

# Prospective Comparison of the Diagnostic Accuracy of MR Imaging versus CT for Acute Appendicitis

Michael D. Repplinger, MD, PhD • Perry J. Pickhardt, MD • Jessica B. Robbins, MD • Douglas R. Kitchin, MD • Tim J. Ziemlewicz, MD • Scott J. Hetzel, MS • Sean K. Golden, BA • John B. Harringa, BS • Scott B. Reeder, MD, PhD

From the Departments of Emergency Medicine (M.D.R., S.K.G., J.B.H., S.B.R.), Radiology (M.D.R., P.J.P., J.B.R., D.R.K., T.J.Z., S.B.R.), Biostatistics and Medical Informatics (S.J.H.), Medicine (S.B.R.), Medical Physics (S.B.R.), and Biomedical Engineering (S.B.R.), University of Wisconsin–Madison, 800 University Bay Dr, Suite 310, Mail Code 9123, Madison, WI 53705. Received August 3, 2017; revision requested September 15; final revision received January 23, 2018; accepted January 29. Address correspondence to M.D.R. (e-mail: [mdreppli@medicine.wisc.edu](mailto:mdreppli@medicine.wisc.edu)).

The authors acknowledge support from the National Institutes of Health, including the National Center for Advancing Translational Sciences (UL1TR000427, KL2TR000428) and the National Institute of Diabetes and Digestive and Kidney Diseases (K08DK111234, K24DK102595). The study also benefited from the University of Wisconsin–Madison's Department of Radiology Research and Development Fund. The University of Wisconsin–Madison Department of Radiology receives research support from GE Healthcare and Bracco Diagnostics. None of the funding sources were involved in any component of the conduct, analysis, or reporting of this study.

Conflicts of interest are listed at the end of this article.

Radiology 2018; 288:467–475 • <https://doi.org/10.1148/radiol.2018171838> • Content codes: **ER** **GI**

**Purpose:** To compare the accuracy of magnetic resonance (MR) imaging with that of computed tomography (CT) for the diagnosis of acute appendicitis in emergency department (ED) patients.

**Materials and Methods:** This was an institutional review board–approved, prospective, observational study of ED patients at an academic medical center (February 2012 to August 2014). Eligible patients were nonpregnant and 12- year-old or older patients in whom a CT study had been ordered for evaluation for appendicitis. After informed consent was obtained, CT and MR imaging (with non–contrast material–enhanced, diffusion-weighted, and intravenous contrast-enhanced sequences) were performed in tandem, and the images were subsequently retrospectively interpreted in random order by three abdominal radiologists who were blinded to the patients' clinical outcomes. Likelihood of appendicitis was rated on a five-point scale for both CT and MR imaging. A composite reference standard of surgical and histopathologic results and clinical follow-up was used, arbitrated by an expert panel of three investigators. Test characteristics were calculated and reported as point estimates with 95% confidence intervals (CIs).

**Results:** Analysis included images of 198 patients (114 women [58%]; mean age, 31.6 years  $\pm$  14.2 [range, 12–81 years]; prevalence of appendicitis, 32.3%). The sensitivity and specificity were 96.9% (95% CI: 88.2%, 99.5%) and 81.3% (95% CI: 73.5%, 87.3%) for MR imaging and 98.4% (95% CI: 90.5%, 99.9%) and 89.6% (95% CI: 82.8%, 94.0%) for CT, respectively, when a cutoff point of 3 or higher was used. The positive and negative likelihood ratios were 5.2 (95% CI: 3.7, 7.7) and 0.04 (95% CI: 0, 0.11) for MR imaging and 9.4 (95% CI: 5.9, 16.4) and 0.02 (95% CI: 0.00, 0.06) for CT, respectively. Receiver operating characteristic curve analysis demonstrated that the optimal cutoff point to maximize accuracy was 4 or higher, at which point there was no difference between MR imaging and CT.

**Conclusion:** The diagnostic accuracy of MR imaging was similar to that of CT for the diagnosis of acute appendicitis.

© RSNA, 2018

Acute appendicitis is a common cause of abdominal pain in patients seen in the emergency department (ED); 267 585 patients were diagnosed in the United States in 2012 (1). The use of history and physical examination alone leads to an incorrect diagnosis in 8%–30% of patients being evaluated for appendicitis, historically yielding a high negative laparotomy rate and even, rarely, death (2). Furthermore, clinical decision instruments do little to improve the accuracy of preoperative diagnosis (3). Failure to diagnose appendicitis has potential severe adverse outcomes, including appendiceal rupture, abscess formation, and, as noted, rarely death. For this reason, emergency medicine physicians and surgeons in the United States routinely rely on cross-sectional imaging to assist in the diagnosis. The American College of Radiology recommends the use of computed tomography (CT) for evaluation of most patients for appendicitis, except for pregnant and pediatric patients, largely because of its high diagnostic accuracy (4,5). The use of routine CT has been shown to decrease negative laparotomy rates (6–8), although this has

led to a substantial increase in the number of CT examinations performed (9–11).

Both scientific and lay publications have paid increased attention to the potential harms associated with the ionizing radiation exposure from CT (9). A typical abdominal and pelvic CT scan exposes patients to 10 mSv of radiation, which has been estimated by some to induce cancer in one patient for every 2000 scanned (12). Furthermore, epidemiologic studies have estimated that 1.5%–2% of all cancers in the United States may be directly attributable to CT scans (9). Consequently, imaging modalities that do not expose patients to ionizing radiation, including ultrasonography (US) and magnetic resonance (MR) imaging, have been proposed as alternatives to achieve the same benefits as CT. Enthusiasm for the use of US for the detection of appendicitis has been dampened by its widely variable, and decidedly inferior, diagnostic accuracy compared with CT (13–15). Although currently recommended as the first-line imaging examination for suspected appendicitis in children, the ability of US to depict the appendix,

## Abbreviations

AUC = area under the ROC curve, CI = confidence interval, ED = emergency department, ROC = receiver operating characteristic

## Summary

The diagnostic accuracy of contrast-enhanced MR imaging for the detection of acute appendicitis was similar to the accuracy previously reported for CT and was not statistically different from the accuracy of concurrently performed CT studies in this study cohort.

