CT of acute appendicitis: can diagnostic accuracy serve as a practical performance metric for readers specialized in abdominal imaging?

Lisa L. Chu, Emily M. Webb, Joseph W. Stengel, Benjamin M. Yeh, Ying Lu, Fergus V. Coakley *

Department of Radiology University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA, 94143–0628

ARTICLE INFO

ABSTRACT

Purpose: To investigate diagnostic accuracy for acute appendicitis at computed tomography (CT) as a performance metric for radiologists specialized in abdominal imaging. **Materials and Methods:** We retrospectively identified six attending abdominal imagers who each independently interpreted over 100 CT studies for suspected acute appendicitis. **Results:** The mean number of studies per reader was 311 (range, 129–386). Mean reader diagnostic accuracy was 95.0% (range, 91.4–97.1%). Only one had a diagnostic accuracy (91.4%) that was significantly lower than all others. **Conclusion:** Diagnostic accuracy for acute appendicitis at CT may be an impractical performance metric for radiologists specialized in abdominal imaging.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Aside from the Mammography Quality Standards Act in breast imaging, quality management and performance metrics have not been extensively reviewed in radiology subspecialties [1–9]. However, the monitoring and reporting of quality in healthcare delivery is becoming more common, given political and public pressure for greater transparency and accountability across all professions and the growing interest in performance-related payment for physicians [1,4,5,8,10–13]. The American College of Radiology now has a Metrics Committee that is devoted to creating metrics to objectively measure the quality of radiology practices [1,14]. Additionally, the Sun Valley Group, an independent think tank of radiologists and individuals experienced in quality improvement efforts, met for the first time in 2005 to develop methods to monitor and improve quality in radiology [15]. Against this backdrop, it has been suggested that diagnostic accuracy for acute appendicitis at computed tomography (CT) could be used as a performance metric for radiologists [5,16], since appendicitis is the commonest cause of an acute surgical abdomen [17,18], CT is performed in about 90% of all patients who undergo surgery for suspected appendicitis [19], and the final diagnosis is generally unequivocally and rapidly established as positive or negative by surgery or clinical outcome. As such, CT for suspected appendicitis lends itself to the objective assessment of interpretative performance [5]. Many studies have reviewed the diagnostic accuracy for acute appendicitis at CT [16,19–29], but to our knowledge, none of these studies have assessed whether this accuracy can be used as a potential performance metric. Therefore, we undertook this study to investigate if diagnostic accuracy for acute appendicitis at CT can serve as a practical performance metric for readers specialized in abdominal imaging.

2. Materials and methods

2.1. Study population

Our retrospective single-institution study was approved by our Committee on Human Research. The study was compliant with the Health Insurance Portability and Accountability Act, and written informed consent was waived. An electronic database search of radiology records was performed to identify all patients 18 years old and older who underwent CT scans performed for suspected appendicitis over a five year period. This database was initially developed for a previous study investigating outcome among patients with an apparent false positive diagnosis of acute appendicitis at CT [30]. The CT reports were sorted by faculty reader and only those reports from radiologists who had read over 100 cases for suspected appendicitis were included in the final study group. An electronic database search of pathology records was also performed to identify all patients 18 years old and older with non-incidental appendectomy specimens during the same interval, and these results were matched with the radiology results to determine surgical outcomes and identify those patients who did not go to surgery. Electronic patient notes were also reviewed to determine patient outcome. The final patient population consisted of 1865 patients, of which 691 were men and 1174 were women with a mean age of 43 years (range, 18–99).

2.2. CT technique

All patients were scanned with multidetector-row CT (4, 16, or 64 detectors; HighSpeed or LightSpeed; General Electric, Milwaukee, WI, USA) using 1.25 or 5 mm slice thickness. CT scans that were performed at 1.25 mm collimation were reconstructed to 5 mm slice thickness. The peak tube voltage was set at 120 kVp, and the mA was automatically adjusted to maintain a noise index of 12. Patients with contrast-enhanced CT ($n = 1661$) had 150 mL of intravenous iohexol (Omnipaque 350, Nycomed Amersham, Princeton, NJ, USA) injected at a rate of 3 to 5 mL/s. Patients routinely received oral contrast, consisting of 800 mL oral diatrizoate meglumine (Hypaque, Nycomed Amersham, Princeton, NJ) administered 45–90 min before the scan.

2.3. Data analysis

A single investigator (———), who was not one of the faculty radiologists being evaluated, classified the radiology reports as positive, negative, or equivocal for the diagnosis of acute appendicitis. Reports of a non-visualized appendix and no secondary signs of inflammation were considered negative. Diagnostic accuracy, sensitivity, and specificity for each reader were determined by comparison of CT reporting diagnoses with the reference standard of surgical pathology or patient outcome. For purposes of analysis, equivocal CT readings received a weight of 0.5 relative to true readings. Differences in diagnostic accuracy between readers were analyzed using a trinomial model and Z-test for statistical significance ($P < .05$). The minimum number of cases per reader needed to significantly differentiate readers by accuracy for appendicitis was determined by the normal approximation test and observed proportion of readers (alpha = .05, 1-beta = .80).

