



# Radiology Residents' Independent Diagnosis of Appendicitis Using 2-mSv Computed Tomography: A Secondary Analysis of a Large Pragmatic Randomized Trial

Jungheum Cho<sup>1</sup>, Hae Young Kim<sup>2</sup>, Seungjae Lee<sup>3</sup>, Ji Hoon Park<sup>1,3,4,5</sup>, Kyoung Ho Lee<sup>1,3,4,5,6</sup>

<sup>1</sup>Department of Radiology, Seoul National University Bundang Hospital, Seongnam, Korea

<sup>2</sup>Department of Radiology, Asan Medical Center, Seoul, Korea

<sup>3</sup>Department of Applied Bioengineering, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea

<sup>4</sup>Department of Radiology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

<sup>5</sup>Department of Medical Device Development, Seoul National University College of Medicine, Seoul, Korea

<sup>6</sup>Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, Korea

**Objective:** To compare the diagnostic performance and clinical outcomes of 2-mSv computed tomography (CT) and conventional-dose CT (CDCT), following radiology residents' interpretation of CT examinations for suspected appendicitis.

**Materials and Methods:** Altogether, 3074 patients with suspected appendicitis aged 15–44 years ( $28 \pm 9$  years, 1672 females) from 20 hospitals were randomly assigned to the 2-mSv CT ( $n = 1535$ ) or CDCT ( $n = 1539$ ) groups in a pragmatic trial from December 2013 and August 2016. Overall, 107 radiology residents participated in the trial as readers in the form of daily practice after online training for 2-mSv CT. They made preliminary CT reports, which were later finalized by attending radiologists via addendum reports, for 640 and 657 patients in the 2-mSv CT and CDCT groups, respectively. We compared the diagnostic performance of the residents, discrepancies between preliminary and addendum reports, and clinical outcomes between the two groups.

**Results:** Patient characteristics were similar between the 640 and 657 patients. Residents' diagnostic performance was not significantly different between the 2-mSv CT and CDCT groups, with a sensitivity of 96.0% and 97.1%, respectively (difference [95% confidence interval {CI}], -1.1% [-4.9%, 2.6%];  $P = 0.69$ ) and specificity of 93.2% and 93.1%, respectively (0.1% [-3.6%, 3.7%];  $P > 0.99$ ). The 2-mSv CT and CDCT groups did not significantly differ in discrepancies between the preliminary and addendum reports regarding the presence of appendicitis (3.3% vs. 5.2%; -1.9% [-4.2%, 0.4%];  $P = 0.12$ ) and alternative diagnosis (5.5% vs. 6.4%; -0.9% [-3.6%, 1.8%];  $P = 0.56$ ). The rates of perforated appendicitis (12.0% vs. 12.6%; -0.6% [-4.3%, 3.1%];  $P = 0.81$ ) and negative appendectomies (1.9% vs. 1.1%; 0.8% [-0.7%, 2.3%];  $P = 0.33$ ) were not significantly different between the two groups.

**Conclusion:** Diagnostic performance and clinical outcomes were not significantly different between the 2-mSv CT and CDCT groups following radiology residents' CT readings for suspected appendicitis.

**Keywords:** Abdomen; Appendicitis; Radiation dosage; Tomography, X-ray computed

## INTRODUCTION

Although evidence that computed tomography (CT) radiation can be reduced to 2 mSv for diagnosing

appendicitis without negative ramifications is compelling [1-4], the implementation of low-dose CT techniques in clinical practice has been unsatisfactory [5]. This may be due to the fact that in many medical centers, especially

**Received:** January 8, 2023 **Revised:** February 28, 2023 **Accepted:** April 4, 2023

**Corresponding author:** Hae Young Kim, MD, PhD, Department of Radiology, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

• E-mail: khysjsh19@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

after hours, the diagnosis of appendicitis is dependent on the autonomous practice of radiology residents who are likely inexperienced with low-dose CT techniques.

In fact, in many academic centers, preliminary reports of CT scans performed after hours are made by radiology residents and are later supplemented by attending radiologists [6,7]. Although around-the-clock coverage by attending radiologists may seem ideal for reporting accuracy, patient throughput, and overall quality of care, such coverage is often practically infeasible due to constrained financial and human resources [7]. For such in-house attendance coverage to be amenable to long-term operations, several competent radiologists who are comfortable with diverse body parts and modalities are required. Moreover, around-the-clock coverage by attending radiologists may have negative repercussions on resident education [7].

Discrepancy rates between preliminary CT reports made by residents and addendum reports made by attending radiologists, which are typically reported to be < 3% [8-16], are often cited as evidence that radiology residents can practice autonomously without causing detrimental impacts on patient care. However, evidence regarding whether residents' performance in diagnosing appendicitis is adequately competent using low-dose CT is scarce. To the best of our knowledge, only two previous studies [17,18] have measured the performance of residents in diagnosing appendicitis on low-dose CT, with both studies reporting slightly superior performance of attending radiologists compared with that of residents on low-dose CT. However, these studies were limited by a small patient sample (14 cases of appendicitis in one study [17] and 48 cases in another [18]) and a retrospective study design. In this study, we compared the diagnostic performance and undesirable clinical outcomes of 2-mSv CT and conventional-dose CT (CDCT) following the autonomous reading of CT examinations by radiology residents in patients with suspected appendicitis. Our use of data from a large pragmatic randomized controlled trial [2] with unprecedentedly large numbers of patients and residents involved in practice provides relevant evidence that closely resembles reality.

## MATERIALS AND METHODS

### Data Source and Study Setting

We performed a secondary analysis of a prospective pragmatic randomized controlled trial (Low-dose CT for Appendicitis Trial, LOCAT; ClinicalTrials.gov number,

NCT01925014) [2,19], which included 3074 patients and > 500 healthcare providers from 20 sites. The institutional review boards of all the sites approved the study (IRB No. SNUBH B-2108-705-104, CAUH 2108-030-19386, GMC 2021-343, DSMC 2021-08-132, JNUH 2021-08-010, SCHSH 2021-08-013, HUKNHH 2021-08-032, YUHS 4-2021-1133, SNUH J-2109-122-1257, KNUH 2021-09-024-001, WKUH 2021-08-027-002, KUAH 2021AS0266, KBSMC 2021-09-024, CNUH 2021-381, WMCSB 202202-11, CMC KC14EIMI0655, KUGH 2021GR0441, HUSHH HALLYM 2021-08-029-001, SCHBC 2021-08-034-002, DMC 2021-09-009). Our patients have been included in previous studies [19,20], including the main report of the trial [2]. Compared to the main trial that aimed to prove the non-inferiority of 2-mSv CT over CDCT with respect to the negative appendectomy rate, this study analyzed the diagnostic performance and clinical outcomes following radiology residents' CT readings. We adhered to reporting guidelines [21-23].