## Implications for Patient Care

- Use of MR imaging for the primary evaluation of acute appendicitis, particularly for patients who do not require sedation, may be appropriate in centers and practices with adequate resources and experience.
- Minimal training was required to instruct radiologists in interpreting abdominal MR images for the evaluation of acute appendicitis.

particularly when the structure is normal, is limited, even in the most experienced hands (16). Contrarily, recent studies evaluating MR imaging in diagnosis of appendicitis have shown promise, with diagnostic accuracy of MR imaging similar to that of CT (17,18).

Use of non-contrast material-enhanced MR imaging to diagnose appendicitis in pregnant women is well established, and in pediatric patients is gaining traction (19–22). Furthermore, recent meta-analyses of the accuracy of MR imaging in the diagnosis or exclusion of appendicitis, and in the identification of alternative diagnoses, in nonpregnant patients have demonstrated high diagnostic performance (23,24). However, no study to date has systematically compared the diagnostic accuracy of CT with MR imaging in a prospectively enrolled cohort, to our knowledge. In addition, the usefulness of gadolinium-based contrast agents with MR imaging has received little attention, yet such agents are regularly used in MR enterography for the follow-up of patients with inflammatory bowel disease. Therefore, the purpose of our prospective study was to determine the diagnostic accuracy of intravenous contrast-enhanced MR imaging, with direct comparison to intravenous contrast-enhanced CT, in diagnosis of acute appendicitis in adults and adolescents presenting in the ED with abdominal pain.

## Materials and Methods

### Study Design and Setting

This was a prospective, observational study of a convenience sample of patients being evaluated for appendicitis at an academic medical center from February 2012 to August 2014. The study was compliant with the Health Insurance Portability and Accountability Act and was approved by the institutional review board. Written informed consent was obtained for all adult subjects. In the case of minors, written informed consent was obtained from at least one parent or guardian, and written informed assent was obtained from the minor.

### Sample Selection

Patients were eligible for enrollment if they were at least 12 years old and had been ordered to undergo CT for evaluation for

appendicitis during study hours (in general, weekdays 7 am to 11 pm and weekends 7 am to 3 pm until May 2014, at which point an in-house MR technologist was available 24 hours per day). Our age cutoff point was selected with input from pediatric emergency medicine physicians to minimize the need for sedation. Exclusion criteria included contraindications to either gadolinium-based contrast material administration or MR imaging (eg, metallic implant) or the inability to provide informed consent or assent. The ED treatment team screened patients for eligibility.

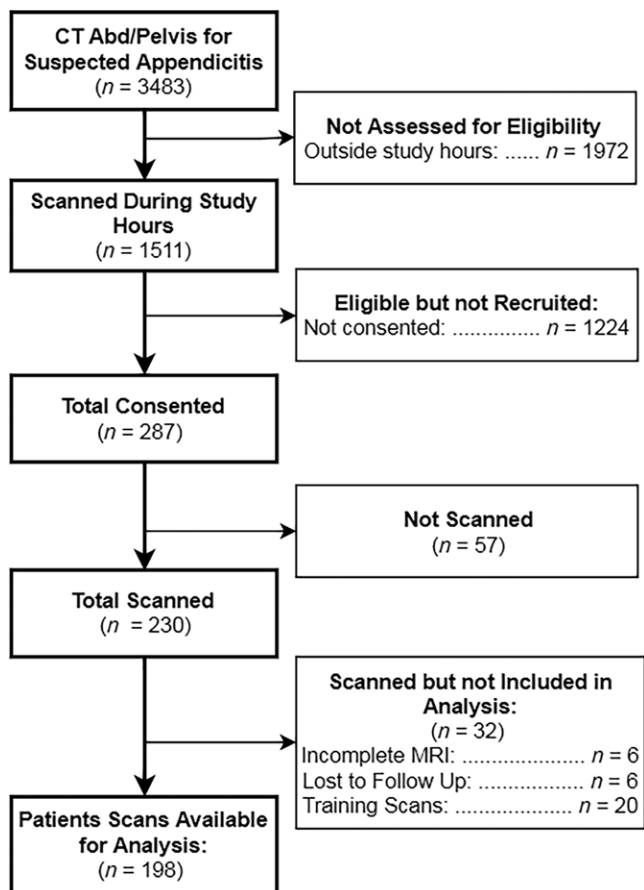
### Study Protocol

All patients underwent intravenous contrast-enhanced CT of the abdomen and pelvis, which for adult patients included ingestion of oral contrast agent (5). After CT, patients underwent the research MR imaging protocol and then proceeded back to the ED. If patients developed any signs of instability during the research study, if they decided to discontinue study participation, or if the surgical services asserted that the MR imaging procedure could negatively affect patient care, the protocol was terminated and these patients were brought back to the ED. All imaging-based clinical decisions in the ED were made solely on the basis of the clinical CT findings. The mean time interval between completion of the CT protocol and beginning of the MR imaging protocol was 61 minutes (95% confidence interval [CI]: 53 minutes, 68 minutes).

Subsequent to the index ED visit, three fellowship-trained abdominal radiologists (J.B.R., T.J.Z., and D.R.K., with 2–4 years of postresidency experience during the study) independently interpreted all MR and CT images, blinded to the patients' clinical information. It should be noted that although the MR and CT images were obtained prospectively, the image analysis was performed retrospectively. The use of three radiologists allowed for a consensus read to be used and for three paired interrater agreement measurements to be calculated for all images. Image headers were stripped of all identifiers, and patients were assigned randomly generated study identification numbers. CT and MR images in the same patient were read at different times, and in random order, by each radiologist. The randomly generated reading order was different for each radiologist.

### CT Protocol

CT examinations of the abdomen and pelvis were performed with a  $64 \times 0.625$ -mm detector configuration scanner (GE Healthcare, Waukesha, Wis) after administration of oral contrast material (Gastrografin; Bracco Diagnostics, Princeton, NJ) and intravenous iohexol (Omnipaque 300; GE Healthcare); imaging was performed in the portal venous phase (SmartPrep with automated scan initiation). Oral contrast agent, which is diluted in 1 L of polyethylene glycol and ingested during 1 hour, is routinely used at our institution for patients 18 years of age or older. Size-specific protocols for small, medium, and large body habitus ranged from 100 to 140 kVp, with modulated milliampere-seconds (Smart mA with mA range of 30–600 and noise index of 15–21). Images were reconstructed with 5-mm section thickness at 3-mm intervals with use of an



**Figure 1:** Flow diagram of study participants. *Abd* = abdomen.

iterative image reconstruction protocol (40% adaptive statistical iterative reconstruction blend with standard filtered back projection) in the axial, sagittal, and coronal planes. The estimated mean CT radiation dose in this patient cohort was 8.20 mSv ± 5.34 (range, 1.62–32.72 mSv).

