Pediatr Radiol
DOI 10.1007/s00247-015-3443-4

ORIGINAL ARTICLE

Development and validation of an ultrasound scoring system for children with suspected acute appendicitis

Sara C. Fallon[1] · Robert C. Orth[2] · R. Paul Guillerman[2] · Martha M. Munden[2] · Wei Zhang[3] · Simone C. Elder[1] · Andrea T. Cruz[4] · Mary L. Brandt[1] · Monica E. Lopez[1] · George S. Bisset[2]

Received: 24 March 2015 / Revised: 28 May 2015 / Accepted: 15 July 2015

Abstract
Background To facilitate consistent, reliable communication among providers, we developed a scoring system (Appy-Score) for reporting limited right lower quadrant ultrasound (US) exams performed for suspected pediatric appendicitis.
Objective The purpose of this study was to evaluate implementation of this scoring system and its ability to risk-stratify children with suspected appendicitis.
Materials and methods In this HIPAA compliant, Institutional Review Board-approved study, the Appy-Score was applied retrospectively to all limited abdominal US exams ordered for suspected pediatric appendicitis through our emergency department during a 5-month pre-implementation period (Jan 1, 2013, to May 31, 2013), and Appy-Score use was tracked prospectively post-implementation (July 1, 2013, to Sept. 30, 2013). Appy-Score strata were: 1=normal completely visualized appendix; 2=normal partially visualized appendix; 3=non-visualized appendix, 4=equivocal, 5a=non-perforated appendicitis and 5b=perforated appendicitis. Appy-Score use, frequency of appendicitis by Appy-Score stratum, and diagnostic performance measures of US exams were computed using operative and clinical finding as reference standards. Secondary outcome measures included rates of CT imaging following US exams and negative appendectomy rates.
Results We identified 1,235 patients in the pre-implementation and 686 patients in the post-implementation groups. Appy-Score use increased from 24% (37/155) in July to 89% (226/254) in September ($P<0.001$). Appendicitis frequency by Appy-Score stratum post-implementation was: 1=0.5%, 2=0%, 3=9.5%, 4=44%, 5a=92.3%, and 5b=100%. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were 96.3% (287/298), 93.9% (880/937), 83.4% (287/344), and 98.8% (880/891) pre-implementation and 93.0% (200/215), 92.6% (436/471), 85.1% (200/235), and 96.7% (436/451) post-implementation – only NPV was statistically different ($P=0.012$). CT imaging after US decreased by 31% between pre- and post-implementation, 8.6% (106/1235) vs. 6.0% (41/686); $P=0.048$). Negative appendectomy rates did not change (4.4% vs. 4.1%, $P=0.8$).
Conclusion A scoring system and structured template for reporting US exam results for suspected pediatric appendicitis was successfully adopted by a pediatric radiology department at a large tertiary children's hospital and stratifies risk for children based on their likelihood of appendicitis.

Keywords Appendicitis · Appendix · Children · Structured reporting · Ultrasound

✉ Robert C. Orth
rcorth@texaschildrens.org

[1] Division of Pediatric Surgery, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA

[2] Edward B. Singleton Department of Pediatric Radiology, Texas Children's Hospital, Baylor College of Medicine, 6701 Fannin St., MC CC470.01, Houston, TX 77030, USA

[3] Surgical Outcomes Center, Texas Children's Hospital, Houston, TX, USA

[4] Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA

Introduction

Abdominal pain is one of the most common presenting complaints in the pediatric emergency department [1]. The diagnosis and management of children with abdominal pain in the emergency department can be complex, as etiologies of the

Published online: 18 August 2015

pain can range from mild constipation to bowel perforation [1]. At our children's hospital, the burden of acute appendicitis is large, with more than 1,000 cases annually, resulting in a high volume of imaging requests.

The initial imaging modality used to evaluate children with suspected appendicitis at our institution is US. In a previous study, we found that 98.7% of patients who presented to our hospital with acute appendicitis had a diagnostic US prior to surgery [2]. Only 8.2% underwent evaluation with CT and the majority of CT exams were obtained after an equivocal US.

During a multidisciplinary discussion of our institution's use of US to evaluate children with suspected appendicitis, providers noted that US exam reports often contained inconsistent or noncommittal language that created uncertainty about the radiologist's impression of the patient's likelihood of acute appendicitis. We were concerned that this lack of clarity might lead to increased rates of follow-up CT exams resulting in increased exposure to ionizing radiation and increased time to definitive treatment.

