Published in final edited form as:

Curr Probl Diagn Radiol. 2017 ; 46(4): 267–274. doi:10.1067/j.cpradiol.2016.07.006.

The Reliability of a Standardized Reporting System for the Diagnosis of Appendicitis

Vlad V. Simianu, MD, MPH^a, Anna Shamitoff, MSN^a, Daniel S. Hippe, MS^b, Benjamin D. Godwin, MD^b, Jabi E. Shriki, MD^b, Frederick T. Drake, MD, MPH^a, Ryan B. O'Malley, MD^b, Suresh Maximin, MD^b, Sarah Bastawrous, DO^b, Mariam Moshiri, MD^b, Jean H. Lee, MD^b, Carlos Cuevas, MD^b, Manjiri Dighe, MD^b, David Flum, MD, MPH^a, and Puneet Bhargava, MD^{b,*}

^aDepartment of Surgery, University of Washington, Seattle, WA

^bDepartment of Radiology, University of Washington, Seattle WA

Abstract

Purpose—Computed tomography (CT) is a fast and ubiquitous tool to evaluate intra-abdominal organs and diagnose appendicitis. However, traditional CT reporting does not necessarily capture the degree of uncertainty and indeterminate findings are still common. The purpose of this study was to evaluate the reproducibility of a standardized CT reporting system for appendicitis across a large population and the system's impact on radiologists' certainty in diagnosing appendicitis.

Methods—Using a previously described standardized reporting system, eight radiologists retrospectively evaluated CT scans in a cohort of 561 patients imaged for possible appendicitis (2010–2014). Based on the original reports, 159 of these scans were selected to be over-read by a second reviewer. Reproducibility of reporting and certainty in the diagnosis of appendicitis was compared between the radiologists, and correlation made with clinical diagnoses.

Results—Appendicitis occurred in 20% of patients (mean age 38, 67% male) in the overall cohort (n=113 of 561). Using the standardized report, radiologists were highly accurate at identifying appendicitis (AUC=0.968, 95%CI: 0.95, 0.99), which occurred in 48% of selected exams (113 of 237). Inter-reader agreement was >80% for most objective findings, and certainty in diagnosing appendicitis was high and reproducible (AUC=0.955 and AUC=0.936 for primary readers and over-readers, respectively). In the subgroup of patients with previously “indeterminate” reports, 57% (24 of 42) were assigned the correct diagnosis using the standardized report.

Conclusions—Using a standardized reporting system resulted in high reproducibility of objective CT findings for appendicitis and achieved high diagnostic accuracy in an at-risk population. Predictive tools based on this reporting system may further improve communication about certainty in diagnosis and guide patient management, especially when CT findings are indeterminate.

---

*Corresponding author: Puneet Bhargava, MD, Department of Radiology, University of Washington, 1959 NE Pacific St, Room BB308, Box 357115, Seattle, WA 98195-7115. bhargp@uw.edu.

Introduction

Appendicitis is the leading cause for acute abdominal surgery, with an estimated lifetime risk of 7–12% and nearly 300,000 appendectomies performed annually in the United States (1,2). Historically, clinicians maintained a high level of suspicion and routinely offered early operative intervention in the hopes of avoiding high-risk complications such as perforation and/or intra-abdominal sepsis. With this approach, negative appendectomy rates of 15–25%, and up to 50% in pregnant patients, were considered acceptable (3,4). The widespread adoption of computed tomography (CT) has significantly reduced the rate of negative appendectomies, with the standard of care currently below 10% in most surgical practices (5,6).

Still, while CT provides fast abdominal imaging and improves accuracy in the diagnosis of appendicitis, there is no standardized reporting system to communicate the likelihood of appendicitis based on imaging findings. Up to 13% of CT scans are interpreted as indeterminate, with up to 73% of these patients subsequently diagnosed with appendicitis (7–11). In Washington State, where our group tracks appendectomy indications and outcomes, over 3300 appendectomies were performed in 2014. Of these, surgical pathology examination revealed that 120 were negative appendectomies (3.6%). Of these negative appendectomies, 83% (90 of the 108 that had pre-operative CT imaging) having discordant CT reports (scans that were reported as either indeterminate or positive for appendicitis (12).