### MR Imaging Protocol

MR imaging was performed with clinical 1.5-T units (Signa HDxt with a CRM or TwinSpeed gradients, Discovery MR450w; GE Healthcare) by using an eight-channel or 12-channel phased-array body coil. Details of the MR imaging protocol have been previously reported (25). In brief, it included multiple orthogonal planes of T2-weighted single-shot fast spin-echo imaging, precontrast and postcontrast T1-weighted three-dimensional spoiled gradient-echo imaging, and diffusion-weighted imaging. For contrast-enhanced T1-weighted imaging, 0.1 mmol/kg of gadobenate dimeglumine (MultiHance; Bracco Diagnostics) was administered intravenously at 2 mL/sec, followed by a 20-mL saline flush injected at the same rate. A typical protocol (non-contrast-enhanced, diffusion-weighted, and contrast-enhanced MR sequences) required about 30 minutes to complete.

### Image Analysis

Radiologists documented multiple end points for each image set (MR imaging and CT), including likelihood that the appen-

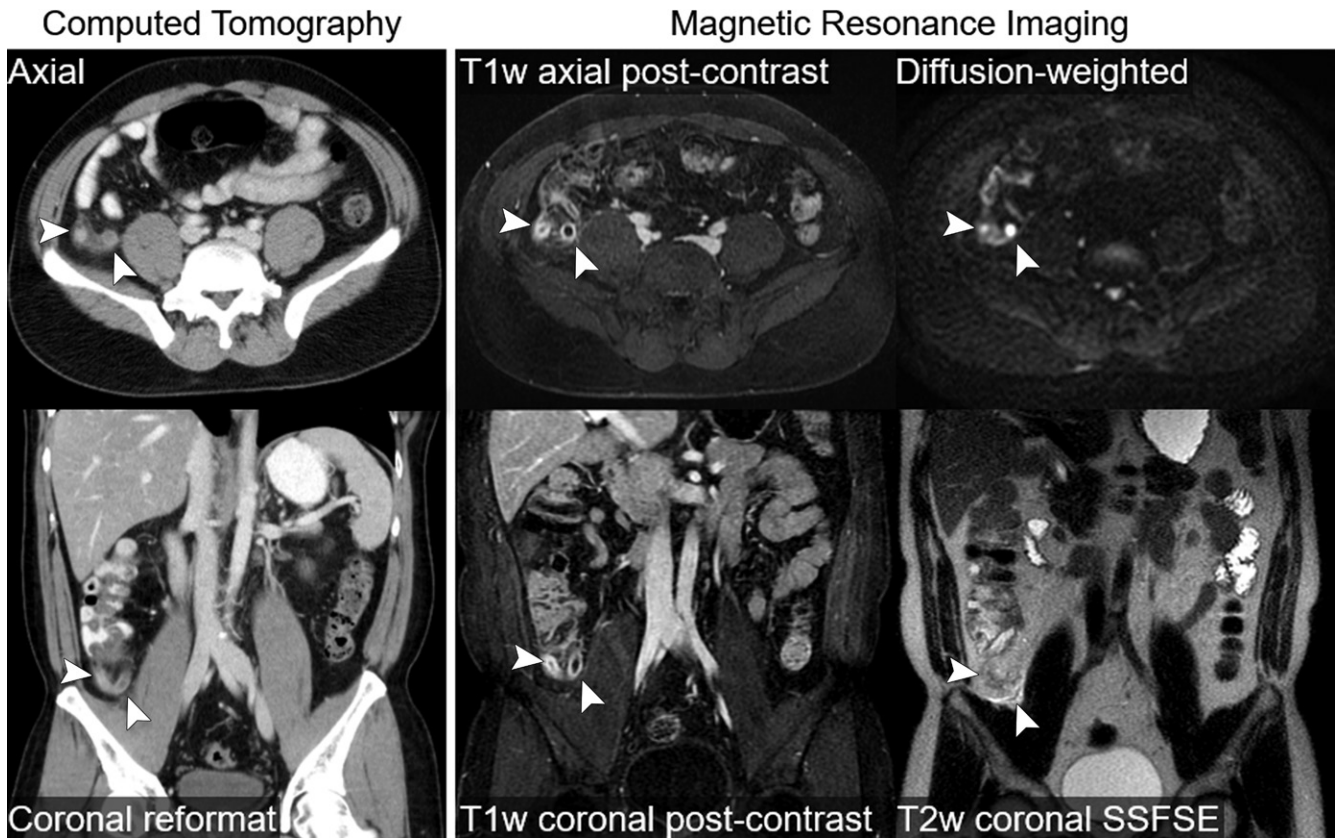
**Table 1: Patient Characteristics and Associated Prevalence of Appendicitis**

Age Group (y) and Sex	No. of Patients	Appendicitis Prevalence (%)
12–20		
F	4/27	15
M	9/20	45
21–30		
F	6/39	15
M	12/24	50
31–40		
F	8/21	38
M	8/22	36
41–50		
F	4/14	29
M	2/7	29
51–60		
F	5/9	56
M	1/5	20
61–70		
F	2/2	100
M	2/3	67
71–81		
F	0/2	0
M	1/3	33

dix was visualized (five-point scale: 1 = definitely not, 2 = probably not, 3 = not sure/possibly, 4 = probably, 5 = definitely); location of the appendix (retrocecal/paracolic, iliac fossa, medial extension, or combination); maximal short-axis width of the appendix (in millimeters); appendiceal wall thickening (yes, no, or unsure); fluid within the appendiceal lumen (yes, no, or unsure); presence of an appendicolith (yes, no, or unsure); degree of periappendiceal inflammation (none, mild, moderate, or severe); increased appendix signal at diffusion-weighted imaging (yes, no, or unsure); and overall likelihood of appendicitis (five-point scale: 1 = definitely not, 2 = probably not, 3 = not sure/possibly, 4 = probably, 5 = definitely). The time required to interpret each imaging study was also recorded. Although we did allow for comments regarding other nonappendicitis image findings, we did not systematically evaluate the accuracy of this MR imaging protocol for nonappendicitis diagnoses.