In response to this perceived opportunity for process improvement in our care of children with abdominal pain, we developed and implemented a scoring system (Appy-Score) and a standardized structured reporting template for limited right lower quadrant abdominal US exams obtained for suspected pediatric appendicitis. The purpose of this study was to evaluate our ability to implement this scoring and reporting system, determine the Appy-Score's ability to stratify risk for patients with suspected pediatric appendicitis, and measure the effect of this scoring system on US diagnostic performance measures, follow-up CT exam rates and negative appendectomy rates.

Materials and methods

Development of the Appy-Score and structured reporting template

The Appy-Score was developed with multidisciplinary input from the Pediatric Radiology, Pediatric Surgery and Pediatric Emergency Medicine departments at a large, tertiary care children's hospital. The purpose of the Appy-Score was to provide a means of clearly communicating the radiologist's impression of the patient's likelihood of acute appendicitis. The Appy-Score is stratified into six categories as shown in Table 1. Appy-Scores 1, 2 and 3 indicate a completely visualized (including the tip), partially visualized or non-visualized appendix, respectively, with no findings of appendiceal or periappendiceal inflammation. An Appy-Score of 4 indicates that the study does not meeting criteria for Appy-Scores 1, 2, 3, 5a or 5b. Examples of equivocal studies include those with periappendiceal inflammatory changes (such as increased echogenicity of periappendiceal fat) or borderline appendiceal enlargement (6-7 mm) with an otherwise normal appendix.

Table 1 Appy-Score strata

Appy-Score	
1	Completely visualized normal-appearing appendix with no ancillary findings to suggest appendicitis
2	Partially visualized normal-appearing appendix with no ancillary findings to suggest appendicitis
3	Non-visualized appendix with no ancillary findings to suggest appendicitis
4	Equivocal
5a	Non-perforated acute appendicitis
5b	Perforated appendicitis

In order for the Appy-Score to be reported consistently and clearly, structured reporting templates were created. The templates were built with input from all radiologists within the body division of the Department of Pediatric Radiology (19 radiologists), and four versions of the templates were tested for clarity, ease of reporting and inclusiveness of all pertinent findings. Six separate structured reporting templates were created, one for each Appy-Score stratum. Final versions of the reporting templates provided the patient's Appy-Score as well as a summary of the findings that led the radiologist to arrive at the score. The reporting template for a 5a Appy-Score and a US image showing acute non-perforated appendicitis are shown in Fig. 1. The Appy-Score system and reporting templates were presented at a faculty meeting and via e-mail, and feedback and suggestions were elicited. The point of greatest debate was whether to include physical examination findings in the structured reporting template. Physical examination findings were not included in the final version of the structured reporting template because an attending radiologist is not available to scan each patient due to overnight interpretations by trainees, some examinations are performed by technologists at remote facilities, and most attending radiologists personally scan only a small minority of patients. The template does allow for additional comments, which could include physical examination findings. After consideration of feedback and suggestions, the scoring systems and reporting templates were finalized, and department leadership mandated implementation.

Ultrasound technique

Our institutional US technique has been described in a prior publication [3]. Briefly, all patients underwent gray-scale and color Doppler US imaging of the right lower quadrant with graded compression. Screening images of the bladder, liver, gallbladder and right kidney are routinely obtained as part of this examination. Exams were performed on LOGIQ E9 imaging systems using 5-MHz curved, 9-MHz linear or 15-MHz linear transducers (General Electric Corp., Waukesha, WI).

a

EXAM: Limited abdominal ultrasound

CLINICAL HISTORY: [Abdominal pain – concern for appendicitis]

PRIOR STUDIES: [None]

FINDINGS:

Appendix:
-Visualized: [Completely]
-Fluid-filled: [Yes]
-Compressible: [No]
-Maximum diameter with compression (outer wall to outer wall): [11 mm]
-Appendicolith: [No]
-Wall
  --Hyperemia: [No]
  --Thickening (>2 mm): [Yes]
  --Loss of mural stratification: [Yes]

Free fluid: [No]

Increased echogenicity of periappendiceal fat: [Yes]

Abscess: [No]

Additional findings: [None]

IMPRESSION:

Appendicitis score: 5a

Alternative/additional diagnosis: [None]