CT offers diagnostic advantages over other imaging modalities (13) and clinical scoring systems (14–16) by demonstrating findings that can be reliably interpreted. Most often, CT demonstrates the presence or absence of multiple findings (9,17–22), that can support or refute a diagnosis of appendicitis (9,23). When this is not the case, as in the 13% of patients whose CT imaging is reported to be “indeterminate”, there are no established guidelines to help radiologists decide which imaging findings can best establish the diagnosis. Thus there is variation in appendicitis reporting, with some readers failing to include all of the pertinent findings that may be helpful in stratifying the risk of appendicitis. This is particularly relevant as the contemporary management of appendicitis is increasingly considering an antibiotics-first approach (24), and CT findings may identify antibiotics “non-responders.” While the antibiotics-first approach to appendicitis is still somewhat controversial, standardizing the reporting of appendicitis may help determine which findings should be considered important for determination of “non-complicated appendicitis – e.g. absence of appendicolith). However, before specific findings can be linked to clinical management recommendations, it is paramount to determine whether standardized enumeration of imaging findings improves certainty in diagnosing appendicitis, and whether this is reproducible among multiple radiologists. The objective of this study was to evaluate the reproducibility of a previously-described a standardized reporting system (25) across by multiple radiologists in a larger population undergoing imaging for possible appendicitis and to determine the impact of this standardized reporting format on certainty in the diagnosis of appendicitis.

Materials and Methods

This study was approved by the Human Subjects Division at the University of Washington institutional review board.

Study population

Using the imaging database for the University of Washington and Harborview Medical Centers, a retrospective cohort of patients that received CT imaging for clinical suspicion of appendicitis was assembled. All adult patients (age range 17–88 years) with CT scans performed between 6/1/2010–6/1/2014 for the indications of “right lower quadrant (RLQ) pain”, “rule out (r/o) appendicitis”, “r/o appy”, or “r/o appi” were considered eligible for inclusion. Two independent reviewers (AS, VS) reviewed all of the CT reports in this cohort. Patients were excluded if they had previous clinical and/or radiologic evidence of appendectomy or trauma as indication for imaging (n=56). In one patient who underwent 2 scans in the same day, only the first scan was included in this study. The final cohort of CT imaging reports (n=561) were categorized by the reviewers as positive (n=97), negative (n=421), or indeterminate (n=42) for appendicitis. Reports were considered “indeterminate” if they explicitly stated that the findings were “indeterminate,” “equivocal” or if they stated the available findings could not reliably diagnose or refute the diagnosis of appendicitis. A cohort size of >500 was selected because to ensure a sufficiently large number of indeterminate scans would be included, which was estimated at approximately 10% (or approximately 50 patients) based on the incidence of indeterminate reports in existing literature.

Imaging review

For feasibility of imaging review, the CT images corresponding to all positive (n=97) and indeterminate (n=42) reports, as well as the randomly-selected sample of the negative CT scans (n=98, to match the number of positive exams), a total of 237 scans, were reviewed using a standardized reporting system by 8 body-fellowship trained board-certified radiologists (Appendix 1). Our group has previously described this reporting system that includes 12 evidence-based objective findings and includes a 5-point assessment score for certainty of appendicitis diagnosis, also enumerated in Appendix 1 (25). The certainty in diagnosis is not linked to any individual findings, rather it reflects the subjective certainty in the diagnosis after enumeration of the imaging findings, and is up to the discretion of individual readers. Acknowledging the time burden of imaging review, all CT scans with indeterminate reports, and a randomly selected sample of positive and negative reports, were read by two radiologists, such that each radiologist reviewed approximately 50 CT scans consisting of a balanced proportion of previously positive-, negative-, and indeterminate-reported studies. This ensured that no reader saw a disproportionate number of positive or negative scans. Radiologists reviewed studies independently and were blinded to patient identifiers and reports. All reading sessions were recorded in a uniform fashion, with the reader having access to all 2.5 mm-thick intravenous contrast-enhanced axial sections reconstructed from images acquired using 0.625 mm collimation along with 5-mm thick coronal, and sagittal reconstructions for review. At our institution enteral (positive oral) contrast is not routinely used in abdominal-pelvic CT scans in cases of suspected

appendicitis, as our published data has shown that it does not improve diagnostic accuracy for appendicitis (26). Readers were blinded to the original report.

Clinical Outcomes

Patient medical records were reviewed to ascertain basic demographics and diagnoses. Patients were adjudicated to be true positives for appendicitis if they met the following criteria: 1) Appendectomy with confirmatory findings on pathology (n=94); 2) Drainage of a periappendiceal abscess (n=8); or 3) Initiation of antibiotic treatment for appendicitis (n=113), either as stand-alone therapy or as bridge to subsequent surgery (n=95) or drainage (n=8). Clinical experience suggests that most re-presentation of untreated appendicitis occurs within days to weeks from initial presentation (27–29). Accordingly, patient records were reviewed for up to one month after the date of their CT scan to ascertain potential “falsely negative” initial diagnosis.