One trained data abstractor (J.B.H.) reviewed the electronic medical records of all enrolled patients by using a standard data extraction protocol. For the first 10 patients, the abstractor trained with the principal investigator to ensure data integrity. After these patients, he independently abstracted the data. For patients who underwent appendectomy, all surgical and pathologic reports were reviewed. For those who did not undergo appendectomy, the findings from all follow-up visits were reviewed to determine if the patient had subsequently been diagnosed with appendicitis or another cause of their symptoms. In addition, these patients underwent a follow-up phone interview 1 month after their index ED visit by means of a standardized script to assess whether they had been evaluated for the same symptoms subsequent to the index ED encounter, had





**Figure 2:** Intravenous contrast-enhanced CT and intravenous contrast-enhanced MR images in a 41-year-old woman with abdominal pain and uncomplicated appendicitis (arrowheads). Thin axial and reformatted coronal CT images acquired after administration of iodinated contrast material are shown. Selected images from MR imaging protocol are also shown, including coronal and axial postcontrast T1-weighted (*T1w*) images acquired after administration of gadolinium-based contrast material, a coronal T2-weighted (*T2w*) single-shot fast spin-echo (*SSFSE*) image, and an axial diffusion-weighted image. Both CT and MR images accurately depict appendiceal wall thickening, periappendiceal stranding, and mucosal enhancement. The T2-weighted MR image also depicts periappendiceal fluid, and the inflamed appendix is very conspicuous on the diffusion-weighted MR image.

undergone interval appendectomy, or had been diagnosed with another cause of their abdominal pain. Three of the authors (two radiologists [P.J.P. and S.B.R.] and one emergency medicine physician [M.D.R.]), none of whom interpreted the images, served as an expert panel to determine true disease state (appendicitis or not) for each patient, taking all follow-up data into account.

### Sample Size Calculation

The primary outcome for our study was the calculated sensitivity and specificity of MR imaging in diagnosis of appendicitis. Previous studies have reported that CT has both sensitivity and specificity ranging from 90% to 98% (4,14,15); the prevalence of appendicitis in the studies of MR test accuracy has typically ranged from 30% to 40% (4,18,26). The following parameters were used to calculate sample size based on the method described by Buderer (27): true sensitivity and specificity, 92%; desired 95% CI width, 5%; and prevalence of disease, 35%. This yielded a projected need for at least 174 patients for the study to be adequately powered.

### Statistical Analysis

The primary outcome was sensitivity and specificity of MR imaging for the diagnosis of appendicitis. Because we designed the

study to include three radiologists' interpretations, the consensus interpretation (meaning that at least two radiologists agreed on the presence or absence of appendicitis) was used as the primary outcome measure. We had a priori set a score of 3 or higher as reflecting a positive test result for image interpretation. However, receiver operating characteristic (ROC) curve analysis with the Youden method for maximizing sensitivity and specificity determined that a score of 4 or higher was the best threshold. We therefore calculated the test characteristics, which are reported as point estimates with 95% CIs, at both the 3 or higher and the 4 or higher thresholds by using the reference standard described earlier. ROC curves for the likelihood of appendicitis are presented, along with their corresponding areas under the curve (AUC). We used the McNemar test to evaluate the difference between the test characteristics of MR imaging and CT.

We also performed several planned secondary analyses. The mean time to interpret each imaging set was estimated with repeated-measures analysis of variance, with imaging method as a fixed effect and radiologist and subject as random effects. Interrater agreement was assessed with the Cohen  $\kappa$  statistic. All analyses were performed with R, version 3.1.0 ([www.r-project.org](http://www.r-project.org)).

