Published in final edited form as:

AJR Am J Roentgenol. 2015 June ; 204(6): 1212–1219. doi:10.2214/AJR.14.13512.

A Novel Reporting System to Improve Accuracy in Appendicitis Imaging

Benjamin D. Godwin¹, Frederick T. Drake², Vlad V. Simianu², Jabi E. Shriki^{1,3}, Daniel S. Hippe¹, Manjiri Dighe¹, Sarah Bastawrous^{1,3}, Carlos Cuevas¹, David Flum², and Puneet Bhargava¹

¹Department of Radiology, University of Washington, 1959 NE Pacific St, UW Mailbox 357115, Seattle, WA 98195

²Department of Surgery, University of Washington, Seattle, WA

³VA Puget Sound Health Care System, Seattle, WA

Abstract

OBJECTIVE—The purpose of this study was to ascertain if standardized radiologic reporting for appendicitis imaging increases diagnostic accuracy.

MATERIALS AND METHODS—We developed a standardized appendicitis reporting system that includes objective imaging findings common in appendicitis and a certainty score ranging from 1 (definitely not appendicitis) through 5 (definitely appendicitis). Four radiologists retrospectively reviewed the preoperative CT scans of 96 appendectomy patients using our reporting system. The presence of appendicitis-specific imaging findings and certainty scores were compared with final pathology. These comparisons were summarized using odds ratios (ORs) and the AUC.

RESULTS—The appendix was visualized on CT in 89 patients, of whom 71 (80%) had pathologically proven appendicitis. Imaging findings associated with appendicitis included appendiceal diameter (odds ratio [OR] = 14 [> 10 vs < 6 mm]; $p = 0.002$), periappendiceal fat stranding (OR = 8.9; $p < 0.001$), and appendiceal mucosal hyperenhancement (OR = 8.7; $p < 0.001$). Of 35 patients whose initial clinical findings were reported as indeterminate, 28 (80%) had appendicitis. In this initially indeterminate group, using the standardized reporting system, radiologists assigned higher certainty scores (4 or 5) in 21 of the 28 patients with appendicitis (75%) and lower scores (1 or 2) in five of the seven patients without appendicitis (71%) (AUC = 0.90; $p = 0.001$).

CONCLUSION—Standardized reporting and grading of objective imaging findings correlated well with postoperative pathology and may decrease the number of CT findings reported as indeterminate for appendicitis. Prospective evaluation of this reporting system on a cohort of patients with clinically suspected appendicitis is currently under way.

Keywords

appendectomy; appendicitis; CT; standardized_reporting

Acute appendicitis is the most common cause of right lower quadrant pain presenting to the emergency department [1] and remains the most frequent indication for urgent abdominal surgery [2]. Before the advent of advanced imaging, the diagnosis of appendicitis was based on clinical symptoms and laboratory markers. Early surgical intervention was preferred, and negative appendectomy rates between 10–40% were routinely accepted to avoid delayed diagnosis and the risk of perforated appendicitis [3–5].

Over the past 20 years, CT has revolutionized the workup of appendicitis, with reported sensitivities and specificities of 88–100% [3–21]. Correspondingly, the number of negative laparotomies for appendicitis has decreased, and rates of less than 10% are considered the current standard of care [5]. Over the same period, the number of cases of perforated appendicitis has remained constant or decreased slightly in retrospective studies [22, 23]. However, the persistently high rate of negative laparotomies suggests the need for additional improvement in our diagnostic ability.

Imaging findings associated with appendicitis have been widely documented in the literature [1, 7, 11, 12, 18, 19, 24–26]. In most cases, CT simultaneously shows multiple findings, enabling confident diagnosis of appendicitis [10, 17]. However, in some cases, CT findings are equivocal or inconclusive, but there is high clinical suspicion for appendicitis. This subset of patients with indeterminate or equivocal imaging findings has been estimated to represent 9–13% of CT studies for workup of appendicitis, with up to 30% of these patients being subsequently diagnosed with appendicitis [17].

Because the diagnosis of appendicitis on CT is not always certain, a clear way of expressing the likelihood of appendicitis is needed. This can be achieved by linking imaging findings with certainty scores [15, 27]; categorizing imaging findings as positive, negative, or indeterminate for appendicitis [17, 28]; or weighing imaging findings on the basis of clinical concern [29]. In this study, we introduce a more systematic approach to the diagnosis of appendicitis on CT. This article represents a feasibility study of our standardized reporting format system, using a cohort of patients who underwent appendectomy at two institutions. We undertook this study to assess the diagnostic accuracy of the system, as well as the usefulness of specific imaging findings, and to evaluate its reproducibility.

Materials and Methods

This study was approved with a waiver of consent from the human subjects division of the University of Washington institutional review board.

