Systematic Review and Meta-Analysis of the Accuracy of MRI to Diagnose Appendicitis in the General Population

Michael D. Repplinger, MD, MS,1,2* Joseph F. Levy, BS,3 Erica Peethumnongsin, MD, PhD,1 Megan E. Gussick, MD,1 James E. Svenson, MD, MS,1 Sean K. Golden, BA,1 William J. Ehlenbach, MD, MSc,4 Ryan P. Westergaard, MD, PhD,4 Scott B. Reeder, MD, PhD,1,2,4,5,6 and David J. Vanness, PhD3

Purpose: To perform a systematic review and meta-analysis of all published studies since 2005 that evaluate the accuracy of magnetic resonance imaging (MRI) for the diagnosis of acute appendicitis in the general population presenting to emergency departments.

Materials and Methods: All retrospective and prospective studies evaluating the accuracy of MRI to diagnose appendicitis published in English and listed in PubMed, Web of Science, Cinahl Plus, and the Cochrane Library since 2005 were included. Excluded studies were those without an explicitly stated reference standard, with insufficient data to calculate the study outcomes, or if the population enrolled was limited to pregnant women or children. Data were abstracted by one investigator and confirmed by another. Data included the number of true positives, true negatives, false positives, false negatives, number of equivocal cases, type of MRI scanner, type of MRI sequence, and demographic data including study setting and gender distribution. Summary test characteristics were calculated. Forest plots and a summary receiver operator characteristic plot were generated.

Results: Ten studies met eligibility criteria, representing patients from seven countries. Nine were prospective and two were multicenter studies. A total of 838 subjects were enrolled; 406 (48%) were women. All studies routinely used unenhanced MR images, although two used intravenous contrast-enhancement and three used diffusion-weighted imaging. Using a bivariate random-effects model the summary sensitivity was 96.6% (95% confidence interval [CI]: 92.3%–98.5%) and summary specificity was 95.9% (95% CI: 89.4%–98.4%).

Conclusion: MRI has a high sensitivity and specificity for the diagnosis of appendicitis, similar to that reported previously for computed tomography.

J. MAGN. RESON. IMAGING 2015;00:000–000.

In the United States in 2005, 38.8 million patients were seen in emergency departments (EDs) for abdominal pain.1 Appendicitis is a frequent cause of such visits, leading to 250,000 appendectomies performed annually.2 Diagnosing appendicitis using only clinical findings is inaccurate in as many as 30% of cases, and may lead to unnecessary surgery.3 Conversely, a missed diagnosis of appendicitis carries significant morbidity. Although initially suggested to be of significant value to aid in the diagnosis of appendicitis, clinical decision instruments like the Alvarado Score have not consistently been shown to be of benefit, even when compared to unstructured clinical gestalt.4,5 For these reasons, current practice relies on imaging to improve the accuracy of the diagnosis of acute appendicitis.

While ultrasound is a safe and generally effective imaging modality, its utility is limited because it is highly...
operator-dependent and has limited sensitivity and specificity for the diagnosis of appendicitis, particularly outside of the pediatric population. 

Emergency physician-performed ultrasound has been reported to have sensitivity 44–67%, specificity 85–98%, and accuracy 67%, although this specifically relates to clinicians with limited training in sonography. However, one study reports that even formal ultrasound performed by medical sonographers was unable to visualize the appendix in 45% of cases, yielding sensitivity 51.8% and specificity 81.4%. Further, although the test accuracy of ultrasound in the case of pediatric abdominal pain concerning appendicitis has long been viewed as superior to that of adults, the sensitivity of ultrasound has been reported to be as low as 35% in centers that do not use the technology often. Finally, nearly half of cases using ultrasound to diagnose appendicitis are either negative or nondiagnostic. Imaging guidelines recommend further evaluation in this situation, further limiting its usefulness.