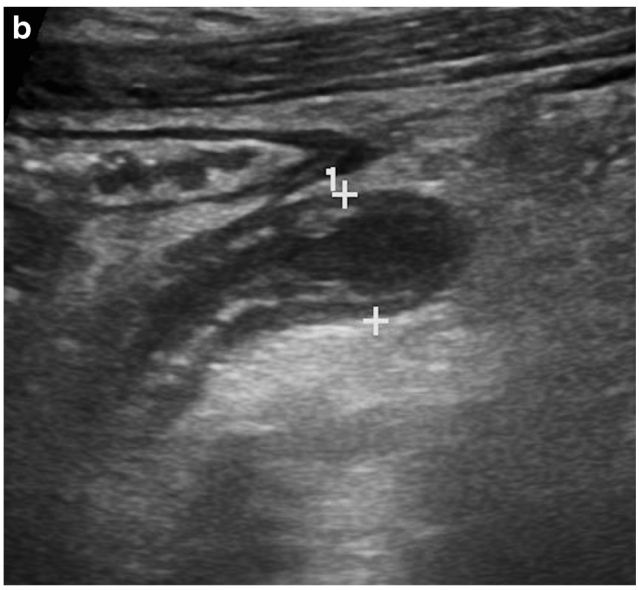

Fig. 1 Reporting template and a US image show acute non-perforated appendicitis. **a** Structured reporting template for Appy-Score 5a. Values for the fields in red can be changed at the discretion of the interpreting radiologist. **b** Gray-scale US image of surgically proven non-perforated acute appendicitis in a 7-year-old girl (Appy-Score 5a) with a diameter of 1.1 cm, wall thickening, loss of mural stratification and increased echogenicitiy of periappendiceal fat

All members of the pediatric radiology staff (n=19, post-residency experience range: 3-40 years) interpreted US exams during the course of the study. US exams were considered positive for acute non-perforated appendicitis (Appy-Score 5a) if the compressed appendix measured >6 mm in maximum outer diameter, and there were associated inflammatory changes such as increased echogenicity of periappendiceal fat, appendiceal wall hyperemia or thickening, and/or periappendiceal fluid but without specific findings of perforation. US exams were considered positive for perforated appendicitis (Appy-Score 5b) if there were marked inflammatory changes in the right lower quadrant with or without visualization of the appendix, an appendicolith without visualization of the appendix, echogenic free fluid, or a fluid collection indicating peritonitis or abscess.

Study design

Institutional Review Board approval (H-33158), which waived the need for informed consent, was obtained for this HIPAA-compliant retrospective and prospective process improvement study. The intervention was the use of the Appy-Score and structured reporting templates. Children who underwent limited abdominal US exams for suspected appendicitis prior to implementation of the Appy-Score (Jan. 1, 2013, to May 31, 2013) were compared to those imaged after implementation of the Appy-Score (July 1, 2013, to Sept. 1, 2013). Data from June 2013 were excluded as we used this month to test and adjust different versions of the reporting template. All limited abdominal US exams performed during the pre- and post-implementation periods were identified through a query of our hospital electronic medical record. All patients <19 years of age who underwent a limited right lower quadrant US exam for suspected appendicitis were included. Patients who underwent an US exam to evaluate for abdominal abscess following laparoscopic appendectomy, cholecystitis, intussusception or pyloric stenosis were excluded. Patients who had imaging performed at another institution and received a subsequent US at our hospital were also excluded. For the post-implementation group, studies that were not reported using both the Appy-Score and structured reporting template were excluded.

Primary outcome measures were the extent of Appy-Score and structured reporting template adoption, appendicitis frequency by Appy-Score stratum pre- and post-implementation, and US diagnostic performance measures (sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV]) pre- and post-implementation. Secondary outcome measures were rates of post-US CT scanning and negative appendectomy rates. Findings leading to equivocal exams (Appy-Score 4) were also recorded for the post-implementation group.

Extent of adoption was calculated for the first month of the post-implementation period (July 2013) and the last month of

the post-implementation period (September 2013) by dividing the number of limited abdominal US exams performed for suspected appendicitis in which the Appy-Score and structured reporting templates were used by the total number of exams performed for suspected appendicitis during each respective month.

