



Diagnostic Performance and Interobserver Variability of Radiologists in CT-Scan Interpretation of Cases Suspected of Acute Appendicitis

Hadi Ahmadi Amoli¹, Abbas Naeef¹, Neda Nilforoushan^{1,2}, Hossein Zabihi Mahmoudabadi^{1,*}, Ehsan Rahimpour¹ and Amir Ashraf-Ganjouei^{1,2,**}

¹Department of Surgery, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran

²Students' Scientific Research Center, Tehran University of Medical Sciences, Tehran, Iran

*Corresponding author: Department of Surgery, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran. Email: hzabihim@tums.ac.ir

**Corresponding author: Students' Scientific Research Center, Tehran University of Medical Sciences, Tehran, Iran. Email: a-ganjouei@student.tums.ac.ir

Received 2019 October 07; Accepted 2020 February 08.

Abstract

Background: Acute appendicitis is one of the most common and sometimes life-threatening conditions in the Emergency Department referrals. Since suspected cases of acute appendicitis require immediate diagnosis and proper intervention, the computed tomography (CT) scan becomes the most frequently used modality for such conditions. However, due to the nature of emergency wards, gastrointestinal (GI) expert radiologists may not be always available.

Objectives: The current study aimed at comparing the interobserver variability of GI expert radiologists, general radiologists, and radiology residents in CT-scan interpretation of cases suspected of acute appendicitis.

Methods: Seventy patients suspected of acute appendicitis admitted to the Emergency Department of our university hospital were included in the study. CT-scan with intravenous contrast was performed on patients that their Alvarado score ranged 5 to 8. Decision for surgical or non-surgical management of patients was made by the routine treatment team of hospital and retrospectively, CT-scan images of all 70 patients were reported blindly by three groups of radiologists.

Results: Out of the 70 cases, 48 had positive confirmatory pathology for appendicitis (69%) and 22 had negative pathology report (31%). The sensitivity of the reports for radiology residents, general radiologists, and GI expert radiologists was 81.3%, 93.8% and 95.8%, respectively. The specificity of the diagnosis in the three groups was 72.7%, 86.4% and 81.8%, respectively.

Conclusions: The study results showed that although the interpretation was not perfect, radiology residents and general radiologists can provide reports with acceptable sensitivity and specificity in the emergency ward.

Keywords: Appendicitis, CT Scan

1. Background

Among the causes of acute abdomen, which is a life-threatening condition, appendicitis is the most common one. About 7% of the population may experience appendicitis during their lives, especially in the second to fourth decades (1). Therefore, it requires immediate diagnosis and proper intervention. Accurate preoperative diagnosis is required to minimize negative appendectomy, which ranges from 6.5% to 45% (2). The diagnostic approach includes patient's medical history, physical examination, laboratory tests, and radiological imaging among which the latter helps to confirm the diagnosis in suspected cases (3).

Three Imaging Modalities Are Available for the Diagnosis: Trans-abdominal sonography, computed tomography (CT)-scan and magnetic resonance imaging (MRI). Trans-abdominal ultrasound (US) is also a noninvasive and cost-

effective tool. Operator dependency, atypical position of appendix, pre-existing peritonitis, obesity, and the presence of intestinal gas are factors limiting the diagnostic information yielded by US (3). CT-scan with sensitivity of 89 - 99% and specificity of 89 - 99% is the most frequently used modality for the diagnosis (4). The common signs of appendicitis on CT-scan evaluation include enlarged appendiceal diameter of more than 6 mm, appendiceal wall thickness more than 2 mm, and abscess formation (5). Under some circumstances such as pregnancy, MRI should be performed instead of CT-scan. In this group of patients, sensitivity and specificity of MRI are 100% and 95%, respectively (6). Due to limited availability, higher costs and longer examination time, MRI is not a method of choice in cases suspected of acute appendicitis.

Although previous studies reported that interpreting CT-scan images by expert radiologists has high specificity

and sensitivity, suspected patients may refer to the Emergency departments when staff with limited experience is available. In 2009, Hof et al. (7), performed a study to compare interpersonal variability of CT-scan interpretation in suspected acute appendicitis. In their study, CT-scan images of patients were interpreted by three groups of radiologists: radiology residents, on-call radiologists, and gastrointestinal (GI) expert radiologists. They concluded that the reports provided by GI expert radiologists had the highest sensitivity and specificity. This implies that interpersonal variability should be taken into account when using CT-scan as a diagnostic tool.