Statistical analysis

We described cohort characteristics and presence of specific imaging findings stratified by original CT report. Categorical variables were summarized as count (percentage), and continuous variables were summarized as mean $\pm$ standard deviation (SD). Categorical and continuous variables were compared between strata using Fisher’s exact test and the Kruskal-Wallis test, respectively. Receiver operating characteristic (ROC) curve analysis and the area under the ROC curve (AUC), including 95% confidence intervals (CI), were used to compare certainty scores in patients with and without appendicitis. The positive predictive value (PPV) of appendicitis for each level of the certainty scale was calculated. Inter-reader agreement for continuous/ordinal variables was summarized using percent agreement, and Cohen’s κ was used for binary variables. Cohen’s Kappa is reported to account for agreement that may be due to chance, with values above 0.4 considered to have moderate agreement, above 0.6 to have substantial agreement, and above 0.8 to have nearly perfect agreement. The ROC analysis, the PPV appendicitis and inter-reader agreement analysis were weighted to account for the stratified sampling design (only 98 of 421 negative scans were reviewed). This gave each patient with a negative exam 421/98 (4.3) times the weight as the other patients, so that they would have the same weight as in the original cohort. The weighted estimates thus properly reflect their corresponding quantities in the cohort of interest. Throughout, two-tailed tests were used with $p < 0.05$ considered statistically significant. All statistical calculations were conducted with the statistical computing language R (version 3.1.1; R Foundation for Statistical Computing, Vienna, Austria).

Results

The overall incidence of appendicitis in this population was 20.1% (113 of 562), though could be considered 48% (113 of 237) in the cohort of selected exams. Patients with appendicitis (mean age 37.6 ± 15 , 67% male) were more frequently younger, male, and had higher WBC counts than those without appendicitis (all $p < 0.025$). The demographics, outcomes, and CT imaging findings of the 237 patients whose CT scans were reviewed are listed in Table 1, stratified by initial CT reports. Recognizing that the original report likely impacted patient management, nearly all patients who had an initial report of appendicitis

were given a clinical diagnosis of appendicitis (99%, n=96 of 97), admitted to the hospital (98%, n=95 of 97), and underwent appendectomy (89%, n=86 of 97). Pathology was available to confirm appendicitis in 81 of 86 appendectomy cases (94%) and demonstrated a normal appendix (negative appendectomy) in only 2 of 86 (2%) of those with initially positive CT reports. Similarly, patients who had negative CT reports were generally managed non-surgically, with only 27% admitted for other diagnoses. Patients with indeterminate reports had a variable course, with 60% diagnosed as not having appendicitis, and 14 patients (33%) undergoing appendectomy.

Retrospective imaging findings suggesting the presence or absence of appendicitis were recorded and are shown in Table 1. Some findings, such as large diameter (>10mm) of the appendix (outer-outer wall or tip) as well as fat stranding, were much more frequent in cases reported as positive for appendicitis. Other findings, including absence of gas in the lumen of the appendix and mucosal hyper-enhancement were found in over 50% of indeterminate cases.

Despite this wide distribution of objective retrospective findings, radiologists in this study consistently and accurately identified appendicitis (Table 2). Readers assigned a certainty score of 4 or 5 to 89% of cases with clinically-confirmed appendicitis, and a score of 1 or 2 to 84% of cases without appendicitis. Overall, the area under the curve (AUC) for predicting appendicitis using the certainty scale was 0.97 (CI: 0.95, 0.99; $p<0.001$). However, in the population with initially indeterminate scans, only 53% of cases of clinically-confirmed appendicitis received scores of 4 or 5 and 60% of true negative scans received scores of 1 or 2. Radiologists' ability to discriminate appendicitis in this group was lower than in the overall cohort with an AUC=0.85 (CI: 0.73, 0.93; $p<0.001$).

Readers' certainty in diagnosis of appendicitis was strongly related to the post-test probability of appendicitis (Table 3). Those patients in which radiologists assigned certainty scores of 1 or 2 (appendicitis definitely absent or unlikely, 70% of the population), had a positive predictive value (PPV) of 1% or lower (95% CI 0.2–2.0%). In patients given a score of 5 (appendicitis definitely present), however, the PPV estimate was 100% (95% CI: 96–100%). An indeterminate score of 3, even after review of all potential findings listed in the report, remained in 8.2% of cases.