3. Results

The six readers specialized in abdominal imaging who each read over 100 CT studies for suspected acute appendicitis over the 5-year period of the study interpreted a total of 1865 CT studies, with a mean number of studies per reader of 311 (range, 129–386). Based on the reference standard, 413 (22%) of the patients had acute appendicitis and 1452 (78%) did not. Of the CT readings, 1313 (70%) were true negative, 388 (21%) were true positive, 141 (8%) were equivocal, 18 (1%) were false positive, and 5 (0.3%) were false negative. Accuracy results for the 6 readers are detailed in Tables 1. The mean reader diagnostic accuracy was 95.0% (range, 91.4–97.1). Only one reader had a diagnostic accuracy (91.4%) that was significantly lower than all the other readers, but the absolute differences from the other radiologists were only 2.2 to 5.7% (Table 1). The minimum number of cases per reader needed to significantly distinguish the least accurate reader from the most accurate reader was 155, whereas the minimum number of cases per reader needed to significantly distinguish the least accurate reader from the second least accurate
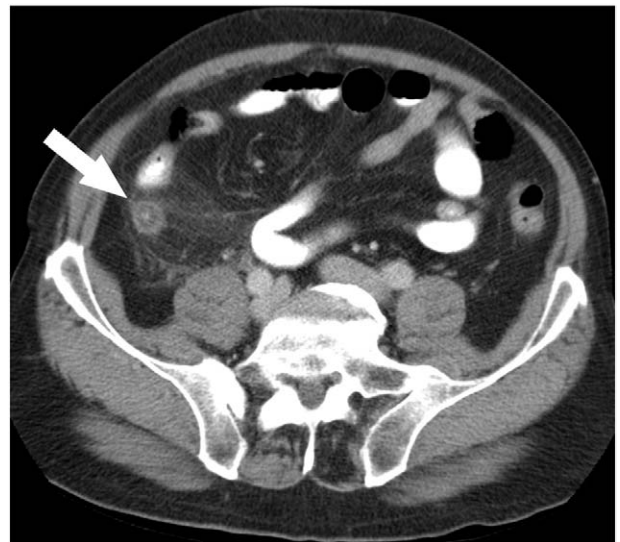

Fig. 1. Axial contrast-enhanced CT in a 45-year-old man with right lower quadrant pain and CT findings of appendiceal thickening and periappendiceal stranding (arrow). True positive diagnosis of acute appendicitis was surgically confirmed.

reader was 1404. Representative examples of positive and equivocal cases are shown in Figs. 1 and 2.

4. Discussion

Our results showed that readers specialized in abdominal imaging appear to have uniformly high accuracy in the diagnosis of acute appendicitis. Our mean CT diagnostic accuracy was 95%, and each reader's accuracy was above 90%, which is consistent with the CT diagnostic accuracy of acute appendicitis found in other studies [16,24,27,31]. To our knowledge, there are no published benchmark data on CT diagnostic accuracy of acute appendicitis, and since we believe our accuracy is similar to those of other abdominal imaging sections, a CT diagnostic accuracy for acute appendicitis in the range of 90–95% could potentially serve as an approximate l benchmark for experienced abdominal imaging subspecialists. Our results showed that reader accuracy differed by no more than 5.7% and the majority of differences between readers were not statistically significant. A previous study noted that interobserver variability and diagnostic
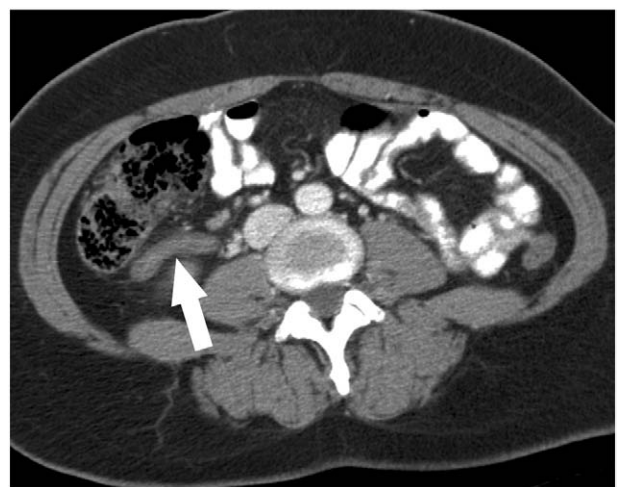

Fig. 2. Axial contrast-enhanced CT in a 57-year-old woman with acute abdominal pain shows borderline appendiceal thickening to a diameter of 10 mm (between arrows). CT findings were considered equivocal for the diagnosis of appendicitis. A normal appendix was found at pathology after surgical resection.

Table 1
Summary of diagnostic accuracy for acute appendicitis at CT for the six readers in the study

Reader	Accuracy (95% CI)	Significance of difference between readers				
		2	3	4	5	6
1	97.1% (95.7–98.5)	NS	NS	NS	$P < .01$	$P < .001$
2	96.5% (94.1–99.0)	–	NS	NS	NS	$P < .05$
3	95.9% (94.3–97.4)	–	–	NS	NS	$P < .01$
4	95.7% (94.1–97.3)	–	–	–	NS	$P < .01$
5	93.6% (91.3–95.8)	–	–	–	–	NS
6	91.4% (89.1–93.7)	–	–	–	–	–