Alternatively, computed tomography (CT), when compared to ultrasound, has sensitivity 94% versus 76% and specificity 81% versus 61%, respectively. Additionally, at least one study reports that the use of CT leads to changes in the treatment decisions of a majority of patients being evaluated for appendicitis. As a result, CT has become widely adopted as the primary imaging modality for detecting appendicitis in the United States, particularly for adults. This has contributed to the dramatic increase in CT use in the United States over the past 30 years, from 2 to 72 million scans annually. In a 5-year period (1999–2004), CT use increased from 51 to 76% for those eventually diagnosed with appendicitis in one study, while another study found that the use of CT in patients presenting to the ED with abdominal pain doubled over a 5-year period (2001–2005) to 22.5%. Although the use of CT in the diagnosis of pediatric appendicitis has been decreasing in recent years, roughly 40% of children are still undergoing CT imaging (2013). This rate of CT utilization has led to a significant increase in the use of ionizing radiation over time, which carries a potential risk of developing cancer, particularly in children and young adults. Specifically, the average effective radiation dose of an abdominopelvic CT for appendicitis is ~10 mSv, corresponding to an estimated excess risk of radiation-induced cancer of 1:2000. Although “low dose” CT protocols are becoming more commonplace, the test characteristics (sensitivity 92.5%, specificity 89%) are inferior to those previously reported with “standard dose” CT; although still superior to ultrasound (sensitivity 82.5%, specificity 82%). However, the negative laparotomy rate for patients undergoing “low dose” CT is similar to that of “standard dose” CT, which has been reported to be as low as 1.7% in a large, retrospective review.

Conversely, magnetic resonance imaging (MRI) is an alternative cross-sectional imaging method that uses no ionizing radiation. Historically, MRI has been limited by its cost and access, particularly from the ED. However, the cost of MRI has become more aligned with CT over time and increasingly available in recent years. In a survey of randomly sampled EDs in the United States, 86% were found to have access to MRI scanners, including 39% with 24/7 MRI availability.

The use of MRI to diagnose numerous emergent conditions, including appendicitis, has been evaluated in multiple recent studies. In fact, at least one study reported that MRI should be used preferentially to ultrasound due to MRI’s superior test characteristics and fewer inconclusive studies. Moreover, innovative techniques, including free-breathing methods and diffusion-weighted imaging, have shown promise to improve the accuracy of MRI to diagnose appendicitis. However, large, multicenter studies that are adequately powered to compare the sensitivity and specificity of MRI with that of CT for the diagnosis of appendicitis have yet to be published.

Therefore, the purpose of this study was to perform a systematic review and meta-analysis of the use of MRI to diagnose appendicitis in the general population, ie, not limited to one subpopulation such as pregnant patients or children. The primary outcomes of interest are sensitivity and specificity of MRI for this indication.

Materials and Methods

Literature Search

The design and results of this systematic review conformed to the recommendations outlined by Leeflang et al as well as the Cochrane Collaboration. Given that our meta-analysis does not qualify as human subjects research, it was exempt from Institutional Review Board (IRB) review.

A comprehensive literature search was performed on PubMed, Web of Science, Cinahl Plus, and the Cochrane Library using the search parameters “magnetic resonance imaging” and “appendicitis.” The search was restricted to articles written in the English language, involving human subjects, and published beginning in the year 2005. We limited studies to those published in the past decade to best represent current imaging protocols including diffusion-weighted imaging and images obtained with free-breathing. Case reports, case series, and review articles were excluded. Ancestral searching was performed by reviewing the bibliographies of articles identified through the original literature search. Those articles that were not already identified were made eligible for inclusion, as long as they fit the previously mentioned restrictions. Studies were only included if they specifically dealt with the diagnosis of acute appendicitis using MRI, although specific imaging sequences were not required for inclusion. Additionally, articles were only included if they had well-defined and acceptable reference standards such as a single imaging comparator or clinical follow-up (surgical findings, histopathological findings,
clinic visits, phone interviews, etc.). Finally, articles were required to provide absolute numbers of true positives, false positives, true negatives, false negatives, and equivocal cases so that pooled statistics could be calculated.

Studies restricted to specific subpopulations, such as children or pregnant women, were excluded because of potentially significant clinical heterogeneity due to the substantial anatomical differences compared to the general population and the potential for spectrum bias. Two authors performed independent reviews of the remaining articles and compiled a list of those meeting eligibility criteria. Articles were included in the final analysis when these authors agreed eligibility was appropriate; discrepancies were discussed and resolved by consensus.

Data were abstracted by one study author and confirmed by another for all included studies. Primary outcome data included true positives, false positives, true negatives, false negatives, number of equivocal cases, and total number of patients enrolled. In addition, authors abstracted the journal of publication, year of publication, gender and age of enrolled patients, years of enrollment, eligibility criteria, whether the study was prospective or retrospective, MRI scanner type, and MRI sequences performed. The list of articles and their abstracted data are listed in Table 1.