2. Objectives

Therefore, the current study aimed at assessing radiologists' interobserver variability, based on their expertise level to diagnose cases suspected of appendicitis in Iran.

3. Methods

The current retrospective, cross sectional study was performed from 2016 to 2018 on 70 patients undergoing appendectomy in a university hospital. At first, patients suspected of acute appendicitis were admitted to the Emergency Department. The Alvarado score was calculated by on-call surgery residents based on anorexia, nausea, vomiting, and periumbilical pain prickled to right lower quadrant (RLQ), RLQ tenderness, RLQ rebound tenderness, fever, and leukocytosis. Ethical standards were observed while taking the medical history of present illness and physical examination; the written informed consent form was signed by the patient and a full description of the intervention was presented. After gathering demographic information, CT-scan with intravenous contrast was performed on the patients that their Alvarado scores ranged 5 to 8. Afterwards, final decision for surgical or non-surgical management of patients was made by the routine treatment team of hospital, based on positive radiological findings or worsening of clinical signs. In the cases that underwent surgery, removed appendixes were sent for histopathological examinations to the laboratory and positive report was considered as the diagnostic gold standard.

Retrospectively, CT-scan images of all 70 patients were reported blindly by three groups of radiologists; first group including second-year radiology residents; second group, general radiologists, and the third group, GI expert radiologists. Finally, histopathology reports were compared with the reports of the radiologists. The obtained data were analyzed using IBM SPSS version 22 (IBM Corp.,

Armonk, N.Y., USA) and to compare the three targeted groups, sensitivity and specificity of their diagnoses were calculated.

4. Results

The number of participants in the study was 70, of which 50 were male (71.4%) and 20 female (28.6%). The mean age \pm standard deviation of the patients was 20.56 ± 6.102 years; ranged 12 to 34. Of the 70 cases, 48 had positive confirmatory pathology for appendicitis (69%) and 22 had negative pathology report (31%). The number of CT scan images interpreted by radiology resident was 70, of which 45 had abnormal findings indicating appendicitis, and 25 were negative and normal. Out of the 70 CT scan images interpreted by general radiologists, 48 had abnormal findings and represented appendicitis and 22 were negative. Of the 70 CT scan images interpreted by the GI expert radiologist, 50 were abnormal cases represented appendicitis and 20 were normal (Table 1).

After statistical analyses, the sensitivity of the reports provided by radiology residents, general radiologists, and GI expert radiologists was 81.3%, 93.8%, and 95.8%, respectively. The specificity of the reports in the three groups was 72.7%, 86.4%, and 81.8%, respectively (Table 2). Based on the confidence intervals obtained, the difference in the interpretation of CT scan images was significant among radiology residents, general radiologists, and GI expert radiologists.

5. Discussion

Most of the diagnostic statistics regarding CT-scan imaging in acute appendicitis are obtained from studies investigating experienced radiologists reports. Since suspected patients refer to Emergency departments at any time of the day and the most experienced members of the medical team may not be available, the diagnostic accuracy of available members should be assessed. Therefore, the current study aimed at comparing the interobserver variability of interpretations, focusing on three different groups of radiology residents, general radiologists, and GI expert radiologists. The calculated sensitivities of diagnosis were 81.3%, 93.8%, and 95.8% and the specificities were 72.7%, 86.4% and 81.8%, respectively.

Considering appendicitis as the most common cause of surgical emergency and its possible complications, it requires prompt and precise preoperative diagnosis. Specific findings on illness history, physical examination, and laboratory test results guide clinicians to appendicitis. In an attempt to increase the benefit from clinical evaluations and

Table 1. CT-Scan Results of Each Observer Group (N = 70)

Observer Group	Pathology	Positive	Negative	Total
Radiology resident	Abnormal	39	9	48
	Normal	6	16	22
On call radiologist	Abnormal	45	3	48
	Normal	3	19	22
GI expert radiologist	Abnormal	46	2	48
	Normal	4	18	22