For pre-intervention patients, an Appy-Score was assigned by a pediatric surgery research fellow (S.F.) based on the US exam report. In cases of ambiguous language or equivocal study results, a pediatric surgery research fellow (S.F.) and a pediatric radiologist (R.C.O. with 6 years' post-fellowship experience) blinded to CT, clinical and surgical outcomes reviewed the US exam reports and assigned the Appy-Score by consensus. To assess for selection bias, Appy-Scores were similarly assigned to post-implementation exams in which the scoring system was not used. The frequency of appendicitis by Appy-Score strata and US diagnostic performance measures for the pre- and post-implementation groups were calculated by comparison to the reference standards of surgical diagnosis for those patients who underwent appendectomies and clinical observation for patients managed nonoperatively who were determined not to have appendicitis. Designation of perforated or non-perforated appendicitis was based on the operative report. In order to accurately capture the number of true negative exams, the caretakers of patients who did not undergo an appendectomy were contacted 30 days after patient presentation to our emergency department and asked if they had subsequently been treated for appendicitis at another institution.

Chi-square testing was used to compare the rates of Appy-Score use between the first and last months of the post-implementation period, appendicitis incidence between post-implementation studies in which the scoring system was used and those in which it was not used, the percentage of patients per Appy-Score stratum between the pre- and post-implementation groups, and the percentage of patients with surgically proven appendicitis by Appy-Score stratum between the pre- and post-implementation groups. The percentage of patients in each Appy-Score stratum pre- and post-implementation was calculated. Fisher exact test was used to compare diagnostic performance measures and chi-square testing was used to compare post-US CT scanning rates and negative appendectomy rates between the pre- and post-implementation groups. A completely visualized normal-appearing appendix (Appy-Score 1), a partially visualized normal-appearing appendix (Appy-Score 2), or a non-visualized appendix without ancillary signs to suggest appendicitis (Appy-Score 3) were considered negative for purposes of calculating test performance measures and equivocal US exams (Appy-Score 4) were considered positive. A receiver operator curve (ROC) was generated to evaluate the diagnostic performance of equivocal US exams (Appy-Score 4) relative to appendiceal size for those cases in which borderline or mild appendiceal enlargement led to the equivocal designation.

Results

We identified 1,235 US exams performed during the pre-implementation period (out of 2,816 limited abdominal US exams performed for all indications [mean age: 9.7 +/- 4.6 years; M:F=1:0.94]) and 686 US exams performed during the post-implementation period that met study criteria (out of 1,791 limited abdominal exams performed for all indications; mean age: 9.7 +/- 4.6 years; M:F=1:1) that met study criteria. Appy-Score use increased from 24% (37/155) in July to 89% (226/254) in September ($P<0.001$). Appendicitis incidence in the pre-implementation group was 23.1%. Of all US exams performed for suspected appendicitis in the post-implementation period, 299 of 985 were reported without using the structured reporting template or scoring system. Appendicitis incidence in these 299 patients was 33.4% (100/299) vs. 41.7% (286/686) in the post-implementation study population, $P=0.015$. Test performance measures for US exams reported without using the structured reporting template or scoring system were sensitivity=94.0% (94/100), specificity=96.5% (192/199), PPV=93.1% (94/101) and NPV=97.0% (192/198).

The reports for 6.6% (81/1,235) of pre-implementation US exams contained ambiguous language or were designated equivocal and required consensus review. The percentage of patients in each Appy-Score stratum and the frequency of surgically proven appendicitis for each Appy-Score stratum pre- and post-implementation are shown in Figs. 2 and 3. The percentage of patients designated Appy-Score 5a increased significantly post-implementation ($P=0.02$). The percentage of patients with surgically proven appendicitis significantly decreased for Appy-Score strata 2 ($P=0.029$) and 5a ($P=0.03$) and increased for strata 3 ($P=0.02$) and 4 ($P=0.04$). We were able to obtain follow-up information after their emergency room visit in 231/470 (49%) of post-implementation patients who did not undergo appendectomy; there was one case of missed appendicitis (0.4%, 1/231).