Standardized Reporting System

Our standardized reporting system was based on consultation with surgical care providers, expert opinion, and literature review. Using these inputs, we developed a standardized reporting format with both objective and subjective components. The objective component

consists of a checklist of imaging findings previously described to have an association with the presence or absence of appendicitis (Appendix 1) [1, 7, 11, 12, 18, 19, 24–26]. The subjective component is a score of the radiologist's certainty of the presence or absence of appendicitis, ranging from 1 (appendicitis definitely absent) through 5 (appendicitis definitely present). The objective imaging findings convey the radiologist's decision-making process to the clinician, and the final score reflects the corresponding implications toward patient management.

Patient Population

Using the Washington State Surgical Care and Outcomes Assessment Program (SCOAP) database [30], we identified patients who underwent appendectomy over a 4-year period, from January 2008 to December 2012, at University of Washington Medical Center and Harborview Medical Center. Most hospitals in Washington State participate in SCOAP, which uses trained chart abstractors to review all laparotomies performed at participating hospitals. Data recorded for appendectomies include preoperative CT findings (positive, negative, or indeterminate for appendicitis), surgical findings, and pathologic information (positive or negative for appendicitis). Determination of whether a CT study was positive, negative, or indeterminate for appendicitis was based on finalized interpretations by body imaging fellowship-trained radiologists. SCOAP metrics and a data dictionary are available at a secure site [26, 30]. Medical records are reviewed for demographic and clinical variables.

The total number of patients who underwent appendectomy over the period of the study was 738. The mean age of this patient population was 35 years (range, 15–94 years). A stratified random sample of these patients was selected for inclusion in the study, where three strata were defined by pathology and CT findings. Specifically, all patients with negative findings for appendicitis on pathology ($n = 17$; 2.3%) were selected as control patients, and all patients with positive findings on pathology but negative ($n = 11$; 1.5%) or indeterminate ($n = 34$; 4.6%) findings on CT were also selected to determine if our standardized reporting system represents an improvement over current clinical practice. Last, we randomly selected 34 of 676 patients who were positive for appendicitis on both pathology and CT, for a total of 96 patients. The total of 96 was chosen to be evenly divisible among four readers, as described in the next section. After selection, medical records of all selected patients were reviewed to confirm the pathologic diagnoses.

Imaging Review

Two separate reviews were performed. For the primary review, preoperative CT scans from the selected patients were reviewed using the standardized reporting system by one of four radiologists. Subsequently, a second review was performed to evaluate interobserver reproducibility. For each reader, six patients whose scans were previously reviewed by another reader were randomly selected to be rereviewed using the standardized reporting system. Selection was restricted such that no patient was selected to be rereviewed by more than one reader. Thus, there were 96 patients in the primary review, of whose studies 24 (25%) were reread in the secondary review by different readers. During both reviews, the radiologists were blinded to the patients' initial imaging reports, pathologic diagnoses, and

other clinical findings. All of the radiologists who reviewed studies were board certified with additional subspecialty fellowship training in abdominal imaging. Radiologists also had between 5 and 10 years of experience in interpreting abdominal CT scans.

Transverse, coronal, and sagittal images were retrospectively reviewed. Radiologists recorded several imaging findings, including the diameter of the appendix, measured in the mid portion and at the tip, with instructions to measure the appendix in a short-axis plane; the thickness of the wall of the appendix, measured in the mid portion from the outer wall to the inner mucosa; the presence or absence of mucosal hyperenhancement (relative to small bowel); the presence or absence of adjacent fat stranding; the presence or absence of an appendicolith; the presence or absence of focal thickening of the cecum, at the base of the appendix; the presence or absence of extraluminal periappendiceal gas; the presence or absence of gas within the lumen of the appendix; and the presence or absence of a right lower quadrant collection, defined as an area of fluid attenuation with an enhancing wall. Radiologists also recorded whether the right ovary was visualized and whether there was an abnormality of the right ovary, such as a cyst or mass that might explain the patient's right lower quadrant pain. Radiologists were also asked to record their certainty of appendicitis using a score of 1–5, as described.

Statistical Analysis

Categorical variables were summarized as number (percent) of patients, and continuous variables were summarized as mean or median (range). Variables were compared between patients with and without appendicitis on pathology using the Fisher exact test (categorical variables) and the Mann-Whitney test (continuous and ordinal variables). The odds ratio (OR) was used to evaluate the strength of association between categorical variables and pathology. The AUC was used to evaluate how well each continuous or ordinal variable could discriminate between those with and without appendicitis on pathology. An AUC of 0.5 is equivalent to a random guess, whereas an AUC of 1.0 corresponds to perfect separation between groups. Precision was summarized using 95% CIs. Interreader agreement was assessed using Cohen's kappa and percent agreement. Percent agreement was calculated as 100 times the number of patients on whose scans readers were agreed divided by the total number of patients whose scans were read by two readers. Cohen's kappa is another measure of agreement that has been adjusted for the level of agreement expected by chance. A value of 0 indicates no agreement beyond what is expected by chance, whereas a value of 1 indicates exact agreement in all cases. All statistical calculations were conducted with the R statistical software package (version 2.14.1, R Foundation for Statistical Computing). Two-tailed tests were used, with $p < 0.05$ denoting statistical significance.