### Patients

The eligibility criteria for the trial were patients aged 15–44 years who were referred from the emergency department for a CT examination under the suspicion of appendicitis [24]. Between December 2013 and August 2016, 3074 patients (mean age  $\pm$  standard deviation,  $28 \pm 9$  years; 1672 females) were randomized into the 2-mSv CT ( $n = 1535$ ) or CDCT groups ( $n = 1539$ ).

### Diagnostic Intervention

As detailed elsewhere [2], 22 CT machines with 16–640 channels from various manufacturers were used. The target effective dose for CDCT was individualized for each CT machine according to the institutional normal dose (median, 7 mSv; interquartile range [IQR], 6–7 mSv). Size-specific dose estimates were 4.1 mGy (IQR, 3.7–4.5) and 14.4 mGy (IQR, 12.9–16.2) for the 2-mSv CT and CDCT groups, respectively. The section thickness was  $\leq 5$  mm, with an overlap of  $\geq 20\%$ . All the patients received intravenous contrast agents, and none received an oral contrast agent.

### Radiologists and Prospectively Issued CT Reports

Altogether, 161 radiologists issued CT reports in their daily practice: 53 attending radiologists, 107 residents, and 1 board-certified on-call radiologist. For CT examinations performed after hours (times other than 0800–1700 h, Monday to Friday), it was a common practice in the catchment area for residents to make preliminary CT reports. All

preliminary reports were confirmed or revised via addendum reports made by attending radiologists, typically on the morning of the next business day. Radiologists could not be blinded to the randomization results because of the image texture differences between 2-mSv CT and CDCT. Although the majority (114 radiologists) had inconsiderable prior experience with 2-mSv CT, all radiologists were encouraged to complete an online training course on 2-mSv CT before trial initiation [17].

The radiologists issued CT reports using a predefined standardized report form (Supplementary Table 1) [25], which included the likelihood score for appendicitis rated on a five-point Likert scale and a free-text section for any alternative diagnosis. The following predefined diagnostic criteria for appendicitis [24] were used. The primary criterion was an enlarged appendix (> 6 mm in diameter) with mural thickening and periappendiceal fat stranding. The secondary diagnostic criteria included abnormal mural enhancement, appendicolith, phlegmon, and abscess [26,27].

### Co-Intervention and Follow-Up

The emergency physicians determined the necessity and timing of imaging studies. In cases of equivocal appendicitis on the initial CT, ultrasonography or CDCT could be used. The surgeons determined the surgical plan for each patient.

The follow-up of all patients included a telephonic interview conducted over 3 months in which patients were asked whether they needed additional hospitalization due to recurrence or exacerbation of symptoms. The site investigators also collected data regarding any serious adverse events that occurred during follow-up and their attributes [24,28].

### Reference Standard

Independent outcome assessors adjudicated the final diagnosis of appendicitis based on medical records and clinical follow-up results [24], according to predefined reference standards, as presented in the Supplementary Material.

### Discrepancy of CT Reports between Residents and Attending Radiologists

Two study coordinators (each with 8 and 7 years of clinical experience) independently reviewed the reports of all CT examinations that were read preliminarily by residents and identified cases with discrepancies between the preliminary CT report and the subsequent addendum report

made by the attending radiologists. First, a discrepancy in the likelihood score for appendicitis was identified, defined as cases with a negative index test result (likelihood score 1 or 2) on a preliminary report that was revised to a positive result (likelihood score 3, 4, or 5) on the addendum report, or vice versa (i.e., positive result revised to negative result). Second, a discrepancy in the alternative diagnosis was identified, defined as any revision regarding the alternative diagnosis that involved diseases with relatively specific imaging findings (e.g., ureter stone or acute right colonic diverticulitis). Of the discrepancies in alternative diagnosis, a "major" discrepancy delimited to revisions that involved diseases that may have required emergency surgery (e.g., ovarian torsion) was additionally identified. Any disagreement between the two study coordinators regarding the presence of discrepancy was resolved by discussion.

### Endpoints

We performed between-group comparisons (i.e., 2-mSv CT vs. CDCT) of the residents' diagnostic performance in terms of sensitivity and specificity. We also compared the 2-mSv CT and CDCT groups with respect to discrepancies between preliminary CT reports made by residents and addendum CT reports made by attending radiologists.

Additionally, between-group comparisons were performed considering undesirable clinical outcomes following the residents' reading. First, in terms of the final diagnosis, we counted the number of patients with negative appendectomy and perforated appendicitis. Negative appendectomy and perforated appendicitis, reciprocal to one another, represent the consequences of false-positive and false-negative (i.e., delayed) diagnoses, respectively [29]. Second, in terms of patient disposition, we counted the number of extended surgeries performed for the treatment of appendicitis (beyond simple appendectomy, such as partial cecectomy) and additional imaging tests. We also measured the interval between CT and discharge without surgery, the interval between CT and appendectomy, and the length of hospital stay associated with appendectomy. Third, the number of adverse events was counted.

In the original trial and other most relevant literatures, the "negative appendectomy rate" and "appendiceal perforation rate" by definition included patients who underwent appendectomy as the denominator. In this study, we used the number of patients in each group as the denominator to calculate the percentage values for the clinical outcomes, for example, the proportion (%) of negative

appendectomies = the number of negative appendectomies ÷ of all patients in each group. This was performed to preserve the effects of randomization as much as possible.

**Statistical Analysis**

Two radiologists (with 8 and 7 years of experience, respectively) and a statistician (with 5 years of experience) planned and performed the analyses after trial data collection. We primarily performed intention-to-treat analyses but also presented the results of per-protocol analyses that were performed by excluding patients with protocol non-adherence regarding eligibility or radiation dose (58 and 42 in the 2-mSv CT and CDCT groups, respectively) [24].