In the subset of the CTs over-read by a second radiologist (n=159), inter-reader agreement was greater or equal to 80% in nearly all binary, objective findings, with over 90% agreement on the presence/absence of appendicolith, RLQ fluid collection, and peri-appendiceal air (Table 4). Continuous measures such as outer wall and tip diameters had slightly lower agreement (67% and 74%, respectively). Inter-reader agreement for the 5-point certainty of diagnosis scale was 56%. However, percent agreement was 80.5% if the certainty of appendicitis present (4 and 5) and certainty of appendicitis absent (1 and 2) categories were combined (Kappa 0.61, 95% CI 0.48–0.73).

The ROC curves corresponding to the primary readers and over-readers are shown in Figure 1. Both the primary readers (AUC=0.955) and over-readers (AUC=0.936) of the same exams had good diagnostic performance. In the subset of CT scans that had previously been

reported as indeterminate, the certainty scale is less discriminatory, with primary readers and over-readers having AUC values of 0.849 and 0.711, respectively.

Discussion

One in five patients in this “at-risk” population undergoing CT scans had appendicitis by our criteria. Systematic review of their CT images using the standardized reporting system demonstrated that some imaging findings were not uniformly reported while making the initial interpretation as positive, negative, or indeterminate. However, enumerating the findings in a standardized fashion resulted in high reproducibility at identifying individual imaging findings between radiologists. Furthermore, readers’ certainty in the “likely” or “definite” presence and absence of appendicitis was highly accurate (PPV for definite appendicitis was 100% and PPV for definitely absent was 0.4%). This high-level of certainty was reproducible across multiple readers.

This study builds on our prior, pilot work describing a standardized report for CT imaging of appendicitis, (25) and applies it across a larger population of “at risk” individuals, who are expected to have a higher proportion of negative scans and are more representative of the distribution of scans in clinical settings. Accordingly, our findings of high reproducibility are important in this setting, and may be used to propose benchmarks for the reproducibility in reported CT-imaging findings for appendicitis. Using this standardized report, readers demonstrated 80% agreement for most binary findings. Lower concordance for continuous measures was expected, where small measurement variability (e.g. appendiceal diameter) may be counted as disagreement, although some of the measurements are closely correlated. Despite high inter-reader agreement for essentially all findings, it is important to acknowledge that agreement was not perfect in certain settings that carry important predictive weight, like visualization of the appendix (which had nearly 20% disagreement). Some studies have reported appendicitis to be as low as 1% when the appendix is not visualized (30, 31).

While there is expected inter-reader variation, the extent to which any given CT is interpreted differently is not well described. The diagnostic certainty scale in this standardized report is designed to reflect this “subjective” interpretation of the constellation of findings by a given reader (25). Our findings are consistent with prior studies describing improved diagnostic accuracy after systematically weighing the presence or absence of specific imaging findings (32). Still, without having to account for *which* findings were most important, the communication of diagnosis using this certainty scale was striking. In the sample, nearly 90% of appendicitis cases were scored 4 or 5 compared with 5% of non-appendicitis cases. Similarly, certainty scores of 1 or 2 were present in 84% of cases without appendicitis. Furthermore, in this cohort of patients imaged for ‘possible appendicitis,’ a score of 1 had a PPV of 0.4%. Importantly, this ability of radiologists to subjectively discriminate across all cases was also reproducible in the subset of over-read CT scans.

However, this ability to correctly identify true positive cases of appendicitis was notably lower in the subset of CT scans previously reported as indeterminate, with nearly 50% of the scale scores listed as indeterminate or being discordant with ultimate clinical diagnosis.

Because appendicitis can be a difficult diagnosis to make – influenced by degree of inflammation, quality of scan, and patient characteristics - this observation may emphasize the scope of the problem with a category of “indeterminate” findings for appendicitis. Indeed, many studies don’t actually comment on a “CT indeterminate” group for appendicitis cases (11,18,21,24,33–35) when discussing diagnostic accuracy. The rate of CT-indeterminate reporting included in this cohort retrospectively was approximately 8% (42 of 561). Furthermore, even after applying the certainty scale across all CT scans, approximately 8% remained with a score of 3 or “indeterminate.” This is consistent with reported rates of indeterminate CT in the literature of 9–13% (9). Variation in “indeterminate” status is likely attributable to an unclear definition of what constitutes an “indeterminate” scan for appendicitis, subjective variation about which constellation of findings should be interpreted as “indeterminate,” and which CT findings, if any, can be used in these settings to further risk stratify patients into a more accurate diagnostic category. This implies that when appendicitis is not frankly obvious with the presence of multiple supportive findings, radiologists and clinicians cannot rely simply on their subjective gestalt to diagnose appendicitis. In this subset of patients, an objective, numerical prediction based on visualized findings may be a more powerful tool to complement clinical data and decision-making. Creation of such a prediction tool based on this standardized imaging report are in development by our group and others (36), but needs to be validated and linked to clinical management results.