Results

Ninety-six patients with complete records of preoperative CT, surgical findings, and pathologic findings were selected from the SCOAP database. After reviewing the medical records of all 96 patients, our team corrected the data of nine patients for whom the SCOAP abstractors had inaccurately recorded pathology results, seven were reclassified from having

positive to having negative findings for appendicitis, and two were reclassified from having negative to having positive findings. During review, one patient was excluded owing to uninterpretable images, leaving 95 patients available for analysis. The mean age of the selected patients was 33 years (range, 18–72 years) (Table 1). Of these patients, 74 (78%) had pathologic confirmation of appendicitis, and 21 (22%) had normal appendixes at pathology. On the basis of the original CT findings, 15 reports (16%) were discordant with pathology, and 39 reports (41%) were indeterminate; thus, a total of 54 patients (57%) were not correctly diagnosed by the CT report.

The appendix could be visualized in 89 of 95 patients (94%). Clinical characteristics and objective findings were compared between these patients with and without appendicitis by pathology (Table 2). WBC counts were significantly higher in cases with positive findings (median, 12 vs $7 \times 10^9/L$; $p = 0.001$). Of the objective CT findings, outer-outer wall diameter (OR = 1.9 [6–10 vs < 6 mm], and OR = 14 [> 10 vs < 6 mm]; $p = 0.002$), tip diameter (OR = 2.7 [6–10 vs < 6 mm], and OR = ∞ [> 10 vs < 6 mm]; $p = 0.007$), and the presence of surrounding fat stranding or thickening of pararenal or lateroconal fascia (OR = 8.9; $p < 0.001$) all correlated significantly with the presence of appendicitis on pathology. The presence of focal cecal thickening (OR = 3.7; $p = 0.057$) was marginally significantly associated with positive findings on pathology. In the 82 patients who were imaged after contrast administration, the presence of mucosal hyperenhancement was also significantly associated with appendicitis (OR = 8.7; $p < 0.001$).

In the 89 patients in whom the appendix could be visualized, the AUC of the appendicitis score for diagnosis of appendicitis was 0.84 (95% CI, 0.74–0.94; $p < 0.001$) (Table 3). In the subset with indeterminate reports ($n = 35$), the AUC was 0.90 (95% CI, 0.80–0.99; $p = 0.002$), which was similar to the AUC in the group ($n = 40$) with reports concordant with pathology (AUC = 0.95 [95% CI, 0.88–0.99]; $p < 0.001$). In the subset for whom the original reports were discordant with pathology ($n = 14$), the AUC was 0.54 (95% CI, 0.22–0.86; $p = 0.8$).

Twenty-four patients were randomly selected to undergo a secondary review to evaluate interobserver reproducibility of our reporting system (Table 4). Despite having more than 50% agreement in all objective findings between readers (including $> 80\%$ reproducibility in findings of extraluminal periappendiceal air, the presence of a right lower quadrant fluid collection, and the presence of an appendicolith), the corresponding Cohen's kappa for the certainty score $\kappa = 0.30$ (95% CI, -0.03 to 0.65 ; $p = 0.07$), and the percent agreement was 58%.

Discussion

Our appendicitis standardized reporting format was created to increase accuracy in the diagnosis of appendicitis and to improve communication between radiologists and clinicians. The objective checklist includes imaging findings that are highly associated with appendicitis. In our study, increased appendiceal diameter, periappendiceal fat stranding, and appendiceal mucosal hyperenhancement had a statistically significant association with appendicitis on pathology. The final subjective score expresses the readers' degree of

certainty regarding the presence or absence of appendicitis and had good sensitivity and specificity for appendicitis in our study.

The need for standardization of reporting has been recognized since the inception of radiology [31]. It was not until the creation of the BI-RADS that this problem was widely and effectively addressed. BI-RADS was created to combine literature-derived objective findings with recommendation to clinicians for further management on the basis of the likelihood of malignancy. BI-RADS has shown that the standardization of terminology and management recommendations improves communication and patient outcomes [32]. The success of BI-RADS led to the creation of the Liver Imaging Reporting and Data System (LI-RADS), with the intention to standardize hepatic diagnostic imaging [33].

Communication of objective findings in a radiology report serves two purposes. First, radiologists are more accurate when systemically weighing the presence or absence of specific imaging findings before coming to a final conclusion. In a series of 53 CT examinations, Kim et al. [28] showed improved diagnostic accuracy for appendicitis cases with previously indeterminate findings when radiologists reevaluated the studies using a 4-point scale. Second, a clear list of findings documents the information the radiologist used to determine the final likelihood of appendicitis, allowing the referring clinician or surgeon to understand the radiologist's decision-making process.