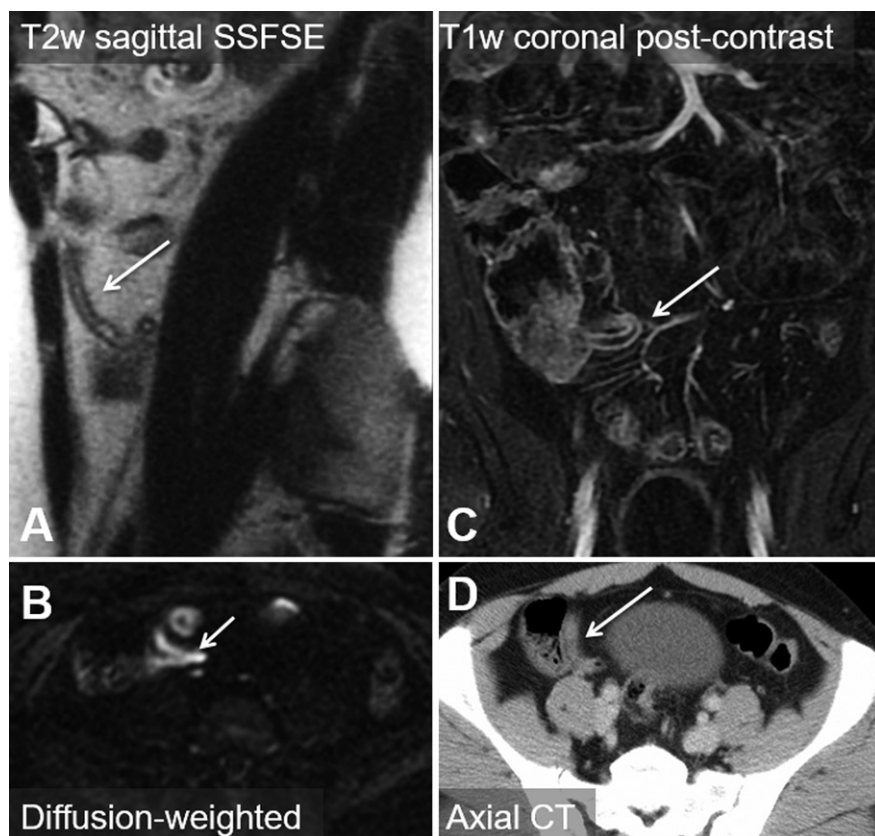
## Results

### Characteristics of Study Participants

We enrolled and imaged 230 patients from February 2012 to August 2014. Image sets from the first 20 subjects were used as a “training set” for our study radiologists. Six subjects were excluded because of incomplete MR imaging examinations, leaving images from 204 patients. Another six subjects were lost to follow-up and were also excluded, leaving 198 subjects for our final analysis (Fig 1). In this cohort, there were 114 women (58%), the mean age was 31.6 years ± 14.2 (range, 12–81 years), and the prevalence of appendicitis was 32.3% (64 of 198). Demographic information is presented in Table 1. During study hours, another 1224 patients were potentially eligible but were not enrolled. Although not systematically evaluated, the frequently cited reason among the ED staff for missing these potential enrollees was competing clinical duties, given that enrollment was primarily done by faculty in the ED. An additional 57 patients consented to enrollment but were not imaged. In this nonenrolled or not scanned cohort, there were 756 women (59%), the mean age was 38.2 years (range, 12–97 years), and the prevalence of appendicitis was 17.9%. There were no adverse events observed in study participants. Example MR and CT images in a single patient are shown in Figure 2. The value of diffusion-weighted imaging is more apparent in the example provided in Figure 3.

### Main Results

At the 3 or higher threshold, the consensus interpretation of the three radiologists had sensitivity and specificity, respectively, of 96.9% (95% CI: 88.2%, 99.5%) and 81.3% (95% CI: 73.5%, 87.3%) for MR imaging and 98.4% (95% CI: 90.5%, 99.9%) and 89.6% (95% CI: 82.8%, 94.0%) for CT. At the 4 or higher level, consensus interpretation had sensitivity and specificity of 96.9% (95% CI: 88.2%, 99.5%) and 89.6% (95% CI: 82.8%, 94.0%) for MR imaging and 98.4% (95% CI: 90.5%, 99.9%) and 93.3% (95% CI: 87.3%, 96.7%) for CT. The 2 × 2 tables for CT and MR imaging at the 4 or higher cutoff level for each radiologist and for consensus interpretation are presented in Table 2. Predictive values and likelihood ratios for each image type are also reported in Table 3. When comparing test characteristics of MR imaging versus CT, there was a statistically significant difference in specificity of 8.2% (95% CI: 1.8%, 14.6%) and a positive likelihood ratio of 0.55 (95% CI: 0.35, 0.57), favoring CT at the a priori cutoff level of 3 or higher, but



**Figure 3:** Images in 24-year old man with acute nontraumatic right lower quadrant pain. A, Sagittal T2-weighted (T2w) single-shot fast spin-echo (SSFSE) MR image, and C, T1-weighted (T1w) coronal postcontrast MR image reveal a very long appendix with only slight contrast agent uptake, fluid filling, and no obvious wall thickening. B, Diffusion-weighted MR image shows restricted diffusion, leading to a diagnosis of appendicitis. D, Corresponding axial CT image. Arrow = appendix.

**Table 2: 2 × 2 Table for Each Radiologist and Consensus Reading for Both CT and MR Imaging**

Reading and Finding	MR Imaging		CT	
	Test (+)	Test (–)	Test (+)	Test (–)
<b>Radiologist 1</b>				
Appendicitis	60	19	60	12
No appendicitis	4	115	4	122
<b>Radiologist 2</b>				
Appendicitis	62	22	62	8
No appendicitis	2	112	2	126
<b>Radiologist 3</b>				
Appendicitis	62	15	62	6
No appendicitis	2	119	2	128
<b>Consensus</b>				
Appendicitis	62	2	62	2
No appendicitis	14	120	9	125

Note.—Data are numbers of patients.

this was not observed at the optimal cutoff level of 4 or higher. There was no statistical difference in the number of scans rated 3 (unsure/possible appendicitis) by consensus for MR imaging

**Table 3: Test Characteristics of MR Imaging and CT**

Threshold	Sensitivity	Specificity	PPV	NPV	LR(+)	LR(-)
<b>Likelihood <math>\geq 3</math></b>						
MR	96.9 (88.2, 99.5)	81.3 (73.5, 87.3)	71.3 (60.4, 80.2)	98.2 (93.0, 99.7)	5.2 (3.7, 7.7)	0.04 (0.00, 0.11)
CT	98.4 (90.5, 99.9)	89.6 (82.8, 94.0)	81.8 (71.0, 89.4)	99.2 (94.8, 100)	9.4 (5.9, 16.4)	0.02 (0.00, 0.06)
<i>P</i> value	.564	.012	.010	.511	.011	.522
<b>Likelihood <math>\geq 4</math></b>						
MR	96.9 (88.2, 99.5)	89.6 (82.8, 94.0)	81.6 (70.7, 89.2)	98.4 (93.6, 99.7)	9.3 (5.9, 16.4)	0.03 (0.00, 0.09)
CT	98.4 (90.5, 99.9)	93.3 (87.3, 96.7)	87.5 (77.1, 93.8)	99.2 (95.0, 100)	14.7 (8.5, 32.7)	0.02 (0.00, 0.06)
<i>P</i> value	.564	.166	.152	.541	.156	.55

Note.—Except where otherwise indicated, data are percentages, with 95% confidence intervals in parentheses. LR(+) = positive likelihood ratio, LR(-) = negative likelihood ratio, NPV = negative predictive value, PPV = positive predictive value.

(15 [7.6%]) and CT (six [3.0%];  $P = .646$ ). We also performed a post-hoc power calculation for a test of noninferiority in the consensus sensitivity between MR imaging and CT. Assuming that an acceptable difference in sensitivity could be no larger than 5% and by using the data that we collected, we would have approximately 91% power with 134 subjects if we were to perform this study again.

### Unenhanced versus Contrast-enhanced MR Accuracy

The test characteristics for non-contrast-enhanced MR images by consensus interpretation at the 3 or higher cutoff level were sensitivity of 95.3% (95% CI: 86%, 98.8%), specificity of 82.8% (95% CI: 75.1%, 88.6%), positive predictive value of 72.6% (95% CI: 61.6%, 81.5%), negative predictive value of 97.4% (95% CI: 91.9%, 99.3%), positive likelihood ratio of 5.6 (95% CI: 4.0, 8.5), and negative likelihood ratio of 0.06 (95% CI: 0, 0.13). In comparing non-contrast-enhanced MR imaging and contrast-enhanced MR imaging at the 3 or higher cutoff level, the sensitivities (95.3% vs 96.9%,  $P = .317$ ), specificities (82.8% vs 81.3%,  $P = .564$ ), positive likelihood ratios (5.2 vs 5.6,  $P = .644$ ), and negative likelihood ratios (0.04 vs 0.06,  $P = .344$ ) had no statistically significant difference. Similar results were found when comparing non-contrast-enhanced MR test characteristics and contrast-enhanced MR imaging characteristics at the 4 or higher cutoff level.