Several limitations should be noted in this study. First, we identified our cohort through our imaging database using indications for CT scans. While our search strategy was less specific than imaging exclusively for “rule out appendicitis,” it was more sensitive than including all studies performed for “abdominal pain.” With an appendicitis prevalence of 20%, our cohort likely represents an important “at-risk” group of patients. Furthermore, our definition of clinical appendicitis was based on best evidence from the medical record. For instance, of the 113 patients who were clinically treated as appendicitis, 84% (n=95) had pathology reports available. Not all cases have pathologic confirmation, as appendicitis can be treated non-operatively with antibiotics and/or drainage. We performed a sensitivity analysis correcting our clinical diagnoses where pathology refuted the diagnosis (2 cases of negative appendectomy and 1 of positive appendectomy) and did not find differences in either measures of subjective scale or inter-reader agreement for those cases. In addition, we may have missed relevant recurrences of clinical appendicitis that occurred outside of our medical system. Second, the time and effort spent to fully evaluate each finding varied across the study’s 8 radiologists, despite standardization of the report and image review sessions. We had anticipated this, and limited each radiologist’s reading burden to 50 scans, balanced across previously positive, negative, and indeterminate reports. Still, each radiologist’s comfort level with making a diagnosis prior to meticulous review of every objective finding may differ, reflecting ‘real world’ variation, where communication with clinicians, additional imaging, or other external factors may all influence time spent evaluating the appendix. Finally, we did not report on the presence of other diagnoses that may have been identified by readers. The presence of alternate diagnoses likely has implications on certainty of appendicitis diagnosis and should be considered in future studies.

In conclusion, our study demonstrated high reproducibility of objective findings on CT for diagnosis of acute appendicitis across 8 readers reviewing imaging from an at-risk population. After enumerating these findings using the standardized report, the radiologists' certainty in diagnosing appendicitis was highly accurate and reproducible. Still, radiologists assigned 'indeterminate' diagnoses to approximately 8% of patients. As such, our group's next steps will focus on generating a better definition of which findings should be associated with this 'indeterminate' category and establishing which constellation of findings can predict the likelihood of appendicitis and success of different clinical management strategies.

References

1. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol.* 1990; 132:910–925. [PubMed: 2239906]
2. Weiss, AJ., Elixhauser, A., Andrews, RM. Characteristics of Operating Room Procedures in U.S. Hospitals, 2011: Statistical Brief #170. Anonymous Healthcare Cost and Utilization Project (HCUP) Statistical Briefs; Rockville (MD): 2006.
3. Neumayer L, Kennedy A. Imaging in appendicitis: a review with special emphasis on the treatment of women. *Obstet Gynecol.* 2003; 102:1404–1409. [PubMed: 14662234]
4. Drake FT, Flum DR. Improvement in the diagnosis of appendicitis. *Adv Surg.* 2013; 47:299–328. [PubMed: 24298858]
5. Tan WJ, Acharyya S, Goh YC, et al. Prospective comparison of the Alvarado score and CT scan in the evaluation of suspected appendicitis: a proposed algorithm to guide CT use. *J Am Coll Surg.* 2015; 220:218–224. [PubMed: 25488354]
6. Jones K, Pena AA, Dunn EL, Nadalo L, Mangram AJ. Are negative appendectomies still acceptable? *Am J Surg.* 2004; 188:748–754. [PubMed: 15619494]
7. Balthazar EJ, Birnbaum BA, Yee J, Megibow AJ, Roshkow J, Gray C. Acute appendicitis: CT and US correlation in 100 patients. *Radiology.* 1994; 190:31–35. [PubMed: 8259423]
8. Balthazar EJ, Megibow AJ, Siegel SE, Birnbaum BA. Appendicitis: prospective evaluation with high-resolution CT. *Radiology.* 1991; 180:21–24. [PubMed: 2052696]
9. Daly CP, Cohan RH, Francis IR, Caoili EM, Ellis JH, Nan B. Incidence of acute appendicitis in patients with equivocal CT findings. *AJR Am J Roentgenol.* 2005; 184:1813–1820. [PubMed: 15908536]
10. Peck J, Peck A, Peck C, Peck J. The clinical role of noncontrast helical computed tomography in the diagnosis of acute appendicitis. *Am J Surg.* 2000; 180:133–136. [PubMed: 11044529]
11. 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]
12. SCOAP Quarterly Report - Non-Elective Appendectomy 2015. [Accessed March 01, 2015] Surgical Care and Outcomes Assessment Program. at <https://cqi.armus.com/reports/SCOAP/index.html>
13. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. *Radiology.* 2006; 241:83–94. [PubMed: 16928974]
14. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg.* 2008; 32:1843–1849. [PubMed: 18553045]
15. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med.* 1986; 15:557–564. [PubMed: 3963537]
16. Kollar D, McCartan DP, Bourke M, Cross KS, Dowdall J. Predicting acute appendicitis? A comparison of the Alvarado score, the appendicitis inflammatory response score and clinical assessment. *World J Surg.* 2015; 39:104–109. [PubMed: 25245432]