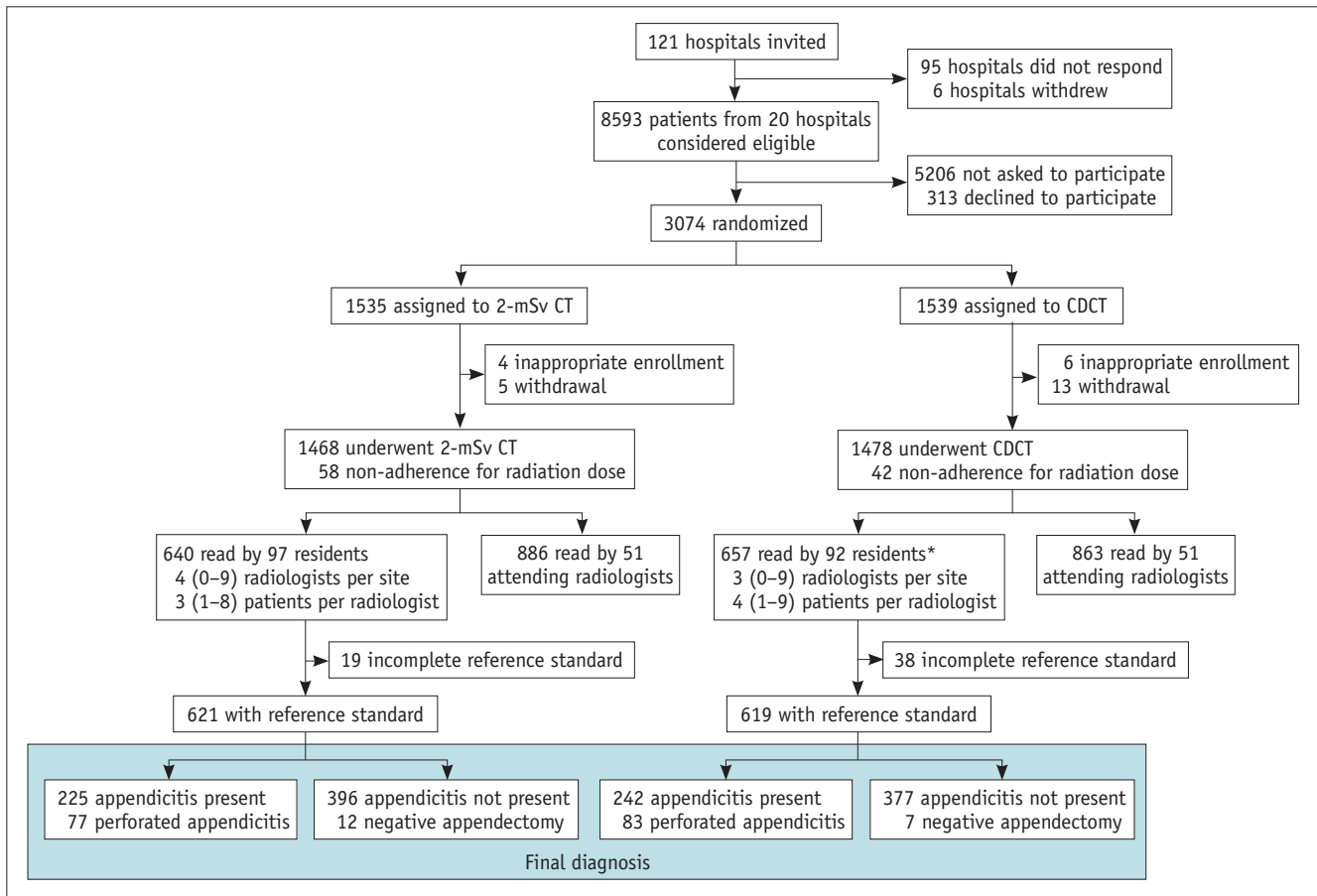
We used the chi-square test and Mann–Whitney U test to compare the 2-mSv CT and CDCT groups. For extremely low event rates, we used Z-pooled unconditional exact tests instead of chi-square tests [30]. We also generated heat maps to visually portray the frequency and distribution

of discrepancies between the residents and attending radiologists when rating the likelihood scores for appendicitis. For all analyses of the likelihood score for appendicitis, grades 3–5 were considered positive for appendicitis.

Statistical significance was defined as a two-sided  $P < 0.05$ . Because missing data were rare, we did not include them in our analysis. All analyses were performed using the R software version 3.6.3 ([www.R-project.org](http://www.R-project.org), The R Foundation for Statistical Computing).

**RESULTS**

As the results of the intention-to-treat and per-protocol analyses were very similar, we have presented intention-to-treat analysis primarily and reported per-protocol analysis in the Supplementary Material. The residents preliminarily read the CT scans of 640 patients (41.7% [640/1535]) in the 2-mSv CT group (Fig. 1) and 657 patients (42.7% [657/1539]) in the CDCT group. The characteristics of the



**Fig. 1.** Patient flow diagram. Unless otherwise indicated, data are numbers of patients. Data in italic are median (interquartile range). \*Included one board-certified on-call radiologist without fellowship training in abdominal radiology. CDCT = conventional-dose computed tomography, CT = computed tomography

**Table 1.** Patient Characteristics

Characteristic	Patients Whose CTs Were Read Preliminarily by Residents		P
	2-mSv CT Group (n = 640)	CDCT Group (n = 657) <sup>‡</sup>	
Age, yr	29 (22–36)	27 (20–35)	0.02
Sex			0.67
Female	362 (56.6)	363 (55.3)	
Male	278 (43.4)	294 (44.7)	
Body size			
Body mass index (kg/m <sup>2</sup> )	21.5 (19.6–24.0)	22.0 (19.9–24.3)	0.12
< 18.5 (underweight)	77 (12.0)	70 (10.7)	
18.5–24.9 (normal)	441 (68.9)	452 (68.8)	
25.0–29.9 (overweight)	102 (15.9)	108 (16.4)	
≥ 30.0 (obese)	14 (2.2)	23 (3.5)	
Body temperature (°C)	36.8 (36.5–37.2)	36.8 (36.5–37.2)	0.49
Blood-test results			
White blood cell (10 <sup>3</sup> /mm <sup>3</sup> )	10.7 (8.0–13.9)	10.3 (8.0–13.6)	0.28
Segmented neutrophil (%)	75.0 (64.0–82.0)	73.0 (63.0–82.0)	0.51
C-reactive protein (mg/dL)	0.7 (0.2–3.3)	0.6 (0.1–3.2)	0.64
Clinical risk scores for appendicitis			
Alvarado score			0.16
Low risk (0–4)	224 (35.0)	263 (40.0)	
Indeterminate risk (5–6)	208 (32.5)	202 (30.7)	
High risk (7–10)	207 (32.3)	191 (29.1)	
Appendicitis inflammatory response score			0.30
Low risk (0–4)	345 (53.9)	377 (57.4)	
Indeterminate risk (5–8)	281 (43.9)	265 (40.3)	
High risk (9–12)	8 (1.3)	12 (1.8)	
Time of CT examination			0.42
Working hours*	93 (14.5)	107 (16.3)	
After hours	547 (85.5)	550 (83.7)	
CT machine			0.92
16-channel	95 (14.8)	92 (14.0)	
64-channel	154 (24.1)	162 (24.7)	
128-channel	269 (42.0)	284 (43.2)	
256- or 640-channel	122 (19.1)	119 (18.1)	
Target effective dose (2-mSv CT vs. CDCT) <sup>†</sup>			NA
2 mSv vs. 3 mSv	0 (0.0)	0 (0.0)	
2 mSv vs. 5 mSv	2 (0.3)	2 (0.3)	
2 mSv vs. 6 mSv	121 (18.9)	138 (21.0)	
2 mSv vs. 7 mSv	288 (45.0)	294 (44.7)	
2 mSv vs. 8 mSv	229 (35.8)	223 (33.9)	
Individual radiation dose			
Dose-length product (mGy·cm)	131 (118–147)	482 (397–555)	< 0.001
Volume CT dose index (mGy)	2.6 (2.3–2.6)	9.1 (7.6–10.3)	< 0.001
Size-specific dose estimate (mGy)	4.0 (3.7–4.5)	14.3 (12.9–15.8)	< 0.001
Iterative reconstruction			< 0.001
Used	328 (51.3)	51 (7.8)	
Not used	312 (48.8)	606 (92.2)	
Site			0.25
2-mSv CT experience in the previous single-center trial			
Yes	100 (15.6)	87 (13.2)	