**Quality Assessment and Data Extraction**
The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) statement was used to rate the quality of each of the included studies.28 Two authors (other than the two who screened for eligibility) used this instrument to assess the quality of each included article independently. Disagreements were resolved by consensus of these two authors.

**Analysis**
To assess the accuracy of MRI to diagnose acute appendicitis correctly, we used a bivariate random-effects meta-analysis that analyzed sensitivity and specificity jointly. This procedure accounts for between-study variation and possible correlation between sensitivity and specificity. Because our data contain studies with zero cell counts (no false positive/negative results) we adopted the generalized linear mixed model approach, which does not require continuity correction.29 In addition, it has been suggested that this approach is preferred to the original bivariate model presented by Reitsma et al when cell counts are low.30

The bivariate generalized linear mixed model produces summary estimates of sensitivity and specificity with 95% confidence intervals (CI). Because all included studies have a common threshold to define positive and negative results, we are able to display a single joint summary point with a 95% elliptical joint confidence region on a summary receiver operating characteristic (ROC) curve space.26,31 Data analysis was performed using the “mada” package in R (R version 3.2.2) and RevMan 5.1.32

**Investigation of Heterogeneity**
We investigated variation across studies by observed study characteristics using subgroup analysis and meta-regression. First, we examined forest plots and ROC curves and then statistically assessed whether inclusion of each covariate in a meta-regression using the bivariate model significantly affected sensitivity and specificity. The continuous sources of heterogeneity examined were proportion of females in the enrolled population, average age of patients, and the observed prevalence of appendicitis in the studies. For visual inspection, the studies were split into subgroups at the observed median value. Statistically, the continuous characteristics were included as covariates in separate regressions and the $P$-values of effect on sensitivity and specificity were checked.

Finally, the magnetic field strength was investigated as a possible source of heterogeneity. Seven studies used 1.5T MRI scanners, while three studies used scanners with lower field strengths (1.0T, 0.5T) or a combination of scanners with field strengths less than 1.5T (0.23T, 0.6T, 1.5T). To assess this statistically, we constructed a categorical variable defined as whether studies only used a 1.5T scanner ($n = 7$) or not ($n = 3$).

**Results**

**Literature Search Results**
The initial PubMed literature search yielded 177 articles. Of these, 81 were unrelated to either the use of MRI or the diagnosis of appendicitis. Of those that were related to this topic, 50 were review articles, six were case reports, one was a description of a study protocol, and one was a discussion of how to teach radiologists to interpret MRI for the diagnosis of appendicitis, but did not examine the accuracy of MRI. Finally, there were two meta-analyses, published in 2010 and 2011. All of these articles were excluded, leaving 36 articles (Fig. 1). Finally, the Web of Science, Cinahl Plus, and Cochrane Library databases were searched for articles using the same search parameters. There were 113 articles found in the Web of Science, 31 in Cinahl Plus, and four in the Cochrane Library, all of which had already been identified by the original PubMed search.

Of the 36 articles that reported the use of MRI to diagnose appendicitis, 10 specifically enrolled children and 14 enrolled only pregnant women, and therefore were excluded from this analysis. Of the 12 remaining articles, three were reports derived from the same study population. Ultimately, 10 articles met all inclusion criteria.

