org/10.1007/s00330-017-5026-2>.
12. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, Meijer GA (2007) Pathology data-banking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 29 (1):19–24. <https://doi.org/10.1155/2007/971816>.
13. Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K, Lopez-Yurda M, Grootcholten C, Beets GL, Snaebjornsson P, Maas M, Mertz M, Veninga V, Bounova G, Broeks A, Beets-Tan RG, de Wijkerslooth TR, van Lent AU, Marsman HA, Nuijten E, Kok NF, Kuiper M, Verbeek WH, Kok M, Van Leerdam ME, Schumacher TN, Voest EE, Haanen JB (2020) Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med* 26 (4):566–576. <https://doi.org/10.1038/s41591-020-0805-8>.
14. Dighe S, Swift I, Magill L, Handley K, Gray R, Quirke P, Morton D, Seymour M, Warren B, Brown G (2012) Accuracy of radiological staging in identifying high-risk colon cancer patients suitable for neoadjuvant chemotherapy: a multicentre experience. *Colorectal Dis* 14 (4):438–444. <https://doi.org/10.1111/j.1463-1318.2011.02638.x>.
15. Hunter C, Siddiqui M, Georgiou Delisle T, Blake H, Jeyadevan N, Abulafi M, Swift I, Toomey P, Brown G (2017) CT and 3-T MRI accurately identify T3c disease in colon cancer, which strongly predicts disease-free survival. *Clin Radiol* 72 (4):307–315. <https://doi.org/10.1016/j.crad.2016.11.014>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.