Table 2. Diagnostic Statistics Based on Different Observer Groups

Observer Group	Values, %	95% Confidence Interval
Sensitivity		
Radiology resident	81.3	67.4 - 91.1%
On call radiologist	93.8	82.8 - 98.7%
GI expert radiologist	95.8	85.7 - 99.5%
Specificity		
Radiology resident	72.7	49.8 - 89.3%
On call radiologist	86.4	65.1 - 97.1%
GI expert radiologist	81.8	59.7 - 94.8%
Positive predictive value		
Radiology resident	86.7	73.2 - 94.9%
On call radiologist	93.8	82.8 - 98.7%
GI expert radiologist	92	80.8 - 97.8%
Negative predictive value		
Radiology resident	64	42.5 - 82%
On call radiologist	86.4	65.1 - 97.1%
GI expert radiologist	90	68.3 - 98.8%

add weight to each finding, some scoring systems were developed, Alvarado score is one of the most popular ones (8). Physician's interpretation and practice setting may affect the accuracy of diagnosis and limit the use of these scores to risk stratification rather than a definitive diagnostic tool (9). It is thus far concluded that radiological imaging plays an inevitable role in the diagnosis of some diseases. Despite all the diagnostic advances, the rate of negative appendectomies still remains remarkable, ranging from 6.5% to 45% (2). Short-term and long-term post-surgical complications that constrain both patients and health systems, clarify the importance of accurate pre-surgical diagnosis.

Since CT-scan plays an important role in the diagnostic approach, several studies are designed to compare the accuracy of different protocols for CT-scan imaging including enhanced vs. unenhanced, and low dose vs. standard dose contrast (3). Intravenous contrast enhance-

ment shows pathognomonic findings for appendicitis or its complications, but in some other conditions such as renal insufficiency or allergic reactions that render contrast administration, the results are contradictory. Kim et al. (10), demonstrated that low-dose CT was equally acceptable as standard-dose CT. The resulted negative appendectomy associated with the low-dose radiation was 3.5% and for the standard-dose was 3.2%. In another study, Seo et al. (11) claimed that even unenhanced low-dose CT with sensitivity of 98.7% and specificity of 95.3% is as suitable as the intravenous-contrast standard-dose CT with the sensitivity of 100% and specificity of 93%. According to these results, in the current study, it was preferred to use the more widely accepted method, which is intravenous-enhanced standard-dose CT-scan.

Furthermore, Albano et al. (12), compared residents and faculty members reports to assess the CT-scan images of patients with acute appendicitis. They assessed 103 patients among which 96 reports were congruent between the two groups and all positive cases reported by residents were positive at surgery as well. They concluded that CT-scan reports of trained residents were matched well with those of faculty members, which can be safe and reliable. In another study similar to the current one, in't Hof et al. (7), compared interobserver variability in CT-scan reports of patients with acute appendicitis. They classified three groups of radiologists based on experience (group A, B, and C), group C the most experienced and group A the least experienced ones. The sensitivity of the reports was 81%, 88%, 95% and specificity was 94%, 94%, and 100% in A, B, and C groups, respectively. The current study compared the reports provided by radiology residents, general radiologists, and GI expert radiologists and the obtained sensitivities were 81.3%, 93.8%, 95.8% and specificities were 72.7%, 86.4%, and 81.8%, respectively. Based on the current study results, the most accurate group was GI expert radiologists, therefore, decision making upon their reports yields optimal results.

Considering the limitations of the current study, more details might have been investigated if residents were ar-

ranged based on the level of expertise or the study was performed in multiple academic hospitals as well. A comparison of similar interpretations between surgeons and radiologists might also help to answer the question that whether surgeons alone are reliable enough to make decision when there might be no expert radiologist accessible or not.

In conclusion, the diagnostic statistics of three different groups of radiologists were assessed, and it was concluded that although GI-expert radiologists were slightly better than the other groups, radiology residents and general radiologist can provide reports with acceptable sensitivity and specificity as well.