17. Rosen MP, Ding A, Blake MA, et al. ACR Appropriateness Criteria(R) right lower quadrant pain--suspected appendicitis. *J Am Coll Radiol.* 2011; 8:749–755. [PubMed: 22051456]
18. 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]
19. 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]
20. 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]
21. Moteiki T, Horikoshi H. New CT criterion for acute appendicitis: maximum depth of intraluminal appendiceal fluid. *AJR Am J Roentgenol.* 2007; 188:1313–1319. [PubMed: 17449776]
22. 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]
23. 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]
24. Flum DR. Clinical practice. Acute appendicitis--appendectomy or the "antibiotics first" strategy. *N Engl J Med.* 2015; 372(20):1937–1943. 14. [PubMed: 25970051]
25. Godwin B, Simianu V, Drake F, et al. A Novel Appendicitis Reporting System to Improve Communication about Uncertainty in Appendicitis Imaging. *AJR Am J Roentgenol.* 2015; 204(6):1212–1219. [PubMed: 26001230]
26. Drake FT, Alfonso R, Bhargava P, et al. Enteral contrast in the computed tomography diagnosis of appendicitis: comparative effectiveness in a prospective surgical cohort. *Ann Surg.* 2014 Aug; 260(2):311–6. [PubMed: 24598250]
27. Temple CL, Huchcroft SA, Temple WJ. The natural history of appendicitis in adults. A prospective study. *Ann Surg.* 1995; 221:278–281. [PubMed: 7717781]
28. Andersson RE. The natural history and traditional management of appendicitis revisited: spontaneous resolution and predominance of prehospital perforations imply that a correct diagnosis is more important than an early diagnosis. *World J Surg.* 2007; 31:86–92. [PubMed: 17180556]
29. Di Saverio S, Sibilio A, Giorgini E, et al. The NOTA Study (Non Operative Treatment for Acute Appendicitis): prospective study on the efficacy and safety of antibiotics (amoxicillin and clavulanic acid) for treating patients with right lower quadrant abdominal pain and long-term follow-up of conservatively treated suspected appendicitis. *Ann Surg.* 2014; 260:109–117. [PubMed: 24646528]
30. Nikolaidis P, Hwang CM, Miller FH, Papanicolaou N. The nonvisualized appendix: incidence of acute appendicitis when secondary inflammatory changes are absent. *AJR Am J Roentgenol.* 2004; 183:889–892. [PubMed: 15385277]
31. Ganguli S, Raptopoulos V, Komlos F, Siewert B, Kruskal JB. Right lower quadrant pain: value of the nonvisualized appendix in patients at multidetector CT. *Radiology.* 2006; 241:175–180. [PubMed: 16928971]
32. 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]
33. 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]
34. 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 i.v. contrast-enhanced CT scans. *AJR Am J Roentgenol.* 2009; 193:96–105. [PubMed: 19542400]
35. 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]

36. Thompson AC, Olcott EW, Poullos PD, et al. Predictors of appendicitis on computed tomography among cases with borderline appendix size. *Emerg Radiol*. 2015 In Press, Corrected Proof.