**Characteristics of Included Studies**
A total of 838 patients were enrolled in the 10 studies included in this meta-analysis. The prevalence of appendicitis ranged from 25–80% with a mean prevalence of 57.7% (95% CI 44.7–70.7%). Women comprised 36–60% of study participants (mean 49.9%, 95% CI 43.6–56.2%). Most studies were prospective (9/10). These studies enrolled 798 (95.2%) of all patients included in this meta-analysis. Convenience sampling was used for most studies, although some used consecutive sampling, particularly when planned appendectomy was part of the inclusion criteria. Patients were enrolled from The Netherlands,33,34 Turkey,35–37 Denmark,38 Germany,39 Japan,40 China,41 and the United States.42 All studies used clinical follow-up as their reference standard for the purposes of calculating the test
<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Location</th>
<th>Study type</th>
<th>Study dates</th>
<th>Reference standard</th>
<th>Total patients</th>
<th>Prevalence (women %)</th>
<th>Mean (range) age (yrs)</th>
<th>Scanor type (manufacturer)</th>
<th>MRI sequences</th>
<th>Reviewer</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitta</td>
<td>2005</td>
<td>Japan</td>
<td>Prospective, single-center</td>
<td>Unknown</td>
<td>Surgical pathology or clinical follow up</td>
<td>37</td>
<td>78%</td>
<td>19 (52)</td>
<td>37.1 (16–69)</td>
<td>0.5T Gyrosan (Philips)</td>
<td>T1 SE, T2 FSE, T2 with fat saturation</td>
<td>Three experienced radiologists, by consensus</td>
</tr>
<tr>
<td>Cobben</td>
<td>2009</td>
<td>Netherlands</td>
<td>Prospective, single-center</td>
<td>1/05–10/06</td>
<td>Surgical pathology or clinical follow up</td>
<td>138</td>
<td>45%</td>
<td>80 (56)</td>
<td>29 (6–80)</td>
<td>1.0T (Siemens)</td>
<td>T1 SGRE, T2 FSE, T2 FSE with fat saturation</td>
<td>One of three GI radiologists (&gt; 5 years experience)</td>
</tr>
<tr>
<td>Singh</td>
<td>2009</td>
<td>USA</td>
<td>Retrospective, multi-center</td>
<td>2001–2007</td>
<td>Final diagnosis at hospital discharge</td>
<td>40</td>
<td>30%</td>
<td>Unknown</td>
<td>34 (11–69)</td>
<td>1.5T Excite T winspeed (GE Medical Systems)</td>
<td>T2 SSFSE with fat saturation, T2 FSE with fat saturation, STIR, pre-gadolinium T1, post-gadolinium T1</td>
<td>Two experienced radiologists (5 &amp; 10 years). A third radiologist (&gt; 20 yrs experience) if lack of consensus</td>
</tr>
<tr>
<td>Inci</td>
<td>2011</td>
<td>Turkey</td>
<td>Prospective, single-center</td>
<td>7-month period</td>
<td>Surgical pathology or clinical follow up</td>
<td>85</td>
<td>67%</td>
<td>40 (47)</td>
<td>26.5 ± 11.3 (14–72)</td>
<td>1.5 T Avanto (Siemens)</td>
<td>T1 FSE, T2 FSE with and without fat saturation</td>
<td>Two abdominal radiologists (8 years and 5 years experience)</td>
</tr>
<tr>
<td>Chabanova</td>
<td>2011</td>
<td>Denmark</td>
<td>Prospective, single-center</td>
<td>Unknown</td>
<td>Surgical pathology or clinical follow up</td>
<td>48</td>
<td>65%</td>
<td>29 (60)</td>
<td>57.1 (18–70)</td>
<td>0.23T and 0.6T Panorama (Philips), 1.5T Infinion (Philips), 1.5T Achieva (Philips)</td>
<td>T1 SE, T2 FSE, STIR</td>
<td>One radiologist (10 years of experience, 6 years experience in abdominal MRI) and a research fellow (3 year experience in abdominal MRI)</td>
</tr>
<tr>
<td>Inci</td>
<td>2011</td>
<td>Turkey</td>
<td>Prospective, single-center</td>
<td>11-month period</td>
<td>Surgical pathology or clinical follow up</td>
<td>119</td>
<td>60%</td>
<td>36</td>
<td>27 (17–72)</td>
<td>1.5 T Avanto (Siemens)</td>
<td>T1 FSE, T2 FSE with and without fat saturation, DWI</td>
<td>Two abdominal radiologists (8 years and 5 years experience)</td>
</tr>
<tr>
<td>Herehagen</td>
<td>2012</td>
<td>Germany</td>
<td>Prospective, single-center</td>
<td>02/2008 - 10/2008</td>
<td>Surgical pathology or clinical follow up</td>
<td>52</td>
<td>25%</td>
<td>21 (40)</td>
<td>44.7 (18–88)</td>
<td>1.5 T Magnetom Sonata (Siemens)</td>
<td>STIR, T2 FSE, bSSFP, 1.5T fat-saturated SGRE (before and after IV contrast), IV butylscopolamin used to prevent peristalsis.</td>
<td>One radiologist (11 years of experience in MRI). A second radiologist with 4 years of experience reviewed cases for inter-rater variability.</td>
</tr>
<tr>
<td>Zhu</td>
<td>2012</td>
<td>China</td>
<td>Prospective, single-center</td>
<td>9/09–9/11</td>
<td>Surgical pathology</td>
<td>41</td>
<td>80%</td>
<td>23 (56)</td>
<td>41.5 ± 11.3</td>
<td>1.5T Achieva Nova Dual (Philips)</td>
<td>T2 FSE, bSSFP with fat saturation</td>
<td>One GI radiologist with &gt;10 years experience.</td>
</tr>
<tr>
<td>Leeuwenburgh</td>
<td>2013</td>
<td>Netherlands</td>
<td>Prospective, multi-center</td>
<td>3/2010 - 9/2010</td>
<td>Expert panel reviewing surgery and clinical follow up</td>
<td>223</td>
<td>52%</td>
<td>132 (59)</td>
<td>35 (IQR 25–50)</td>
<td>1.5T Magnetom Avanto (Siemens)</td>
<td>T2 FSE with and without fat saturation, DWI</td>
<td>Two experienced radiologists (14 and 16 years experience in consensus)</td>
</tr>
<tr>
<td>Avcu</td>
<td>2013</td>
<td>Turkey</td>
<td>Prospective, single-center</td>
<td>03/2009 - 02/2010</td>
<td>Surgical pathology or clinical follow up</td>
<td>55</td>
<td>71%</td>
<td>26 (43)</td>
<td>35.6 ± 15.5 (17–83)</td>
<td>1.5T Magnetom Symphony (Siemens)</td>
<td>DWL bSSFP, STIR</td>
<td>One radiologist (experience level not given)</td>
</tr>
</tbody>
</table>