Footnotes

Authors' Contribution: Study concept and design: Hadi Ahmadi Amoli, Abbas Naej, Neda Nilforoushan, Hossein Zabihi Mahmoudabadi, Ehsan Rahimpour, and Amir Ashraf-Ganjouei. Acquisition of data and statistical analysis: Abbas Naej and Hossein Zabihi Mahmoudabadi. Drafting of the manuscript: Neda Nilforoushan and Amir Ashraf-Ganjouei. Critical revision of the manuscript: Hadi Ahmadi Amoli, Hossein Zabihi Mahmoudabadi, and Ehsan Rahimpour. Final approval of the manuscript: Hadi Ahmadi Amoli, Abbas Naej, Neda Nilforoushan, Hossein Zabihi Mahmoudabadi, Ehsan Rahimpour, and Amir Ashraf-Ganjouei.

Conflict of Interests: Nothing to declared.

Funding/Support: This piece of research did not receive any fund.

References

- Hosseini A, Omidian J, Nazarzadeh R. Investigating diagnostic value of ultrasonography in acute appendicitis. *Adv Biomed Res.* 2018;7:113. doi: 10.4103/abr.abr_79_18. [PubMed: 30123787]. [PubMed Central: PMC6071446].
- Wise SW, Labuski MR, Kasales CJ, Blebea JS, Meilstrup JW, Holley GP, et al. Comparative assessment of CT and sonographic techniques for appendiceal imaging. *AJR Am J Roentgenol.* 2001;176(4):933-41. doi: 10.2214/ajr.176.4.1760933. [PubMed: 11264081].
- Karul M, Berliner C, Keller S, Tsui TY, Yamamura J. Imaging of appendicitis in adults. *Rofa.* 2014;186(6):551-8. doi: 10.1055/s-0034-1366074. [PubMed: 24760428].
- Yu YR, Shah SR. Can the diagnosis of appendicitis be made without a computed tomography scan? *Adv Surg.* 2017;51(1):11-28. doi: 10.1016/j.yasu.2017.03.002. [PubMed: 28797333].
- Patel RR, Javors BR. Intramural vesicular fat—an uncommon CT finding. *Clin Imaging.* 2012;36(1):75-6. doi: 10.1016/j.clinimag.2011.04.015. [PubMed: 22226449].
- Wi SA, Kim DJ, Cho ES, Kim KA. Diagnostic performance of MRI for pregnant patients with clinically suspected appendicitis. *Abdom Radiol (NY).* 2018;43(12):3456-61. doi: 10.1007/s00261-018-1654-5. [PubMed: 29869102].
- in't Hof KH, Krestin GP, Steijgerberg EW, Bonjer HJ, Lange JF, Becking WB, et al. Interobserver variability in CT scan interpretation for suspected acute appendicitis. *Emerg Med J.* 2009;26(2):92-4. doi: 10.1136/emj.2008.058990. [PubMed: 19164615].
- Owen TD, Williams H, Stiff G, Jenkinson LR, Rees BI. Evaluation of the Alvarado score in acute appendicitis. *J R Soc Med.* 1992;85(2):87-8. [PubMed: 1489366]. [PubMed Central: PMC1294889].
- Chong CF, Adi MI, Thien A, Suyoi A, Mackie AJ, Tin AS, et al. Development of the RIPASA score: A new appendicitis scoring system for the diagnosis of acute appendicitis. *Singapore Med J.* 2010;51(3):220-5. [PubMed: 20428744].
- Kim SY, Lee KH, Kim K, Kim TY, Lee HS, Hwang SS, et al. Acute appendicitis in young adults: Low-versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. *Radiology.* 2011;260(2):437-45. doi: 10.1148/radiol.11102247. [PubMed: 21633052].
- Seo H, Lee KH, Kim HJ, Kim K, Kang SB, Kim SY, et al. Diagnosis of acute appendicitis with sliding slab ray-sum interpretation of low-dose unenhanced CT and standard-dose i.v. contrast-enhanced CT scans. *AJR Am J Roentgenol.* 2009;193(1):96-105. doi: 10.2214/AJR.08.1237. [PubMed: 19542400].
- Albano MC, Ross GW, Ditchek JJ, Duke GL, Teeger S, Sostman HD, et al. Resident interpretation of emergency CT scans in the evaluation of acute appendicitis. *Acad Radiol.* 2001;8(9):915-8. doi: 10.1016/s1076-6332(03)80772-9. [PubMed: 11724048].