### Individual Radiologist Accuracy

For a threshold of 3 or higher, the individual sensitivity and specificity of MR imaging ranged from 95.3% to 98.4% and from 77.6% to 80.6%, respectively, among the three radiologists. The corresponding individual sensitivity and specificity values for CT were 95.3%–98.4% and 85.8%–89.6%, respectively. At the 4 or higher threshold, the sensitivity and specificity of the MR imaging protocol ranged from 93.8% to 96.9% and from 83.6% to 88.8% for each radiologist, and from 93.8% to 96.9% and from 91.0% to 95.5% for CT, respectively.

### Interrater Agreement Measurement

We used the Cohen  $\kappa$  statistic to measure interrater agreement of MR imaging and CT for diagnosis of appendicitis. At a threshold of 3 or higher, the estimates for  $\kappa$  values ranged

from 0.69 to 0.77 for MR imaging and from 0.78 to 0.87 for CT; they ranged from 0.78 to 0.82 for MR imaging and from 0.85 to 0.91 for CT at the 4 or higher cutoff value. The full comparison is shown in Table 4.

### ROC Curve Analysis

Individual radiologist AUCs for the MR imaging ROC curves (Fig 4) had values of 0.91–0.96 (AUC for all radiologists = 0.93), and the AUCs for the CT ROC curves had values of 0.96–0.98 (AUC for all radiologists = 0.97).

### Time for Interpretation

The mean time required to interpret the contrast-enhanced MR protocol was 4.33 minutes (95% CI: 4.21, 4.45), whereas CT findings took 2.06 minutes (95% CI: 1.95, 2.18) to interpret (difference: 2.27 minutes [95% CI: 2.12, 2.41];  $P < .001$ ).

### Discussion

In our study, we demonstrated that the diagnostic accuracy of intravenous contrast-enhanced MR imaging for the diagnosis of acute appendicitis was similar to the accuracy previously reported for CT and was not statistically different from the accuracy of concurrently performed CT examinations in our study cohort. In particular, the sensitivity and specificity of MR imaging and CT were not statistically different at the optimal cutoff value of 4 or higher in our prospectively enrolled population (although study interpretations were performed retrospectively, after the index ED visit). Moreover, the interrater agreement for both CT and MR imaging was substantial among our three radiologists, supporting the assertion that radiologists can reliably interpret these images in a consistent manner. Based on these results, we propose that intravenous contrast-enhanced MR imaging may be a viable initial imaging option for the evaluation of appendicitis in the ED in nonpregnant patients.

Although several studies have evaluated the efficacy of MR imaging in the diagnosis of appendicitis, the majority were conducted in select populations including pregnant patients, pediatric patients, and patients already scheduled to undergo appendectomy. Our study addressed this by keeping our enrollment criteria broad: We limited recruitment



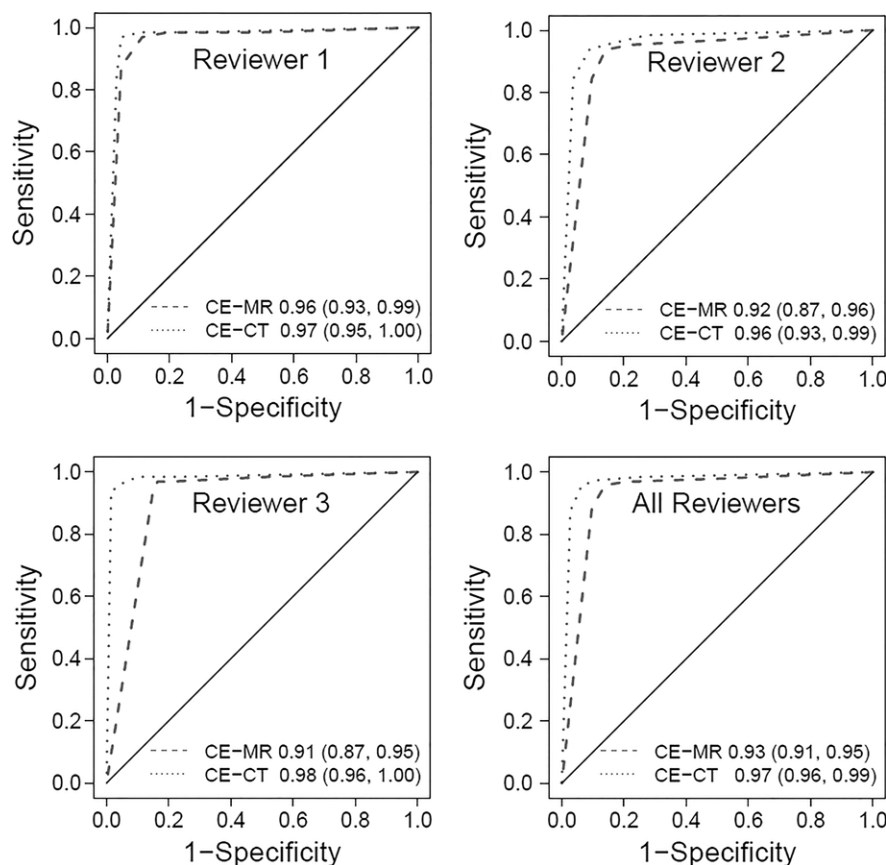
**Table 4: Interrater Agreement Reported as  $\kappa$  Values**

Modality and Radiologist	$\geq 3$ Threshold		$\geq 4$ Threshold	
	Radiologist 1	Radiologist 2	Radiologist 1	Radiologist 2
<b>MR imaging</b>				
Radiologist 2	0.78 (0.64, 0.92)	...	0.82 (0.68, 0.96)	...
Radiologist 3	0.76 (0.62, 0.90)	0.69 (0.56, 0.83)	0.81 (0.67, 0.95)	0.78 (0.64, 0.92)
<b>CT</b>				
Radiologist 2	0.82 (0.68, 0.96)	...	0.85 (0.71, 0.99)	...
Radiologist 3	0.78 (0.64, 0.92)	0.87 (0.73, 1.00)	0.85 (0.71, 0.99)	0.91 (0.77, 1.00)