**Table 1.** Patient Characteristics (continued)

Characteristic	Patients Whose CTs Were Read Preliminarily by Residents		P
	2-mSv CT Group (n = 640)	CDCT Group (n = 657) <sup>‡</sup>	
No	540 (84.4)	570 (86.8)	
Annual number of appendectomies			0.97
< 150	12 (1.9)	12 (1.8)	
150–299	108 (16.9)	114 (17.4)	
300–449	181 (28.3)	178 (27.1)	
≥ 450	339 (53.0)	353 (53.7)	

Data are n (%) or median (interquartile range) unless otherwise indicated. For each characteristic, there were missing data in less than 2.5% of all randomized patients. \*0800–1700 h, working days, <sup>†</sup>Target effective dose for conventional-dose computed tomography (CDCT) was individualized for each computed tomography (CT) machine following the institutional normal dose, <sup>‡</sup>Included one board-certified on-call radiologist without fellowship training in abdominal radiology. NA = not applicable

**Table 2.** Diagnostic Performance of Residents

Endpoints	2-mSv CT Group (n = 621)	CDCT Group (n = 619)	Difference (95% CI)	P
Sensitivity	96.0% (216/225)	97.1% (235/242)	-1.1% (-4.9%–2.6%)	0.69
Specificity	93.2% (369/396)	93.1% (351/377)	0.1% (-3.6%–3.7%)	> 0.99

There were 19 and 38 patients with incomplete reference standard in the 2-mSv computed tomography (CT) and conventional-dose computed tomography (CDCT) group, respectively. CI = confidence interval

**Table 3.** Discrepancy between Preliminary CT Reports Made by Residents vs. Addendum Reports Made by Attending Radiologists

Discrepancy	2-mSv CT Group (n = 640)	CDCT Group <sup>†</sup> (n = 657)	Difference (95% CI)	P
<b>Appendicitis</b>				
Likelihood scores for appendicitis*	21 (3.3)	34 (5.2)	-1.9% (-4.2%–0.4%)	0.12
Negative index test result revised to positive result in addendum report	7 (1.1)	11 (1.7)	-0.6% (-2.0%–0.8%)	0.51
Likelihood score 1 or 2 revised to score 3	2 (0.3)	6 (0.9)		
Likelihood score 1 or 2 revised to score 4 or 5	5 (0.8)	5 (0.8)		
Positive index test result revised to negative result in addendum report	14 (2.2)	23 (3.5)	-1.3% (-3.3%–0.6%)	0.21
Likelihood score 3 revised to score 1 or 2	5 (0.8)	13 (2.0)		
Likelihood score 4 or 5 revised to score 1 or 2	9 (1.4)	10 (1.5)		
<b>Alternative diagnosis</b>	35 (5.5)	42 (6.4)	-0.9% (-3.6%–1.8%)	0.56
Significant discrepancy that may have required an emergency surgery	0 (0.0)	2 (0.3)	-0.3% (-1.1%–0.3%)	0.21

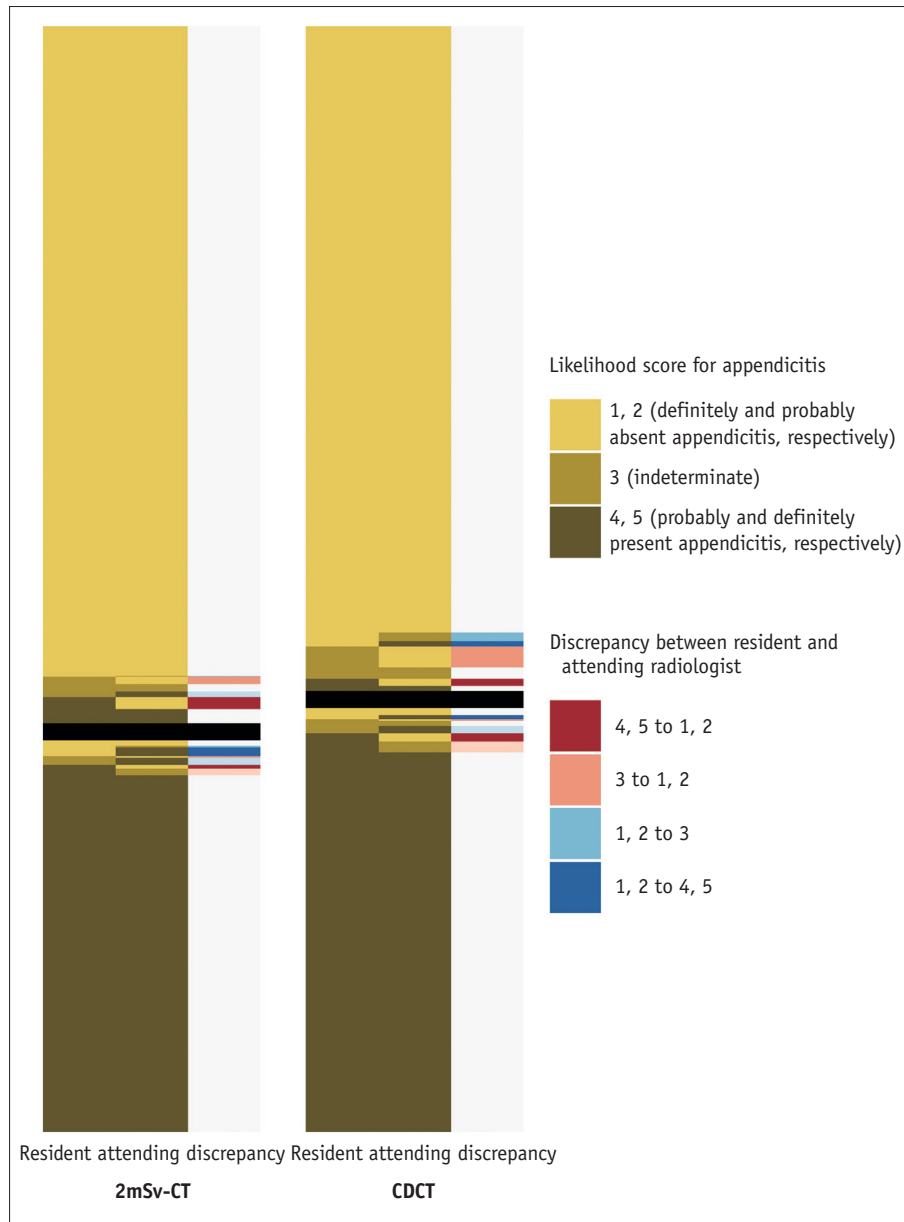
Data are n (%) (proportion out of patients whose computed tomographies [CTs] were read preliminarily by residents), unless specified otherwise. Likelihood score 3–5 was considered as index test positive for appendicitis. \*Likelihood score 1 and 2 indicated definitely and probably absent appendicitis, respectively; 3 indicated indeterminate; 4 and 5 indicated probably and definitely present appendicitis, respectively, <sup>†</sup>For one CT report that lacked addendum report, discrepancy was assumed as not present. CDCT = conventional-dose computed tomography, CI = confidence interval