Our results confirm that the CT findings of increased appendiceal body diameter, increased appendiceal tip diameter, mucosal hyperenhancement, and periappendiceal fat stranding or lateroconal fascia thickening were significantly associated with appendicitis. These findings have been frequently documented in the literature [1, 12, 15, 18, 34], and the present study adds to the available data. In our study, imaging findings suggesting perforated appendicitis—including periappendiceal air and a right lower quadrant fluid collection—were always associated with appendicitis at pathology, but the number of cases where these findings were present was too low to reach statistical significance. Similarly, there were insufficient cases of ovarian pathology to evaluate an association. The presence of appendiceal wall thickening and gas in the lumen of the appendix have previously been reported as having associations with appendicitis [8, 18, 24, 25, 34] but did not have a significant association in our study.

One encouraging result seen in this study is that, through the standardized reporting approach, the appendicitis score had a good correspondence with pathology, even in the 35 cases with indeterminate diagnoses on initial CT, where the AUC was 0.90. However, 13% (12 of 89) of our cases were assigned a score of 3, indeterminate for appendicitis. Most other published studies evaluating the diagnostic accuracy of CT have not included such a CT-indeterminate group [6, 10–12, 18, 24, 27, 35–38]. Other studies have also acknowledged the difficulty of decreasing the number of indeterminate cases, using the presence or absence of imaging findings [17, 26, 28]. Daly et al. [17] found that 10% of CT examinations will remain equivocal even after thorough review, similar to our results. Our readers used the standardized report to successfully reduce the number of CT scans deemed indeterminate while improving accuracy.

Despite the strong correlations with pathologic findings, as well as an incremental improvement in accuracy where initial clinical interpretations are indeterminate, correlation between multiple readers was variable in our study. Several objective imaging findings in our standardized report had high agreement between readers, including the presence of an appendicolith, focal cecal thickening, and the presence of gas within the appendix (percent agreement, 90%, 70%, and 85%, respectively). Agreement was also high on findings suggesting perforated appendicitis (including extraluminal periappendiceal gas and right lower quadrant fluid collections) and for the presence or absence of right ovarian abnormalities; however, because these imaging findings were uncommon, Cohen's kappa values were low ($\kappa = -0.05, 0.06, \text{ and } 0.62$, respectively). Interestingly, the presence of fat stranding or fascial thickening, mucosal hyperenhancement, and increased appendiceal diameter had only moderate interobserver agreement (percent agreement, 58%, 56%, and 60%, respectively) and low Cohen's kappa values ($\kappa = 0.17, 0.01, \text{ and } 0.36$, respectively). Given that these imaging findings have been reported as being associated with appendicitis both by several previous studies [7, 8, 10, 12, 15, 17–19, 24, 35, 36] and in our analysis, the lack of agreement is notable.

Furthermore, despite the high agreement in many objective findings, the agreement of subjective scores, assigned by our four readers, was only moderate, at 58% ($\kappa = 0.30$). This may reflect the subjective nature of our scoring system. In clinical settings, some of the variability in subjective scoring might be related to differences in detection of specific findings, but this did not appear to be a significant contributing factor in our analysis, given that readers reported objective findings with similar rates of detection. Mucosal hyperenhancement and periappendiceal fat stranding are often obvious, but in more subtle cases, radiologists will have differing thresholds for reporting a finding. Variation can be mitigated by establishing clearer guidelines. Future studies will also include several training cases, so that readers can practice using this reporting system.

There were several limitations to our study. First, because our study population was restricted to patients who underwent appendectomy, that introduces a selection bias compared with the population of patients being evaluated for possible appendicitis. Our patients may have been more likely to have clinical features of appendicitis and a high pretest probability, because all underwent laparotomy and appendectomy. This limitation was somewhat mitigated given that we preferentially included analysis of CT reports that were interpreted as being either indeterminate or negative for the presence of appendicitis. We chose to evaluate patients who underwent surgery as an initial test of the reporting system, so that comparison could be made to the pathologic reference standard of appendicitis. To assess the generalizability of this reporting system, we are currently evaluating it prospectively in a cohort of patients undergoing CT evaluation for the indication of ruling out appendicitis. Another potential limitation may arise from differences in reader experience; radiologists of different training levels provided the initial clinical reports, whereas all of the readers in our retrospective analysis were fellowship-trained radiologists. The improvement in accuracy that we found could be attributed to the use of a systematic approach to objective and subjective findings, or it could reflect the higher experience level of our readers [17, 25, 29]. Furthermore, the determination of whether a CT report is positive, negative, or indeterminate for appendicitis was made by chart review

performed by a non-radiologist and nonphysician abstractor as part of the SCOAP initiative. It is possible that CT findings considered as indeterminate for appendicitis by a SCOAP abstractor might have been considered as positive findings for appendicitis by a trained surgeon or emergency department physician. This potential pitfall in our study highlights the need for standardization of radiology reports and is a reason for the creation of our standardized format.