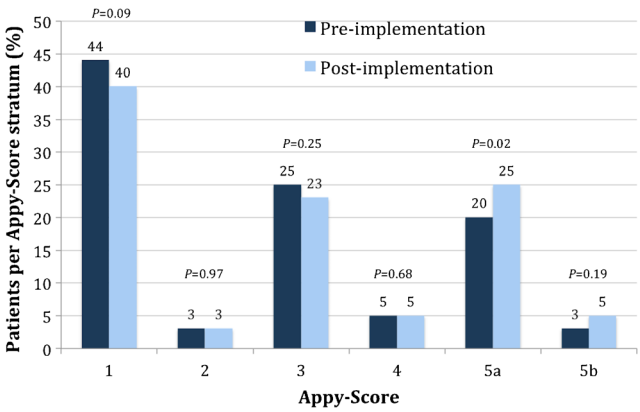

Fig. 2 Percentage of patients in the pre- and post-implementation groups within each Appy-Score stratum. The numerical percentage value of each bar is given above the bar and the *P*-value for comparisons between percentages pre- and post-implementation are given for each Appy-Score stratum

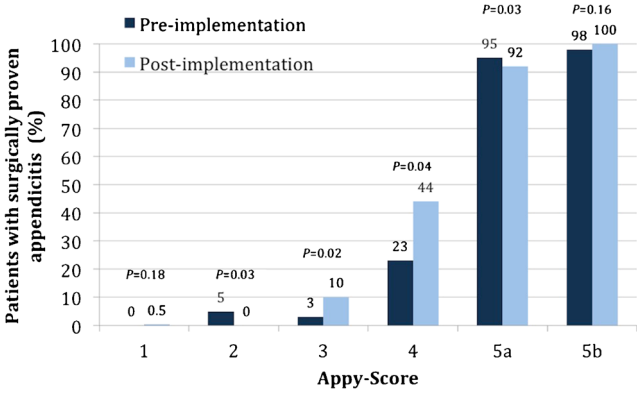

Fig. 3 Frequency of surgically proven appendicitis for each Appy-Score stratum pre- and post-implementation. The numerical percentage value of each bar is given above the bar and the *P*-value for comparisons between percentages pre- and post-implementation are given for each Appy-Score stratum

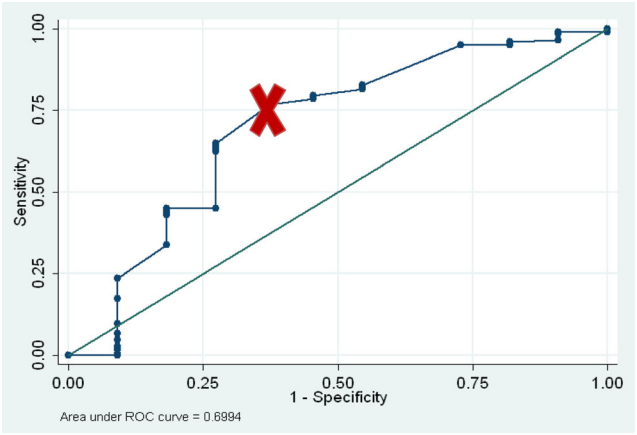

Fig. 4 Receiver operator curve evaluating appendiceal diameter thresholds for acute appendicitis

The use of the Appy-Score system did not significantly change the diagnostic performance of US exams between the pre- and post-implementation groups with the exception of the NPV, which decreased from 98.8% (95% CI: 97.7%-99.3%) pre-implementation to 96.7% (95% CI: 94.5%-98.1%) post-implementation, P=0.012 (Table 2). The rate of follow-up CT scanning after US decreased from 8.6% (106/1,235) to 6.0% (41/686), P=0.048, and negative appendectomy rates were not statistically different – 4.4% (54/1,235) vs. 4.1% (28/686), P=0.8.

Further investigation of our post-implementation patients with Appy-Score 4, which represents 5% of patients, revealed that the majority of patients (27/33) had a score of 4 on the basis of borderline enlargement of the appendix without secondary signs to suggest appendicitis. Inflammatory changes in the periappendiceal fat, tenderness on exam, non-compressibility of the appendix and the presence of abnormal periappendiceal fluid were secondary signs of appendicitis that led to an equivocal designation in the remaining 6/33. Of the patients who were scored a 4 on the basis of a borderline enlarged appendix in the post-implementation group, 10/27 (37%) had appendicitis, compared to 4/6 (66%) for patients with secondary signs of inflammation. We evaluated the diagnostic performance of the size of the appendix using a receiver operator curve (Fig. 4) and found that a size threshold of 7.5 mm provided the best balance between sensitivity and specificity, which is slightly higher than our initial size threshold of 6.0 mm [4, 5].