patients were similar between both groups (Table 1). Most CT examinations that were preliminarily read by the residents were performed after hours (85.5% and 83.7% in the 2-mSv CT and CDCT groups, respectively). Reference standard information was available for 621 and 619 patients in the 2-mSv CT and CDCT groups, respectively (Fig. 1).

No significant difference was observed in the performance of residents in diagnosing appendicitis, both in terms of sensitivity (96.0% [216/225] vs. 97.1% [235/242]; difference [95% confidence interval {CI}], -1.1% [-4.9%, 2.6%]; *P* = 0.69) and specificity (93.2% [369/396] vs.

93.1% [351/377]; 0.1% [-3.6%, 3.7%]; *P* > 0.99) (Table 2, Supplementary Table 2) between the 2-mSv CT and CDCT groups. By contrast, the sensitivity and specificity of the attending radiologists based on their addendum reports were 97.3% and 95.7%, respectively, in the 2-mSv CT group, and 95.5% and 95.5%, respectively, in the CDCT group (Supplementary Table 3).

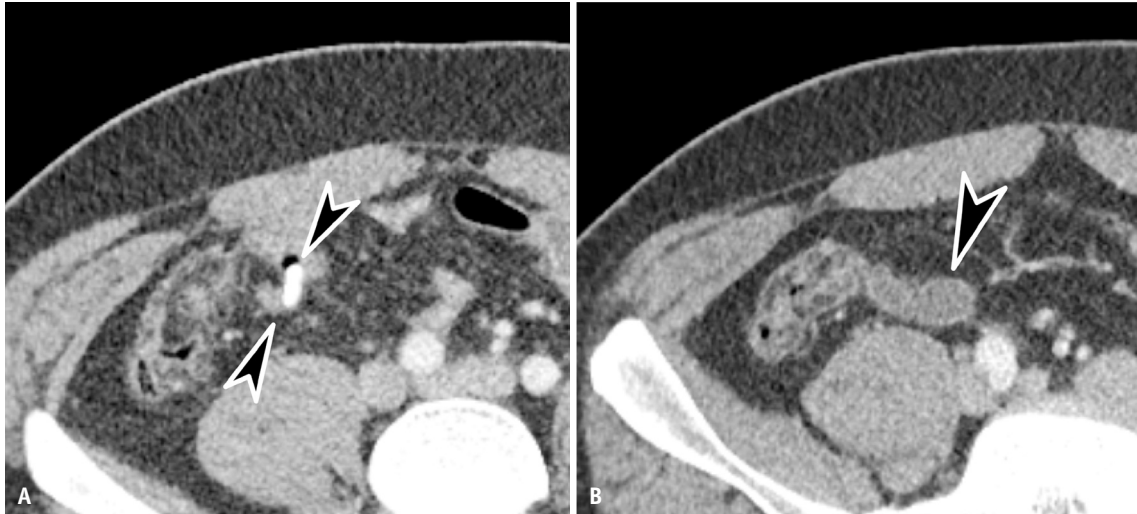
No significant between-group differences were noted for discrepancies in CT reports between residents and attending radiologists (Table 3, Supplementary Table 4, Fig. 2). Regarding the likelihood score for appendicitis,



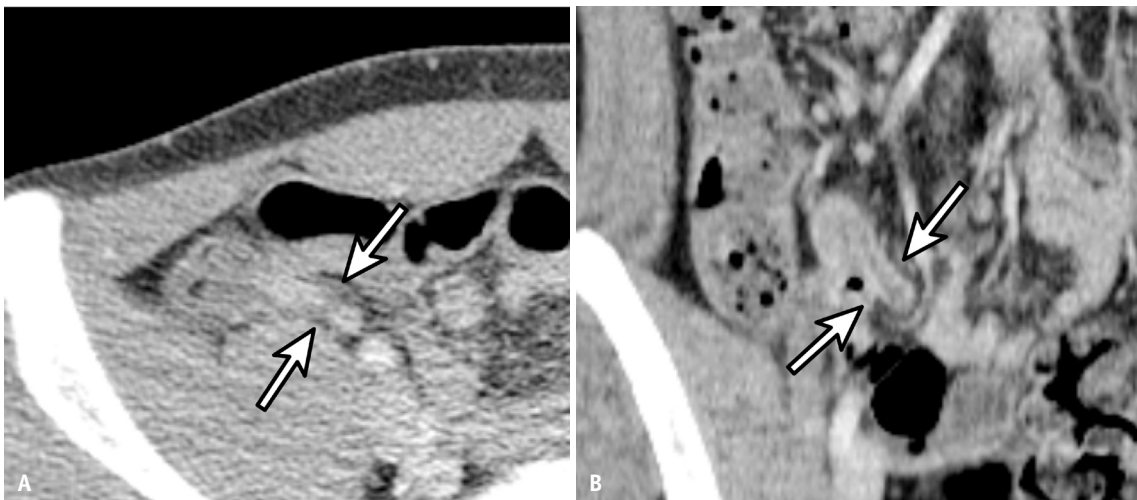
**Fig. 2.** Heat maps demonstrating results of likelihood for appendicitis. Likelihood for appendicitis rated by residents (first column) and attending radiologists (second column), and discrepancies between the two (third column). Each row represents datum of each patient. In patients with appendicitis as the final diagnosis (data are presented below the black spacing), discrepancies coded in blue shades refer to incorrect diagnoses made by residents. Meanwhile, in patients without appendicitis as the final diagnosis (data are presented above the black spacing), discrepancies coded in blue shades refer to incorrect diagnoses made by attending radiologists. CDCT = conventional-dose computed tomography, CT = computed tomography