In summary, our standardized reporting system for appendicitis performed well in identifying cases of appendicitis. Future retrospective and prospective studies with larger patient populations will guide in the development of the system and will likely provide a large enough dataset for multivariate analysis of appendicitis-specific imaging findings. For now, our data suggest that the likelihood for appendicitis can be more confidently based on the presence or absence of specific imaging findings, rather than relying on overall gestalt. Although CT has revolutionized the work-up of appendicitis with a substantial drop in the number of missed appendicitis cases and the number of negative appendectomies [5, 7, 8, 11, 12, 15, 19, 24], further development and adoption of a standardized reporting format may continue to improve accuracy of appendicitis diagnosis, increase the quality of radiologist-clinician communication, and ultimately optimize patient management.

Acknowledgments

V. V. Simianu and F. T. Drake are supported by a National Institute of Diabetes and Digestive and Kidney Diseases training grant (5T32DK070555). The Surgical Care and Outcomes Assessment Program (SCOAP) is a Coordinated Quality Improvement Program of the Foundation for Health Care Quality (FHCQ).

References

1. Rosen MP, Ding A, Blake MA, et al. ACR Appropriateness Criteria® right lower quadrant pain: suspected appendicitis. *J Am Coll Radiol*. 2011; 8:749–755. [PubMed: 22051456]
2. Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. *Arch Surg*. 2002; 137:799–804. discussion, 804. [PubMed: 12093335]
3. Hale DA, Molloy M, Pearl RH, Schutt DC, Jaques DP. Appendectomy: a contemporary appraisal. *Ann Surg*. 1997; 225:252–261. [PubMed: 9060580]
4. Krajewski S, Brown J, Phang PT, Raval M, Brown CJ. Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. *Can J Surg*. 2011; 54:43–53. [PubMed: 21251432]
5. Drake FT, Flum DR. Improvement in the diagnosis of appendicitis. *Adv Surg*. 2013; 47:299–328.
6. Walker S, Haun W, Clark J, McMillin K, Zeren F, Gilliland T. The value of limited computed tomography with rectal contrast in the diagnosis of acute appendicitis. *Am J Surg*. 2000; 180:450–454. discussion, 454–455. [PubMed: 11182396]
7. Ives EP, Sung S, McCue P, Durrani H, Halpern EJ. Independent predictors of acute appendicitis on CT with pathologic correlation. *Acad Radiol*. 2008; 15:996–1003. [PubMed: 18620120]
8. Horton MD, Counter SF, Florence MG, Hart MJ. A prospective trial of computed tomography and ultrasonography for diagnosing appendicitis in the atypical patient. *Am J Surg*. 2000; 179:379–381. [PubMed: 10930484]
9. Stroman DL, Bayouth CV, Kuhn JA, et al. The role of computed tomography in the diagnosis of acute appendicitis. *Am J Surg*. 1999; 178:485–489. [PubMed: 10670858]
10. Kamel IR, Goldberg SN, Keogan MT, Rosen MP, Raptopoulos V. Right lower quadrant pain and suspected appendicitis: nonfocused appendiceal CT—review of 100 cases. *Radiology*. 2000; 217:159–163. [PubMed: 11012439]