Appendix 1

Standardized imaging report for appendicitis. [25]

1. Appendix visualized	☐ Yes	☐ No
2. Outer-outer wall diameter:	_____mm	
3. Tip Diameter	_____mm	
4. Single wall thickness	_____mm	
5. Mucosal hyper-enhancement	☐ Yes	☐ No
6. Surrounding fat stranding / thickening of pararenal / 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 / phlegmon / collection / abscess	☐ Yes	☐ No
11. Gas in lumen of appendix.	☐ Yes	☐ No
12. Right Ovary (in female patients)	☐ Normal	☐ Abnormal:
13. Certainty scale:	☐ 1: Appendicitis definitely absent ☐ 2: Appendicitis unlikely ☐ 3: Indeterminate ☐ 4: Appendicitis likely ☐ 5: Appendicitis definitely present	

Advances in Knowledge

Despite a wide distribution of objective findings, radiologists consistently and accurately identified appendicitis.

Readers' certainty in diagnosis of appendicitis is strongly related to the post-test probability of appendicitis.

Inter-reader agreement was greater or equal to 80% in nearly all objective findings on the standardized report, with over 90% agreement on the presence/absence of appendicolith, right lower quadrant fluid collection, and peri-appendiceal air.

Implications for Patient Care

After enumerating these findings using the appendicitis standardized report, the radiologists' certainty in diagnosing appendicitis was highly accurate and reproducible. Still, approximately 8% of patients have indeterminate CT reports. This suggests that future work should be directed at evaluating which constellation of findings can be used to better predict the likelihood of appendicitis and link these findings to success of clinical therapy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Summary statement

Using a standardized reporting system resulted in high reproducibility of objective CT findings for appendicitis and achieved high diagnostic accuracy in an at-risk population.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. ROC curves for appendicitis certainty scale from primary reads (AUC=0.955) and over-reads (AUC=0.936) based on 159 exams which were over-read.

Table 1
Cohort demographics and clinical outcomes, stratified by original CT report. (N=237 exams).

Clinical Variable	Original CT Report*			P-value
	Appendicitis N=97	Indeterminate N=42	Not Appendicitis N=98	
Sex	Female	24 (57)	53 (54)	0.003
	Male	18 (43)	45 (46)	
Age, years	37.6 ± 14.8	33.8 ± 16.2	41.1 ± 16.9	0.025
WBC count	12.3 ± 4.2	10.4 ± 4.0	9.4 ± 4.4	<0.001
Admitted	95 (98)	25 (60)	26 (27)	<0.001
Appendectomy	86 (89)	14 (33)	0 (0)	<0.001
Drain placed	7 (7)	1 (2)	1 (1)	0.053
Antibiotics given	93 (96)	17 (40)	0 (0)	<0.001
Clinical diagnosis	Appendicitis	17 (40)	0 (0)	<0.001
	Not appendicitis	1 (1)	98 (100)	
Imaging Variable [§]				
1. Appendix visualized	92 (95)	37 (88)	70 (71)	<0.001
2. Outer-outer wall diameter [‡]	> 10 mm	8 (22)	1 (1)	<0.001
	6–10 mm	21 (23)	23 (33)	
	6 mm	1 (1)	46 (66)	
3. Tip diameter [‡]	> 10 mm	30 (33)	2 (6)	<0.001
	6–10 mm	50 (54)	20 (56)	
	6 mm	12 (13)	14 (39)	<0.001
4. Single wall thickness > 3 mm [‡]	57 (62)	8 (23)	6 (9)	<0.001

Clinical Variable	Original CT Report*			P-value
	Appendicitis N=97	Indeterminate N=42	Not Appendicitis N=98	
5. Mucosal hyper-enhancement [†]	69 (78)	17 (52)	7 (11)	<0.001
6. Fat stranding	92 (96)	19 (45)	19 (20)	<0.001
7. Appendicolith [‡]	30 (33)	9 (24)	3 (4)	<0.001
8. Cecal thickening	60 (62)	7 (17)	7 (7)	<0.001
9. Periappendiceal air [‡]	8 (9)	0 (0)	0 (0)	0.008
10. RLQ fluid/phlegmon	11 (11)	2 (5)	5 (5)	0.25
11. Absence of gas in lumen of appendix [‡]	62 (67)	21 (57)	21 (30)	<0.001
12. Absence of right ovary abnormality	94 (98)	37 (88)	85 (89)	0.020
13. Certainty scale	Appendicitis definitely absent (1)	7 (17)	56 (57)	<0.001
	Appendicitis unlikely (2)	8 (19)	33 (34)	
	Indeterminate (3)	15 (36)	7 (7)	
	Appendicitis likely (4)	9 (21)	2 (2)	
	Appendicitis definitely present (5)	3 (7)	0 (0)	

* Values are no. (%) or mean $\pm$ SD;

[§] Enumerated as per Standardized imaging report

[†] Only reported by radiologists when appendix was visualized

Table 2

Diagnostic performance of certainty scale.