discrepancies occurred in 21 of 640 reports (3.3%) on 2-mSv CT and in 34 of 657 reports (5.2%) on CDCT (difference [95% CI], -1.9% [-4.2%–0.4%];  $P = 0.12$ ). In both the 2-mSv CT and CDCT groups, the discrepancies were more often revisions of a positive index test result (i.e., likelihood score of 3, 4, or 5) to a negative index test result (i.e., score of 1 or 2) than vice-versa (Fig. 3). Many discrepancies in rating the likelihood score for appendicitis involved

a score of 3, which was defined as an indeterminate CT finding for appendicitis (7 in the 2-mSv CT group and 19 in the CDCT group). Some discrepancies were because of incorrect diagnoses made by the attending radiologists rather than by the residents (Fig. 4). Regarding alternative diagnoses, discrepancies occurred in 35 reports (5.5%) on 2-mSv CT and in 42 reports (6.4%) on CDCT (difference [95% CI], -0.9% [-3.6%–1.8%];  $P = 0.56$ ) (Table 3). Major



**Fig. 3.** A 24-year-old male with a final diagnosis of acute appendicitis. Contrast-enhanced 2-mSv 4-mm-thick transverse computed tomography (CT) images indicate (A) appendicolith (arrowheads), and (B) mild appendiceal dilatation (arrowhead). A resident-made preliminary CT report, where the likelihood score for appendicitis was rated as 2 (probably absent appendicitis). An attending radiologist revised the score as 5 (definitely present appendicitis) on the addendum report.



**Fig. 4.** A 16-year-old male with a final diagnosis of acute appendicitis. Contrast-enhanced 2-mSv 4-mm-thick transverse (A) and coronal (B) computed tomography (CT) images indicate mild appendiceal dilatation (arrows) with minimal periappendiceal fat infiltration. A resident made preliminary CT report, where the likelihood score for appendicitis was rated as 4 (probably present appendicitis). An attending radiologist revised the score as 2 (probably absent appendicitis) on the addendum report.

discrepancies in alternative diagnoses were rare in both the groups (0 cases in the 2-mSv CT group and 2 cases in the CDCT group).

No significant difference was observed in undesirable clinical outcomes following residents' reading (Table 4, Supplementary Table 5), including perforated appendicitis (12.0% [77/640] vs. 12.6% [83/657]; difference [95% CI], -0.6% [-4.3%–3.1%];  $P = 0.81$ ) and negative appendectomies (1.9% [12/640] vs. 1.1% [7/657]; difference [95% CI], 0.8% [-0.7%–2.3%];  $P = 0.33$ )

between the 2-mSv CT and CDCT groups. In terms of patient disposition, the need for extended surgery to treat appendicitis (1.4% [9/640] vs. 1.2% [8/657]; difference [95% CI], 0.2% [-1.2%–1.6%];  $P = 0.96$ ), and additional imaging tests (2.2% [14/640] vs. 2.6% [17/657]; -0.4% [-2.2%–1.4%];  $P = 0.77$ ) were not significantly different between the two groups. The interval between CT and discharge without surgery (median, 1.6 vs. 1.6 h in 372 and 354 patients, respectively,  $P = 0.83$ ), interval between CT and appendectomy (5.8 vs. 6.1 h in 237 and



**Table 4.** Undesirable Clinical Outcomes Following Residents' Reading

Undesirable Clinical Outcomes Following Residents' Reading	2-mSv CT Group (n = 640)	CDCT Group (n = 657)	Difference (95% CI)	P
Final diagnosis				
Perforated appendicitis	77 (12.0)	83 (12.6)	-0.6% (-4.3%–3.1%)	0.81
Negative appendectomies	12 (1.9)	7 (1.1)	0.8% (-0.7%–2.3%)	0.33
Patient disposition				
Extended surgery (beyond simple appendectomy)	9 (1.4)	8 (1.2)	0.2% (-1.2%–1.6%)	0.96
Delay in patient disposition				
Additional imaging test	14 (2.2)	17 (2.6)	-0.4% (-2.2%–1.4%)	0.77
Ultrasonography	13 (2.0)	15 (2.3)	-0.3% (-2.0%–1.5%)	0.90
CDCT	1 (0.2)	2 (0.3)	-0.1% (-1.0%–0.7%)	0.68
Interval between CT and discharge without surgery, hr*	1.6 (0.9–3.5)	1.6 (0.9–3.9)	0.0 (-0.2–0.2)	0.83
Interval between CT and appendectomy, hr	5.8 (3.1–10.9)	6.1 (3.6–12.5)	-0.3 (-1.2–0.3)	0.21
Length of hospital stay associated with appendectomy, day*	3.0 (2.4–3.9)	2.9 (2.3–3.7)	0.1 (-0.1–0.3)	0.47

Data are presented as n (%) (proportion calculated with all patients in either group as the denominator) or median (interquartile range). There were 19 and 38 patients with incomplete reference standard in the 2-mSv computed tomography (CT) and conventional-dose computed tomography (CDCT) group, respectively. \*In patients discharged without hospitalization, time of discharge was defined as the time of departure from the emergency department. For patients admitted to wards, time of discharge was defined as noon on the day of discharge. CI = confidence interval

249 patients, respectively,  $P = 0.21$ ), and length of hospital stay associated with appendectomy (3.0 vs. 2.9 d in 237 and 249 patients, respectively,  $P = 0.47$ ) also did not indicate a significant difference. In the 2-mSv CT and CDCT groups, 19 and 17 adverse events occurred, respectively, among patients whose CT scans were preliminarily read by residents (Supplementary Table 6). All adverse events resolved without any sequelae.

## DISCUSSION

In this retrospective analysis of a randomized pragmatic trial that included adolescents and young adults with suspected appendicitis, no significant difference was observed in the performance of residents in diagnosing appendicitis or the discrepancies between preliminary reports made by residents and addendum reports made by attending radiologists between the 2-mSv CT and CDCT groups. No significant differences were recorded in undesirable clinical outcomes following the residents' preliminary reading of CT examinations between the two groups.