Note.—Data in parentheses are 95% confidence intervals for each point estimate.

only to patients aged 12 years or older to minimize the need for sedation. It should be noted, however, that we recently published an analysis of the test accuracy of our MR imaging protocol in patients younger than 21 years (28). Those patients were also included in this study cohort, although that analysis also included the interpretations of three pediatric radiologists. Another distinction between our study and others is that unlike several previous studies in which a prevalence of appendicitis exceeding 60% was reported (29–31), our cohort had a prevalence of 32%, which is similar to our institutionally observed prevalence of 25% in patients undergoing CT for suspected appendicitis (4). We also attempted to assess the value of adding intravenous contrast enhancement to the MR imaging protocol, which is not standard among centers in which MR imaging is used to diagnose appendicitis (usually for pregnant patients and children). Although this study was underpowered for such an analysis, we found no statistically significant difference.

We directly compared prospectively acquired MR images with CT scans for all patients by using the retrospective interpretations of three faculty radiologists who were blinded to the clinical outcomes of the patients and who interpreted the images in random order. The largest competing study, to our knowledge, by Leeuwenburgh et al (17), did not uniformly obtain CT scans for comparison in all patients. Furthermore, unlike previous studies in which image interpretation was provided by only one radiologist, two radiologists who were able to consult with each other to resolve disagreements, or a pool of radiologists, our study required that the same three radiologists interpret all MR and CT images, permitting evaluation of interradiologist agreement for both MR imaging and



**Figure 4:** Graphs show receiver operating characteristic (ROC) curves for each radiologist and for all three radiologists together. Graphs were generated by using a five-point scale described in the Materials and Methods section (1 = definitely not appendicitis, 2 = probably not appendicitis, 3 = unsure/possible appendicitis, 4 = probably appendicitis, 5 = definitely appendicitis). Areas under the ROC curve and 95% confidence intervals are reported in the graphs. CE = contrast enhanced.

CT. We were also able to assess the time required to interpret these images and found a small difference between the interpretation times, with MR image interpretation taking about 2 minutes longer than CT interpretation. However, because both examinations were performed in tandem in every patient, we were unable to assess the time it takes to complete the examination (ie, from order to image acquisition completion), which is likely more important from a clinical perspective. This should be an area of study moving forward.

Our study had several limitations. First, we enrolled a relatively small number of patients (given the prevalence of appendicitis) by using convenience sampling, which led to a study population that was younger and had a higher prevalence of appendicitis than those patients who were eligible but who were not enrolled or not imaged. We believe this was largely because screening and frequently even the obtaining of patient consent were performed by emergency medicine physicians, who had to balance study recruitment with their clinical duties. We also limited enrollment to study hours, which in general did not include nights or weekends. Another limitation was the fact that the study was performed at a single academic medical center with a strong MR imaging presence, which may limit generalizability. However, although our study radiologists were fellowship trained in abdominal imaging, we do not routinely perform MR imaging to diagnose appendicitis, so they required some training. We do not know if additional training would have further improved their performance, although there is some evidence that this is the case (32). Third, our analysis focused on the performance of intravenous contrast-enhanced MR imaging for the diagnosis of appendicitis but did not examine the accuracy for the diagnosis of other causes of abdominal pain. This is a highly relevant and important avenue of research but is beyond the scope of this article. Fourth, the fact that only CT was used for clinical purposes (not MR imaging) yields the potential for incorporation bias. Furthermore, because some mild cases of appendicitis may self-resolve, there is the potential for differential verification bias (ie, no further evaluation was conducted for patients who had resolution of symptoms but who actually did have appendicitis at the index encounter). Finally, all patients underwent CT before MR imaging per our study design. For most, that mandated that oral contrast material be ingested. The effect of this oral contrast material on the diagnostic accuracy observed for the MR imaging protocol was not evaluated in our study; this is another important factor to investigate in the future.

In summary, the accuracy of contrast-enhanced MR imaging for the diagnosis of acute appendicitis in our prospectively enrolled cohort of ED patients with suspected appendicitis was similar to the accuracy previously reported for CT. We observed no statistically significant differences in the test characteristics of MR imaging versus CT at the 4 or higher cutoff level. We conclude that MR imaging may be a suitable alternative for the evaluation of acute appendicitis when MR imaging availability and expertise exist, particularly in patients who are not expected to require sedation.

**Author contributions:** Guarantors of integrity of entire study, M.D.R., S.B.R.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, M.D.R., P.J.P., S.B.R.; clinical studies, M.D.R., P.J.P., J.B.R., D.R.K., T.J.Z., S.K.G., S.B.R.; statistical analysis, S.J.H., J.B.H., S.B.R.; and manuscript editing, M.D.R., P.J.P., J.B.R., D.R.K., T.J.Z., S.J.H., J.B.H., S.B.R.

**Disclosures of Conflicts of Interest:** M.D.R. Activities related to the present article: disclosed no relevant relationships. Activities not related to the

present article: institution received Knowledge Translation grant in 2016 from the Emergency Medicine Foundation, which included travel for M.D.R. to Washington, DC for a conference with other researchers and visits to the National Institutes of Health. Other relationships: disclosed no relevant relationships. P.J.P. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Bracco and Check-Cap; is the cofounder of VirtuotCTC; owns stock or stock options in Celectar Biosciences, SHINE, and Elucet. Other relationships: disclosed no relevant relationships. J.B.R. disclosed no relevant relationships. D.R.K. disclosed no relevant relationships. T.J.Z. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Neuwave Medical. Other relationships: disclosed no relevant relationships. S.J.H. disclosed no relevant relationships. S.K.G. disclosed no relevant relationships. J.B.H. disclosed no relevant relationships. S.B.R. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is the cofounder of Calimetrix; is a shareholder in Celectar Biosciences and Elucet Medical; is a consultant for Parexel International. Other relationships: disclosed no relevant relationships.