11. Rao PM, Rhea JT, Novelline RA, et al. Helical CT technique for the diagnosis of appendicitis: prospective evaluation of a focused appendix CT examination. *Radiology*. 1997; 202:139–144. [PubMed: 8988203]
12. Rao PM, Rhea JT, Novelline RA. Sensitivity and specificity of the individual CT signs of appendicitis: experience with 200 helical appendiceal CT examinations. *J Comput Assist Tomogr*. 1997; 21:686–692. [PubMed: 9294553]
13. Raja AS, Wright C, Sodickson AD, et al. Negative appendectomy rate in the era of CT: an 18-year perspective. *Radiology*. 2010; 256:460–465. [PubMed: 20529988]
14. Raman SS, Kadell BM, Vodopich DJ, Sayer J, Cryer H, Lu DS. Patient gender-related performance of nonfocused helical computed tomography in the diagnosis of acute appendicitis. *J Comput Assist Tomogr*. 2003; 27:583–589. [PubMed: 12886148]
15. Jacobs JE, Birnbaum BA, Macari M, et al. Acute appendicitis: comparison of helical CT diagnosis focused technique with oral contrast material versus nonfocused technique with oral and intravenous contrast material. *Radiology*. 2001; 220:683–690. [PubMed: 11526267]
16. Drake FT, Florence MG, Johnson MG, et al. Progress in the diagnosis of appendicitis: a report from Washington State's Surgical Care and Outcomes Assessment Program. *Ann Surg*. 2012; 256:586–594. [PubMed: 22964731]
17. Daly CP, Cohan RH, Francis IR, Caoili EM, Ellis JH, Nan B. Incidence of acute appendicitis in patients with equivocal CT findings. *AJR*. 2005; 184:1813–1820. [PubMed: 15908536]
18. Choi D, Park H, Lee YR, et al. The most useful findings for diagnosing acute appendicitis on contrast-enhanced helical CT. *Acta Radiol*. 2003; 44:574–582. [PubMed: 14616200]
19. Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology*. 2000; 215:337–348. [PubMed: 10796905]
20. Flum DR, McClure TD, Morris A, Koepsell T. Misdiagnosis of appendicitis and the use of diagnostic imaging. *J Am Coll Surg*. 2005; 201:933–939. [PubMed: 16310698]
21. Pickhardt PJ, Lawrence EM, Pooler BD, Bruce RJ. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. *Ann Intern Med*. 2011; 154:789–796. [PubMed: 21690593]
22. Seetahal SA, Bolorunduro OB, Sookdeo TC, et al. Negative appendectomy: a 10-year review of a nationally representative sample. *Am J Surg*. 2011; 201:433–437. [PubMed: 21421095]
23. Teixeira PG, Demetriades D. Appendicitis: changing perspectives. *Adv Surg*. 2013; 47:119–140.
24. Moteki T, Horikoshi H. New CT criterion for acute appendicitis: maximum depth of intraluminal appendiceal fluid. *AJR*. 2007; 188:1313–1319. [PubMed: 17449776]
25. Cabarrus M, Sun YL, Courtier JL, Stengel JW, Coakley FV, Webb EM. The prevalence and patterns of intraluminal air in acute appendicitis at CT. *Emerg Radiol*. 2013; 20:51–56. [PubMed: 22996072]
26. Webb EM, Wang ZJ, Coakley FV, Poder L, Westphalen AC, Yeh BM. The equivocal appendix at CT: prevalence in a control population. *Emerg Radiol*. 2010; 17:57–61. [PubMed: 19597855]
27. Wise SW, Labuski MR, Kasales CJ, et al. Comparative assessment of CT and sonographic techniques for appendiceal imaging. *AJR*. 2001; 176:933–941.
28. Kim HC, Yang DM, Kim SW, Park SJ. Reassessment of CT images to improve diagnostic accuracy in patients with suspected acute appendicitis and an equivocal preoperative CT interpretation. *Eur Radiol*. 2012; 22:1178–1185. [PubMed: 22193372]
29. Blackmore CC, Terasawa T. Optimizing the interpretation of CT for appendicitis: modeling health utilities for clinical practice. *J Am Coll Radiol*. 2006; 3:115–121. [PubMed: 17412021]
30. SCOAP: Surgical Care and Outcomes Assessment Program. [Accessed March 3, 2015] www.SCOAP.org/
31. Reiner BI, Knight N, Siegel EL. Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol*. 2007; 4:313–319. [PubMed: 17467614]
32. Reiner BI. Medical imaging data reconciliation. Part 4. Reconciliation of radiology reports and clinical outcomes data. *J Am Coll Radiol*. 2011; 8:858–862. [PubMed: 22137004]

33. Purysko AS, Remer EM, Coppa CP, Leao Filho HM, Thupili CR, Veniero JC. LI-RADS: a case-based review of the new categorization of liver findings in patients with end-stage liver disease. *RadioGraphics*. 2012; 32:1977–1995. [PubMed: 23150853]
34. van Randen A, Lameris W, van Es HW, et al. Profiles of US and CT imaging features with a high probability of appendicitis. *Eur Radiol*. 2010; 20:1657–1666. [PubMed: 20119730]
35. Krishnamoorthi R, Ramarajan N, Wang NE, et al. Effectiveness of a staged US and CT protocol for the diagnosis of pediatric appendicitis: reducing radiation exposure in the age of ALARA. *Radiology*. 2011; 259:231–239. [PubMed: 21324843]
36. Seo H, Lee KH, Kim HJ, et al. Diagnosis of acute appendicitis with sliding slab ray-sum interpretation of low-dose unenhanced CT and standard-dose IV contrast-enhanced CT scans. *AJR*. 2009; 193:96–105. [PubMed: 19542400]
37. Soyer P, Dohan A, Eveno C, et al. Pitfalls and mimickers at 64-section helical CT that cause negative appendectomy: an analysis from 1057 appendectomies. *Clin Imaging*. 2013; 37:895–901. [PubMed: 23845254]
38. Weyant MJ, Eachempati SR, Maluccio MA, et al. Interpretation of computed tomography does not correlate with laboratory or pathologic findings in surgically confirmed acute appendicitis. *Surgery*. 2000; 128:145–152. [PubMed: 10922984]