These results have important clinical implications. First, it bolsters the evidence that 2-mSv CT can be used for diagnosing appendicitis after online training, regardless of the radiologist's experience level. The 2-mSv CT and CDCT groups did not demonstrate any significant differences in the diagnostic performance of residents or discrepancies between the preliminary and addendum reports. Second,

our study provides supporting evidence that unsupervised practice by radiology residents rarely leads to detrimental clinical outcomes when managing patients with suspected appendicitis. Our comparison between the preliminary and addendum reports revealed that discrepancies regarding the likelihood score for appendicitis were reasonably low (3.3% and 5.2% in the 2-mSv CT and CDCT groups, respectively). Based on visual assessment of our heat map, some of these discrepancies were incorrect diagnoses made by the attending radiologists. Moreover, major discrepancies in the alternative diagnoses were extremely rare (0 and 2 cases in the 2-mSv and CDCT groups, respectively). Measures of diagnostic performance, including sensitivity and specificity, were also similar between the preliminary and addendum reports. Whether radiology residents can diagnose appendicitis independently is an important clinical question that is closely relevant to many medical centers where reliance on unsupervised after-hours practice by residents is inevitable owing to limited resources. However, only a few studies with small sample sizes have provided data pertinent to this question, with results often discrepant from one another. To date, the largest previous study [31] included 322 patients whose CTs were read autonomously by residents, demonstrating that the performance of residents was not significantly different from that of more experienced radiologists. Meanwhile, a few small studies have suggested that the performance of radiologists in diagnosing appendicitis on CT is affected by

their experience level [18,32,33].

Most diagnoses made by residents regarding the presence of appendicitis were either true positive or true negative. Among incorrect diagnoses, false-positives were made more frequently than false-negatives. This tendency was also observed in the comparison between the preliminary and addendum reports, where it was more frequent for the addendum reports to negate the presence of appendicitis that had been suggested in the preliminary report (2.2% on 2-mSv CT and 3.5% on CDCT) rather than vice versa (1.1% on 2-mSv CT and 1.7% on CDCT). This reflects the residents' relative lack of diagnostic confidence and their concern regarding patients' immediate discharge following a false-negative diagnosis. False-positive diagnoses may not directly result in a negative appendectomy because some preparation time is generally required preoperatively, during which an attending radiologist may revise the diagnosis via an addendum report. Such a lack of diagnostic confidence is also reflected in the fact that a substantial portion of the discrepancies between preliminary and addendum reports involved an appendiceal likelihood score of 3 (indeterminate regarding the presence of appendicitis).

Although we observed that the sensitivity and specificity of the preliminary reports were similar to those of the addendum reports, we did not perform a formal statistical comparison between the preliminary and addendum reports. This was because some addendum reports may have been issued after patient disposition was already determined. For example, an attending radiologist may have issued an addendum report with the surgical report available for reference.

We did not perform non-inferiority or equivalence tests for our comparisons because such analyses are meaningless without a predetermined non-inferiority or equivalence margin before data collection [34,35]. Although no statistically significant between-group difference was present in any of the analyzed endpoints, we observed a slightly higher number of negative appendectomies performed in the 2-mSv CT group than those in the CDCT group, with 95% CI of the between-group difference straddling asymmetrically around zero (e.g., -0.7%–2.3%). This trend was also observed in the main trial [2] and is likely attributable to inherent differences in radiation dose rather than to differences in residents' performance relative to that of attending radiologists.

Our study has several limitations. First, despite the large-scale and pragmatic nature of the trial, the generalizability of our data may still be compromised to some extent.

The catchment area is limited to South Korea, where an extremely large body habitus is rare, CT is highly utilized, and non-surgical treatment for appendicitis is rarely performed. All participating sites were teaching hospitals. All potential radiologists were encouraged to complete an online training course on 2-mSv CT before the trial initiation. Second, our consideration of a likelihood score of 3 (indeterminate likelihood for appendicitis) as a positive index test result for appendicitis may have affected the diagnostic performance to some degree, particularly in the direction of inflating sensitivity. Third, verification bias may have affected the 2-mSv CT and CDCT groups differently as underestimation of incipient appendicitis with mild inflammation that resolves without appendectomy [36,37] is likely to occur more frequently with 2-mSv CT than that with CDCT. Fourth, although we used data from a randomized trial, our comparison was made between subsets of patients whose CT scans were initially read by residents in each of the 2-mSv CT and CDCT groups. To preserve between-group comparability achieved via randomization as much as possible, we counted the number of negative appendectomy and appendiceal perforation among all patients instead of using the conventional definition of "negative appendectomy rate" and "appendiceal perforation rate." Fifth, radiologists were not randomized between the 2-mSv CT and CDCT groups. However, because the radiologists at each site read 2-mSv CT or CDCT in their daily practice, we assumed that the radiologists in the two groups were virtually the same.

In conclusion, no significant differences were identified in the diagnostic performance and clinical outcomes between the 2-mSv CT and CDCT groups following the preliminary reading of CT examinations by radiology residents in patients with suspected appendicitis. Our study bolsters the evidence that 2-mSv CT can be used to diagnose appendicitis after online training even when the involved radiologists are residents.

## Supplement

The Supplement is available with this article at <https://doi.org/10.3348/kjr.2023.0015>.

## Availability of Data and Material

The datasets generated or analyzed during the current study are available on reasonable request. Please refer to the website; [www.locat.org/data\\_sharing](http://www.locat.org/data_sharing).

### Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

### Author Contributions

Conceptualization: Jungheum Cho, Hae Young Kim. Data curation: all authors. Formal analysis: Seungjae Lee. Funding acquisition: Kyoung Ho Lee. Investigation: all authors. Methodology: Jungheum Cho, Hae Young Kim. Project administration: Jungheum Cho, Hae Young Kim. Resources: all authors. Supervision: Jungheum Cho, Hae Young Kim. Writing—original draft: all authors. Writing—review & editing: all authors.

### ORCID iDs

Jungheum Cho

<https://orcid.org/0000-0003-3412-0863>

Hae Young Kim

<https://orcid.org/0000-0002-9508-4280>

Seungjae Lee

<https://orcid.org/0000-0001-5508-8634>

Ji Hoon Park

<https://orcid.org/0000-0002-6794-4909>

Kyoung Ho Lee

<https://orcid.org/0000-0001-6045-765X>

### Funding Statement

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1A2B5B01001804).

### Acknowledgments

The data in our study were obtained from the database of a randomized controlled trial by the LOCAT Group.