## References

1. HCUPnet: A tool for identifying, tracking, and analyzing national hospital statistics. <http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=1E1EE590A4752BF5&Form=DispTab&JS=Y&Action=Accept>. Accessed August 17, 2015.
2. Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology* 2000; 215(2):337–348.
3. Golden SK, Harringa JB, Pickhardt PJ, et al. Prospective evaluation of the ability of clinical scoring systems and physician-determined likelihood of appendicitis to obviate the need for CT. *Emerg Med J* 2016;33(7):458–464.
4. Pickhardt PJ, Lawrence EM, Pooler BD, Bruce RJ. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. *Ann Intern Med* 2011;154(12):789–796, W-291.
5. Smith MP, Katz DS, Lalani T, et al. ACR Appropriateness Criteria right lower quadrant pain—suspected appendicitis. *Ultrasound Q* 2015;31(2):85–91.
6. Coursey CA, Nelson RC, Patel MB, et al. Making the diagnosis of acute appendicitis: do more preoperative CT scans mean fewer negative appendectomies? a 10-year study. *Radiology* 2010;254(2):460–468.
7. Raja AS, Wright C, Sodickson AD, et al. Negative appendectomy rate in the era of CT: an 18-year perspective. *Radiology* 2010;256(2):460–465.
8. Rao PM, Rhea JT, Novelline RA, Mostafavi AA, McCabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *N Engl J Med* 1998;338(3):141–146.
9. Brenner DJ, Hall EJ. Computed tomography: an increasing source of radiation exposure. *N Engl J Med* 2007;357(22):2277–2284.
10. Broder J, Warshauer DM. Increasing utilization of computed tomography in the adult emergency department, 2000–2005. *Emerg Radiol* 2006;13(1):25–30.
11. Repplinger MD, Weber AC, Pickhardt PJ, et al. Trends in the use of medical imaging to diagnose appendicitis at an academic medical center. *J Am Coll Radiol* 2016;13(9):1050–1056.
12. Dixon AK, Dendy P. Spiral CT: how much does radiation dose matter? *Lancet* 1998;352(9134):1082–1083.
13. Pickuth D, Heywang-Köbrunner SH, Spielmann RP. Suspected acute appendicitis: is ultrasonography or computed tomography the preferred imaging technique? *Eur J Surg* 2000;166(4):315–319.
14. Al-Khayal KA, Al-Omran MA. Computed tomography and ultrasonography in the diagnosis of equivocal acute appendicitis: a meta-analysis. *Saudi Med J* 2007;28(2):173–180.
15. van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester MA. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. *Radiology* 2008;249(1):97–106.
16. Binkovitz LA, Unsdorfer KM, Thapa P, et al. Pediatric appendiceal ultrasound: accuracy, determinacy and clinical outcomes. *Pediatr Radiol* 2015;45(13):1934–1944.
17. Leeuwenburgh MM, Wiarda BM, Wiezer MJ, et al. Comparison of imaging strategies with conditional contrast-enhanced CT and unenhanced MR imaging in patients suspected of having appendicitis: a multicenter diagnostic performance study. *Radiology* 2013;268(1):135–143.
18. Cobben L, Groot I, Kingma L, Coerkamp E, Puylaert J, Blickman J. A simple MRI protocol in patients with clinically suspected appendicitis: results in 138 patients and effect on outcome of appendectomy. *Eur Radiol* 2009;19(5):1175–1183.
19. Pedrosa I, Laforana M, Pandharipande PV, Goldsmith JD, Rofsky NM. Pregnant patients suspected of having acute appendicitis: effect of MR imaging on negative laparotomy rate and appendiceal perforation rate. *Radiology* 2009;250(3):749–757.
20. Rapp EJ, Naim F, Kadivar K, Davarpanah A, Cornfeld D. Integrating MR imaging into the clinical workup of pregnant patients suspected of having appendicitis is associated with a lower negative laparotomy rate: single-institution study. *Radiology* 2013;267(1):137–144.
21. Aspelund G, Fingeret A, Gross E, et al. Ultrasonography/MRI versus CT for diagnosing appendicitis. *Pediatrics* 2014;133(4):586–593.
22. Thieme ME, Leeuwenburgh MM, Valdehueza ZD, et al. Diagnostic accuracy and patient acceptance of MRI in children with suspected appendicitis. *Eur Radiol* 2014;24(3):630–637.



23. Duke E, Kalb B, Arif-Tiwari H, et al. A systematic review and meta-analysis of diagnostic performance of MRI for evaluation of acute appendicitis. *AJR Am J Roentgenol* 2016;206(3):508–517.
24. Replinger MD, Levy JF, Peethumnongsin E, et al. Systematic review and meta-analysis of the accuracy of MRI to diagnose appendicitis in the general population. *J Magn Reson Imaging* 2016;43(6):1346–1354.
25. Kinner S, Replinger MD, Pickhardt PJ, Reeder SB. Contrast-enhanced abdominal MRI for suspected appendicitis: how we do it. *AJR Am J Roentgenol* 2016;207(1):49–57.
26. Singh AK, Desai H, Novelline RA. Emergency MRI of acute pelvic pain: MR protocol with no oral contrast. *Emerg Radiol* 2009;16(2):133–141.
27. Buderer NM. Statistical methodology. I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;3(9):895–900.
28. Kinner S, Pickhardt PJ, Riedesel EL, et al. Diagnostic accuracy of MRI versus CT for the evaluation of acute appendicitis in children and young adults. *AJR Am J Roentgenol* 2017;209(4):911–919.
29. Avcu S, Çetin FA, Arslan H, Kemik Ö, Dülger AC. The value of diffusion-weighted imaging and apparent diffusion coefficient quantification in the diagnosis of perforated and nonperforated appendicitis. *Diagn Interv Radiol* 2013;19(2):106–110.
30. Nitta N, Takahashi M, Furukawa A, Murata K, Mori M, Fukushima M. MR imaging of the normal appendix and acute appendicitis. *J Magn Reson Imaging* 2005;21(2):156–165.
31. Zhu B, Zhang B, Li M, Xi S, Yu D, Ding Y. An evaluation of a superfast MRI sequence in the diagnosis of suspected acute appendicitis. *Quant Imaging Med Surg* 2012;2(4):280–287.
32. Leeuwenburgh MM, Wiarda BM, Bipat S, et al. Acute appendicitis on abdominal MR images: training readers to improve diagnostic accuracy. *Radiology* 2012;264(2):455–463.