Discussion

Due to concerns over the adverse effects of ionizing radiation and the relatively high cost of CT scanning, tertiary care children's hospitals are developing abdominal pain protocols that encourage the use of US as the preferred initial imaging modality for suspected appendicitis [1, 6, 7]. A similar emphasis has not been necessarily seen in community hospitals, and the type of hospital has been found to be an independent predictor of preferential CT use [8]. However, significant variation in preoperative appendicitis imaging exists even among specialized children's hospitals [9].

While the diagnostic performance characteristics of CT and US can be similar, the success of US is more often dependent on the skills of the technologist, with additional costs related to the manpower hours necessary to perform the studies [8]. Other data suggest US use may be more cost-effective [10]. The sensitivity of US ranges from 42% to 90% in recently reported studies [4, 11]. The sensitivity of CT is more often reported as high, with a range of 85-95% [12–14]. Specificity is greater than 90% for both CT and US, and negative

Table 2 Test performance measures of US pre- and post-implementation

	Pre-implementation	Post-implementation	*P*-value
Sensitivity	96.3% (287/298), 95% CI=93.3%-98.0%	93.0% (200/215), 95% CI=88.5%-95.9%)	0.105
Specificity	93.9% (880/937), 95% CI=92.1%-95.3%	92.6% (436/471), 95% CI=89.7%-94.7%)	0.361
PPV	83.4% (287/344), 95% CI=79.0%-87.1%	85.1% (200/235), 95% CI=79.8%-89.3%)	0.644
NPV	98.8% (880/891), 95% CI=97.7%-99.3%	96.7% (436/451), 95% CI=94.5%-98.1%)	0.012

95% confidence intervals in parentheses

PPV positive predictive value, *NPV* negative predictive value

predictive values are high [14–16]. A barrier to more universal adoption of US as a first-line imaging exam for suspected pediatric appendicitis is the perception of increased variability related to provider interpretation. Our study provides a standardized method for streamlined US interpretation and reporting that may encourage other institutions to implement non-CT-based diagnostic algorithms.

As US is already established as the preferred initial imaging modality at our institution, the process improvement initiative in this study was the development of a risk stratification scoring system and structured reporting template for children with suspected appendicitis. Implementation of the Appy-Score and structured reporting template was highly successful at our institution, with its clinical adoption approaching 90% only 3 months after its introduction. Factors influencing this rapid adoption included full departmental and multidisciplinary commitment, as well as an emphasis on integrating the use of the reporting template into the routine workflow of the radiologists. The success of this process improvement initiative has led our institution to investigate other disease processes that would lend themselves to a similar standardization of radiology reporting templates to improve clarity and effective communication; projects regarding neonatal abdominal radiography for suspected necrotizing enterocolitis and ultrasonography to evaluate for pyloric stenosis are in development.

The Appy-Score was also successful in providing accurate risk stratification of those patients with suspected appendicitis (for example, those with equivocal US exams and a corresponding Appy-Score of 4 had a nearly 50% chance of having appendicitis). Implementation of the Appy-Score did not change the test performance characteristics of US with the exception of a statistically significant but numerically small decrease in NPV. The CT scan rate decreased during the post-implementation study period without increasing the negative appendectomy rate. Future study regarding the effect of this streamlined communication on other processes in the emergency department, such as the time to surgical consultation, admission or operation, is needed to understand the broader effects of this scoring and reporting system.

The number of equivocal US exams with surgically proven appendicitis nearly doubled between the pre- and post-implementation groups, 23% to 44%. One possibility for this is that some of the US exams designated equivocal in the pre-implementation group were misclassified. For example, several pre-implementation examinations were reported as "Probable appendicitis" and retrospectively designated Appy-Score 5. However, "probable" is an ambiguous term that some radiologists might consider Appy-Score 5 and others might consider Appy-Score 4. Similarly, misclassification of some pre-implementation exams may at least partially explain the increase in surgically proven cases of appendicitis with Appy-Score 3 exams and the decrease in with Appy-Scores 2 and 5a exams.

A similar 5-category scheme for reporting US examinations for suspected appendicitis was recently reported by Larson et al. [17], who designated two equivocal categories – one in which the appendix was visualized and another in which the appendix was not visualized. Although the percentage of equivocal US exams was higher than in our study and the rate of appendicitis in equivocal cases was lower than in our study, both studies show the importance of including an equivocal category for accurate risk stratification.