## REFERENCES

- Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med* 2012;366:1596-1605
- LOCAT Group. Low-dose CT for the diagnosis of appendicitis in adolescents and young adults (LOCAT): a pragmatic, multicentre, randomised controlled non-inferiority trial. *Lancet Gastroenterol Hepatol* 2017;2:793-804
- Rud B, Vejborg TS, Rappoport ED, Reitsma JB, Wille-Jørgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev* 2019;2019:CD009977
- Park JH, Salminen P, Tannaphai P, Lee KH. Low-dose abdominal CT for evaluating suspected appendicitis in adolescents and young adults: review of evidence. *Korean J Radiol* 2022;23:517-528
- Kim HJ, Lee KH, Kim MJ, Park SB, Ko Y; LOCAT Group. Using 2-mSv appendiceal CT in usual practice for adolescents and young adults: willingness survey of 579 radiologists, emergency physicians, and surgeons from 20 hospitals. *Korean J Radiol* 2020;21:68-76
- Hunter TB, Taljanovic MS, Krupinski E, Ovitt T, Stubbs AY. Academic radiologists' on-call and late-evening duties. *J Am Coll Radiol* 2007;4:716-719
- Bruno MA, Duncan JR, Bierhals AJ, Tappouni R. Overnight resident versus 24-hour attending radiologist coverage in academic medical centers. *Radiology* 2018;289:809-813
- Carney E, Kempf J, DeCarvalho V, Yudd A, Nosher J. Preliminary interpretations of after-hours CT and sonography by radiology residents versus final interpretations by body imaging radiologists at a level 1 trauma center. *AJR Am J Roentgenol* 2003;181:367-373
- Branstetter BF 4th, Morgan MB, Nesbit CE, Phillips JA, Lionetti DM, Chang PJ, et al. Preliminary reports in the emergency department: is a subspecialist radiologist more accurate than a radiology resident? *Acad Radiol* 2007;14:201-206
- Cooper VF, Goodhart LA, Nemcek AA Jr, Ryu RK. Radiology resident interpretations of on-call imaging studies: the incidence of major discrepancies. *Acad Radiol* 2008;15:1198-1204
- Kung JW, Melenevsky Y, Hochman MG, Didolkar MM, Yablon CM, Eisenberg RL, et al. On-call musculoskeletal radiographs: discrepancy rates between radiology residents and musculoskeletal radiologists. *AJR Am J Roentgenol* 2013;200:856-859
- Tomich J, Retrouvey M, Shaves S. Emergency imaging discrepancy rates at a level 1 trauma center: identifying the most common on-call resident "misses". *Emerg Radiol* 2013;20:499-505
- Issa G, Taslakian B, Itani M, Hitti E, Batley N, Saliba M, et al. The discrepancy rate between preliminary and official reports of emergency radiology studies: a performance indicator and quality improvement method. *Acta Radiol* 2015;56:598-604
- Weinberg BD, Richter MD, Champine JG, Morriss MC, Browning T. Radiology resident preliminary reporting in an independent call environment: multiyear assessment of volume, timeliness, and accuracy. *J Am Coll Radiol* 2015;12:95-100
- Mellnick V, Raptis C, McWilliams S, Picus D, Wahl R. On-call radiology resident discrepancies: categorization by patient location and severity. *J Am Coll Radiol* 2016;13:1233-1238
- Vaattovaara E, Nikki M, Nevalainen M, Ilmarinen M, Tervonen O. Discrepancies in interpretation of night-time emergency computed tomography scans by radiology residents. *Acta Radiol Open* 2018;7:2058460118807234
- Yang HK, Ko Y, Lee MH, Woo H, Ahn S, Kim B, et al. Initial performance of radiologists and radiology residents in interpreting low-dose (2-mSv) appendiceal CT. *AJR Am J Roentgenol* 2015;205:W594-W611
- Chang CC, Wong YC, Wu CH, Chen HW, Wang LJ, Lee YH, et al.

- Diagnostic performance on low dose computed tomography for acute appendicitis among attending and resident radiologists. *Iran J Radiol* 2016;13:e33222
19. Low-dose CT of Appendicitis Trial (LOCAT). What is LOCAT? LOCAT.com. Web site. <https://www.locat.org>. Published June 13, 2017. Accessed January 5, 2023
  20. Cho J, Kim Y, Lee S, Min HD, Ko Y, Chee CG, et al. Appendiceal visualization on 2-mSv CT vs. conventional-dose CT in adolescents and young adults with suspected appendicitis: an analysis of large pragmatic randomized trial data. *Korean J Radiol* 2022;23:413-425
  21. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P; CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295-309
  22. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390
  23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 2015;277:826-832
  24. Ahn S, LOCAT Group. LOCAT (low-dose computed tomography for appendicitis trial) comparing clinical outcomes following low- vs standard-dose computed tomography as the first-line imaging test in adolescents and young adults with suspected acute appendicitis: study protocol for a randomized controlled trial. *Trials* 2014;15:28
  25. Park SB, Kim MJ, Ko Y, Sim JY, Kim HJ, Lee KH; LOCAT Group. Structured reporting versus free-text reporting for appendiceal computed tomography in adolescents and young adults: preference survey of 594 referring physicians, surgeons, and radiologists from 20 hospitals. *Korean J Radiol* 2019;20:246-255
  26. Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology* 2000;215:337-348
  27. Pinto Leite N, Pereira JM, Cunha R, Pinto P, Sirlin C. CT evaluation of appendicitis and its complications: imaging techniques and key diagnostic findings. *AJR Am J Roentgenol* 2005;185:406-417
  28. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) v4.0. National Institutes of Health. com Web site. [http://ctep.cancer.gov/protocolDevelopment/electronic\\_applications/ctc.htm#ctc\\_40](http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_40). Published June 14, 2010. Accessed January 5, 2023
  29. Velanovich V, Satava R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. *Am Surg* 1992;58:264-269
  30. Mehrotra DV, Chan IS, Berger RL. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics* 2003;59:441-450
  31. Lietzén E, Salminen P, Rinta-Kiikka I, Paaanen H, Rautio T, Nordström P, et al. The accuracy of the computed tomography diagnosis of acute appendicitis: does the experience of the radiologist matter? *Scand J Surg* 2018;107:43-47
  32. Ceydeli A, Lavotshkin S, Yu J, Wise L. When should we order a CT scan and when should we rely on the results to diagnose an acute appendicitis? *Curr Surg* 2006;63:464-468
  33. Poortman P, Lohle PN, Schoemaker CM, Cuesta MA, Oostvogel HJ, de Lange-de Klerk ES, et al. Improving the false-negative rate of CT in acute appendicitis-reassessment of CT images by body imaging radiologists: a blinded prospective study. *Eur J Radiol* 2010;74:67-70
  34. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594-2604
  35. Park SH, Ahn S, Hong N, Park JH, Hwang SS, Lee KH. Quality of reporting noninferiority/similarity in research studies of diagnostic imaging. *Radiology* 2014;270:241-247
  36. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet* 2015;386:1278-1287
  37. Sippola S, Grönroos J, Sallinen V, Rautio T, Nordström P, Rantanen T, et al. A randomised placebo-controlled double-blind multicentre trial comparing antibiotic therapy with placebo in the treatment of uncomplicated acute appendicitis: APPAC III trial study protocol. *BMJ Open* 2018;8:e023623