TABLE 1

Patient Characteristics, Imaging Findings, and Appendicitis Score Godwin et al. for 95 Patients

Demographic and Clinical Parameters	Result	N
Sex		95
Male	53 (56)	
Female	42 (44)	
Age, y	30 (18–72)	95
Estimated weight, kg	71 (53–165)	85
Height, cm	171 (142–198)	66
WBC count, $\times 10^9/L$	11 (3–24)	95
Pathology and original CT group		95
Positive on pathology		
Positive imaging findings	35 (37)	
Indeterminate imaging findings	30 (32)	
Negative imaging findings	9 (9)	
Negative on pathology		
Positive imaging findings	6 (6)	
Indeterminate imaging findings	9 (9)	
Negative imaging findings	6 (6)	
Perforated appendicitis	8 (8)	95
Standardized report for appendicitis		
Appendix was visualized	89 (94)	95
Outer-outer wall diameter ^a , mm	10 (3–19)	89
< 6 mm	14 (16)	
6–10 mm	36 (40)	
> 10 mm	39 (44)	
Tip diameter ^a , mm	8 (2–23)	89
< 6 mm	27 (30)	
6–10 mm	45 (51)	
> 10 mm	17 (19)	
Single wall thickness ^a , mm	3 (1–7)	87
3 mm	36 (41)	
> 3 mm	51 (59)	
Mucosal hyperenhancement ^{a,b}	56 (68)	82
Surrounding fat stranding or thickening of pararenal or lateroconal fascia	66 (69)	95
Appendicolith ^a	28 (31)	89
Focal cecal thickening at base	35 (37)	94
Periappendiceal air	7 (8)	92
Right lower quadrant fluid collection, phlegmon, or abscess	9 (9)	95
Gas in lumen of appendix ^a	30 (34)	89
Right ovary abnormality	4 (4)	93

Demographic and Clinical Parameters	Result	<i>N</i>
Appendicitis score		95
1	15 (16)	
2	7 (7)	
3	12 (13)	
4	10 (11)	
5	51 (54)	

Note—Except where otherwise indicated, data are given as no. (%) or median (range).

^aEvaluated only in cases where appendix was visualized.

^bExcluding seven patients who were not administered contrast material.

Associations Between Clinical Characteristics and Objective Imaging Findings Among 89 Patients With and Without Appendicitis in Whom Appendix Was Visualized

TABLE 2

Demographics and Clinical Results	Result		Odds Ratio or AUC ^a (95% CI)	p ^b	N
	Positive Findings (n = 71)	Negative Findings (n = 18)			
Male sex	39 (55)	10 (56)	1.0 (0.3–3.1)	> 0.99	89
Age, y	30 (18–72)	28 (19–53)	0.56 (0.42–0.70)	0.43	89
Weight, kg	72 (53–134)	71 (54–134)	0.52 (0.37–0.66)	0.84	81
Height, cm	170 (142–198)	172 (150–193)	0.53 (0.36–0.70)	0.71	63
WBC count	12 (3–24)	7 (5–16)	0.75 (0.62–0.87)	0.001	89
Outer-outer wall diameter				0.002	89
< 6 mm	8 (11)	6 (33)	Reference		
6–10 mm	26 (37)	10 (56)	1.9 (0.5–7.1)		
> 10 mm	37 (52)	2 (11)	14 (2.4–82)		
Tip diameter				0.007	89
< 6 mm	17 (24)	10 (56)	Reference		
6–10 mm	37 (52)	8 (44)	2.7 (0.9–8.1)		
> 10 mm	17 (24)	0 (0)	∞		
Single wall thickness > 3 mm	31 (44)	5 (31)	1.7 (0.5–6.9)	0.41	87
Mucosal hyperenhancement ^c	51 (78)	5 (29)	8.7 (2.3–36)	< 0.001	82
Surrounding fat stranding or thickening of pararenal or lateroconal fascia	58 (82)	6 (33)	8.9 (2.5–34)	< 0.001	89
Appendicolith	24 (34)	4 (22)	1.8 (0.5–8.2)	0.41	89
Focal cecal thickening at base	30 (42)	3 (17)	3.7 (0.9–21)	0.057	89
Periappendiceal air	5 (7)	0 (0)	∞	0.58	89
Right lower quadrant fluid collection, phlegmon, or abscess	7 (10)	0 (0)	∞	0.34	89
Gas in lumen of appendix	26 (37)	4 (22)	2.0 (0.6–9.3)	0.28	89
Absence of right ovary abnormality ^d	29 (94)	5 (71)	5.8 (0.3–92)	0.15	38

Note—Except where otherwise indicated, data are given as no. (%) or median (range).

^a Odds ratio was calculated for binary variables and AUC for continuous variables.