Other studies have investigated the issue of equivocal US exams for suspected appendicitis, although the definition of "equivocal" varies. One study examining outcomes in those with an "incompletely visualized" appendix found that 15% required an operation, with only a 0.3% missed appendicitis rate [11]. In a study of US exams in which the appendix was not identified, the presence of secondary signs of inflammation increased the odds ratio of having appendicitis, and an increased number of secondary signs further increased these odds [18]. Of additional concern is the optimal cutoff criterion for an enlarged appendix. Historically, enlargement has been defined as a maximal outer diameter exceeding 6 mm with compression. However, for equivocal studies, we found that a 7.5-mm threshold may be more optimal for diagnosing appendicitis, in contrast to a recent publication suggesting that 6 mm may be the appropriate cutoff for normal appendiceal diameter [19]. The 6-mm threshold has been challenged in three recently reported studies evaluating the optimal size threshold for the diagnosis of acute appendicitis [3, 20, 21]. Our data suggest that secondary signs of inflammation may be more predictive of appendicitis than isolated enlargement when an US exam is found to be equivocal by the radiologist, but further investigation of a greater number of exams (only 33 were available for review in this study) is required to make definitive conclusions and change institutional practice definitions of appendiceal enlargement.

An issue this study did not address was the ability of US to differentiate simple from perforated appendicitis. This differentiation has clinical implications, as new data suggest that antibiotic therapy alone may be adequate treatment for non-perforated appendicitis in children [22]. As a randomized trial found that nonoperative management of perforated appendicitis with an associated phlegmon in children results in worse outcomes than primary operative management, the ability to distinguish simple from perforated appendicitis is critical if practice patterns eventually trend away from operative management of simple appendicitis [23].

There is potential selection bias given that not all US exams performed for suspected appendicitis in the post-implementation period were reported using the Appy-Score. The incidence of appendicitis differed between post-implementation patients reported without and with the Appy-Score system, 33.4% (100/299) vs. 41.6% (286/686), respectively. Excluding normal patients from a cohort would

be expected to increase the PPV and decrease the NPV. While this could explain the changes in predictive values between the pre- and post-implementation groups in our study, other factors may have been responsible for the small predictive values differences including the increase in surgically proven appendicitis cases among the Appy-Score 3 exams in the post-implementation group. Although we cannot exclude possible effects of selection bias on study results, the large increase in Appy-Score use between July (24%, 37/155) and September (89%, 226/254) suggests that most of the US exams reported without use of the scoring system were the result of incomplete implementation, i.e. lack of knowledge or understanding of the scoring system rather than selection based on exam characteristics.

There are several limitations of this study. While we did attempt to contact all patients discharged from the emergency department to gain an accurate assessment of the true negatives in our study, it is possible that the false-negative population is underrepresented. There may be bias in the assignation of the Appy-Score to the pre-implementation ultrasounds. However, this was mitigated by a consensus review in equivocal studies, and the proportion of patients in each Appy-Score stratum did not change after implementation. Finally, the findings in this study may not be generalizable to other institutions that have differing levels of expertise in conducting and interpreting pediatric ultrasonography, that do not use US as a first-line diagnostic study for suspected appendicitis or that do not have 24-h US capabilities.

Future study and monitoring of the effect of this Appy-Score process improvement initiative are needed. We are investigating the effect of the Appy-Score on other emergency department processes, including the time to consultation and operation. We are examining those findings in patients with an Appy-Score of 4 to determine predictors of appendicitis when the study is found to be equivocal in order to provide additional clarity regarding the likelihood of appendicitis. Finally, we are working to improve and clarify our criteria for perforated appendicitis as this may have a future impact on appendicitis treatment pathways for those with early, simple disease.

Conclusion

Use of a risk-stratification scoring system and structured template for reporting US exam results for suspected pediatric appendicitis allows communication of appendicitis likelihood to the treating physician. The categorical data from these structured reports may be used in future longitudinal studies to evaluate findings on equivocal US exams and determine the accuracy of US for differentiating perforated and non-perforated appendicitis.

Acknowledgments Dr. Robert C. Orth is funded in part by the Association of University Radiologists GE Radiology Research Academic Fellowship.