Certainty Scale	All Exams (N=237)				Indeterminate Exams (N=42)			
	Clinical Appendicitis (N=113)	Clinical Not Appendicitis (N=124)	AUC (95% CI)	P-value	Clinical Appendicitis (N=17)	Clinical Not Appendicitis (N=25)	AUC (95% CI)	P-value
1: Appendicitis definitely absent	1 (1)	63 (51)	0.968	<0.001	0 (0)	7 (28)	0.846	<0.001
2: Appendicitis unlikely	3 (3)	41 (33)	(0.95–0.99)		0 (0)	8 (32)	(0.73–0.93)	
3: Indeterminate	9 (8)	14 (11)			8 (47)	7 (28)		
4: Appendicitis likely	20 (18)	6 (5)			6 (35)	3 (12)		
5: Appendicitis definitely present	80 (71)	0 (0)			3 (18)	0 (0)		

CI = confidence interval; PPV = positive predictive value;

* Values are no. (%) and are not weighted for the sampling design.

Table 3

Estimated probability of appendicitis for each level of certainty scale using all exams (N=237).

Certainly Scale	No. in Sample	Estimated % in Population	PPV for Appendicitis	(95% CI)
1: Appendicitis definitely absent	64	44.4%	0.4%	(0.0–1.3%)
2: Appendicitis unlikely	44	27.3%	2.0%	(0.0–4.7%)
3: Indeterminate	23	8.2%	19.5%	(8.6–37.7%)
4: Appendicitis likely	26	5.8%	61.4%	(39.3–89.7%)
5: Appendicitis definitely present	80	14.3%	100.0%	(95.5–100%)*

CI = confidence interval; PPV = positive predictive value;

* CI is unweighted for the sampling design; a weighted CI could not be computed when PPV = 100%;

Table 4

Inter-reader agreement of imaging-based findings based on over-read exams (N=159).

Variable [§]	Reader [†]		Inter-reader Agreement		
	No. *	Primary Reader	Over-Reader	Cohen's Kappa	% Agree (95% CI)
13. Certainty scale	159	40 (25)	40 (25)	0.39	(0.27–0.52)
Appendicitis definitely absent (1)		32 (20)	28 (18)		55.7
Appendicitis unlikely (2)		21 (13)	19 (12)		
Indeterminate (3)		20 (13)	16 (10)		
Appendicitis likely (4)		46 (29)	56 (35)		
Appendicitis definitely present (5)					
Individual Imaging Findings					
1. Appendix visualized	159	133 (84)	140 (88)	0.42	(0.19–0.65)
2. Outer-outer wall diameter	126	48 (38)	51 (40)	0.50	(0.34–0.65)
> 10 mm		47 (37)	49 (39)		67.4
6–10 mm		31 (25)	26 (21)		
6 mm					
3. Tip diameter	125	24 (19)	24 (19)	0.51	(0.35–0.67)
> 10 mm		52 (42)	50 (40)		73.5
6–10 mm		49 (39)	51 (41)		
6 mm					
4. Single wall thickness > 3 mm	123	45 (37)	42 (34)	0.38	(0.20–0.56)
5. Mucosal hyper-enhancement	114	56 (49)	61 (54)	0.67	(0.53–0.82)
6. Fat stranding	157	84 (54)	87 (55)	0.65	(0.50–0.79)
7. Appendicolith	126	35 (28)	34 (27)	0.85	(0.72–0.99)
8. Cecal thickening	158	46 (29)	50 (32)	0.52	(0.35–0.70)
9. Periappendiceal air	126	4 (3)	6 (5)	0.59	(0.23–0.96)
10. RLQ fluid/phlegmon	159	13 (8)	12 (8)	0.49	(0.18–0.80)

Variable [§]	Reader [‡]		Inter-reader Agreement		
	No. *	Primary Reader	Over-Reader	Cohen's Kappa	% Agree (95% CI)
11. Absence of gas in lumen of appendix	126	68 (54)	72 (57)	0.72	(0.56–0.88) 86.4
12. Absence of right ovary abnormality	155	143 (92)	138 (89)	0.46	(0.18–0.73) 88.4

CI = confidence interval;

[§]Enumerated as per Standardized imaging report

^{*}The number of exams differs from total (n=159) to represent total where finding could be assessed by both readers (for instance, outer wall diameter reported only on subset of CT scans when both radiologists agreed appendix could be visualized);

[‡]Values are no. (%) and are not weighted for the sampling design.