^b Fisher exact test was used for binary variables.

Author Manuscript

^cIn only those patients administered contrast material.
^dIn female patients only.

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 3
ROC Analysis of Appendicitis Score for 89 Patients in Whom Appendix Was Visualized and Subgroups

Group	Appendicitis Score						AUC (95% CI)	p
	Mean (Range)	1	2	3	4	5		
All cases (N = 89)							0.84 (0.74–0.94)	< 0.001
Positive (n = 71)	4.3 (1–5)	6 (8)	1 (1)	8 (11)	8 (11)	48 (68)		
Negative (n = 18)	2.3 (1–5)	6 (33)	5 (28)	4 (22)	1 (6)	2 (11)		
Original report concordant with pathology (N = 40)							0.95 (0.88–0.99)	< 0.001
Positive (n = 35)	4.6 (1–5)	1 (3)	1 (3)	2 (6)	2 (6)	29 (83)		
Negative (n = 5)	2.0 (1–4)	2 (40)	2 (40)	0 (0)	1 (20)	0 (0)		
Original report indeterminate (N = 35)							0.90 (0.80–0.99)	0.001
Positive (n = 28)	4.1 (1–5)	3 (11)	0 (0)	4 (14)	5 (18)	16 (57)		
Negative (n = 7)	1.9 (1–3)	3 (43)	2 (29)	2 (29)	0 (0)	0 (0)		

Note—Except where otherwise indicated, data are given as no. (%).

TABLE 4
Interreader Agreement of the Standardized Reporting Format in 24 Patients Randomly Selected for Secondary Review

Parameter	Result		κ (95% CI)	% Agreement	N ^a
	Primary Reader	Secondary Reader			
Appendicitis score					
1–2	8 (33)	10 (42)	0.30 (–0.03 to 0.63)	58	24
3	3 (12)	3 (12)			
4–5	13 (54)	11 (46)			
Appendix visualized	22 (92)	21 (88)	0.33 (–0.37 to 0.99)	88	24
Outer-outer wall diameter			0.36 (0.02–0.70)	60	20
< 6 mm	5 (25)	3 (15)			
6–10 mm	9 (45)	11 (55)			
> 10 mm	6 (30)	6 (30)			
Tip diameter			0.20 (–0.18 to 0.59)	55	20
< 6 mm	6 (30)	6 (30)			
6–10 mm	11 (55)	12 (60)			
> 10 mm	3 (15)	2 (10)			
Single wall thickness > 3 mm	7 (39)	3 (17)	–0.04 (–0.58 to 0.50)	56	18
Mucosal hyperenhancement ^b	13 (72)	11 (61)	0.01 (–0.50 to 0.52)	56	18
Surrounding fat stranding or thickening of pararenal or lateroconal fascia	13 (54)	11 (46)	0.17 (–0.22 to 0.56)	58	24
Appendicolith	5 (25)	5 (25)	0.73 (0.38–0.99)	90	20
Focal cecal thickening at base	9 (39)	6 (26)	0.32 (–0.10 to 0.74)	70	23
Periappendiceal air	1 (4)	1 (4)	–0.05 (–1.43 to 0.99)	91	23
Right lower quadrant fluid collection, phlegmon, or abscess	1 (4)	2 (8)	–0.06 (–1.18 to 0.99)	88	24
Gas in lumen of appendix	5 (25)	6 (30)	0.62 (0.23–0.99)	85	20
Right ovary abnormality ^c	1 (9)	2 (18)	0.62 (–0.09 to 0.99)	91	11

Note—Except where otherwise indicated, data are given as no. (%).

^aNo. of patients for whom both readers reported measurements.

^bIn only those patients administered contrast material.

^cIn female patients only.

APPENDIX 1**Standardized Imaging Report for Appendicitis**

1. Appendix visualized?	☐ Yes	☐ No
2. Outer-outer wall diameter	_____ mm	
3. Tip diameter	_____ mm	
4. Single wall thickness	_____ mm	
5. Mucosal hyperenhancement?	☐ Yes	☐ No
6. Surrounding fat stranding or thickening of pararenal or lateroconal fascia?	☐ Yes	☐ No
7. Appendicolith?	☐ Yes	☐ No
8. Focal cecal thickening at base of appendix?	☐ Yes	☐ No
9. Periappendiceal air?	☐ Yes	☐ No
10. Right lower quadrant fluid collection, phlegmon, or abscess?	☐ Yes	☐ No
11. Gas in lumen of appendix?	☐ Yes	☐ No
12. Right ovary (in female patients)	☐ Normal	☐ Abnormal
13. Certainty score	☐ 1, Appendicitis definitely absent ☐ 2, Appendicitis unlikely ☐ 3, Indeterminate ☐ 4, Appendicitis likely ☐ 5, Appendicitis definitely present	