Conflicts of interest None

References

1. Saucier A, Huang EY, Emeremni CA et al (2014) Prospective evaluation of a clinical pathway for suspected appendicitis. Pediatrics 133:e88–e95
2. Kim ME, Orth RC, Fallon SC et al (2015) Performance of CT examinations in children with suspected acute appendicitis in the community setting: a need for more education. AJR Am J Roentgenol 204:857–860
3. Orth RC, Guillerman RP, Zhang W et al (2014) Prospective comparison of MR imaging and US for the diagnosis of pediatric appendicitis. Radiology 272:233–240
4. Goldin AB, Khanna P, Thapa M et al (2011) Revised ultrasound criteria for appendicitis in children improve diagnostic accuracy. Pediatr Radiol 41:993–999
5. Worrell JA, Drolshagen LF, Kelly TC et al (1990) Graded compression ultrasound in the diagnosis of appendicitis. A comparison of diagnostic criteria. J Ultrasound Med 9:145–150
6. Mittal MK, Dayan PS, Macias CG et al (2013) Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. Acad Emerg Med 20:697–702
7. Polites SF, Mohamed MI, Habermann EB et al (2014) A simple algorithm reduces computed tomography use in the diagnosis of appendicitis in children. Surgery 156:448–454
8. Saito JM, Yan Y, Evashwick TW et al (2013) Use and accuracy of diagnostic imaging by hospital type in pediatric appendicitis. Pediatrics 131:e37–e44
9. Rice-Townsend S, Barnes JN, Hall M et al (2014) Variation in practice and resource utilization associated with the diagnosis and management of appendicitis at freestanding children's hospitals: implications for value-based comparative analysis. Ann Surg 259: 1228–1234
10. Wan MJ, Krahn M, Ungar WJ et al (2009) Acute appendicitis in young children: cost-effectiveness of US versus CT in diagnosis–a Markov decision analytic model. Radiology 250:378–386
11. Ross MJ, Liu H, Netherton SJ et al (2014) Outcomes of children with suspected appendicitis and incompletely visualized appendix on ultrasound. Acad Emerg Med 21:538–542
12. Srinivasan A, Servaes S, Pena A et al (2015) Utility of CT after sonography for suspected appendicitis in children: integration of a clinical scoring system with a staged imaging protocol. Emerg Radiol 22:31–42
13. Dilley A, Wesson D, Munden M et al (2001) The impact of ultrasound examinations on the management of children with suspected appendicitis: a 3-year analysis. J Pediatr Surg 36:303–308
14. Axelrod DA, Sonnad SS, Hirschl RB (2000) An economic evaluation of sonographic examination of children with suspected appendicitis. J Pediatr Surg 35:1236–1241
15. Elikashvili I, Tay ET, Tsung JW (2014) The effect of point-of-care ultrasonography on emergency department length of stay and computed tomography utilization in children with suspected appendicitis. Acad Emerg Med 21:163–170
16. Pacharn P, Ying J, Linam LE et al (2010) Sonography in the evaluation of acute appendicitis: are negative sonographic findings good enough? J Ultrasound Med 29:1749–1755
17. Larson DB, Trout AT, Fierke SR et al (2015) Improvement in diagnostic accuracy of ultrasound of the pediatric appendix through

the use of equivocal interpretive categories. AJR Am J Roentgenol 204:849–856
18. Estey A, Poonai N, Lim R (2013) Appendix not seen: the predictive value of secondary inflammatory sonographic signs. Pediatr Emerg Care 29:435–439
19. Coyne SM, Zhang B, Trout AT (2014) Does appendiceal diameter change with age? A sonographic study. AJR Am J Roentgenol 203: 936–945
20. Prendergast PM, Poonai N, Lynch T et al (2014) Acute appendicitis: investigating an optimal outer appendiceal diameter cut-point in a pediatric population. J Emerg Med 46:157–164
21. Trout AT, Towbin AJ, Zhang B (2014) Journal club: the pediatric appendix: defining normal. AJR Am J Roentgenol 202:936–945
22. Minneci PC, Sulkowski JP, Nacion KM et al (2014) Feasibility of a nonoperative management strategy for uncomplicated acute appendicitis in children. J Am Coll Surg 219:272–279
23. Blakely ML, Williams R, Dassinger MS et al (2011) Early vs interval appendectomy for children with perforated appendicitis. Arch Surg 146:660–665