For reference, Prev = prevalence of appendicitis in the study population. Under the sequence column, RARE = rapid acquisition with relaxation enhancement; SPAIR = spectral selection attenuated inversion recovery; SENSE = sequence, sensitivity encoding; BTFE = balanced turbo field echo; SPIR = spectral pre-saturation and inversion recovery; SS-FSE = single shot fast spin echo; FSE = fast spin echo; STIR = short tau inversion recovery; SE = spin echo; DWI = diffusion-weighted imaging; True-FISP = T2-weighted True-fast imaging with steady state precession; TIRM = Turbo inversion-recovery in magnitude; TSE = Turbo spin echo, HF = half Fourier, GRE = gradient echo, FLASH = fast low angle shot, DWI = diffusion-weighted image.
characteristics of MRI to diagnose appendicitis. Using the QUADAS-2 assessment tool, the included studies were generally deemed to be at low or uncertain risk of bias (Fig. 2).

The majority of studies (7/10) used 1.5T MRI scanners, with Siemens being the most common manufacturer. The study from Denmark used a variety of scanners including a 0.23T Philips Panorama, 0.6T Philips Panorama, 1.5T Phillips Infinion, and a 1.5T Philips Achieva.38 Cobben et al used a 1.0T Siemens scanner.34 The study from Japan used a Philips 0.5T Gyroscan.40 Almost all studies used unenhanced imaging protocols, although some used contrast enhancement and diffusion-weighted imaging. For a complete list of sequences, please refer to Table 1.

Main Results
Forest plots of sensitivity and specificity depict individual study data in Fig. 3. Using a bivariate random-effects model to estimate sensitivity and specificity jointly, the pooled sensitivity was 96.6% (95% CI: 92.3–98.5%) and pooled specificity was 95.8% (95% CI: 89.4–98.4%). Figure 4 depicts all 10 studies' sensitivity and specificity on an ROC curve space, with square sizes scaled proportionately to the number of patients in each study. The summary point and uncertainty ellipse depict the joint summary estimate of the meta-analysis and the corresponding uncertainty around it.

We assessed outliers visually and by Cook's distance. The results from Chabanova et al38 (sensitivity of 85.5%, specificity 60.5%) had a Cook's distance greater than 4, suggesting a potential outlier. To assess its impact, we performed a robustness check removing the study, following the same approach as our full estimate. This yielded a pooled sensitivity of 97.4% (95% CI: 94.6–98.7) and pooled specificity of 96.0% (95% CI: 92.2–97.9).

In addition to calculating the pooled sensitivity and specificity of the included studies, the pooled likelihood ratios and predictive values were also calculated (Table 2).
Given that our analysis revealed an outlier among the data, we present both the results of all 10 articles and the results with the outlier data excluded.

**Heterogeneity**

For each observable study characteristic, we created subgroups split at the median value of the included studies. These categories (and values) were: proportion of females enrolled (51%), average age of study participants (35.6 years), and prevalence of appendicitis (64.5%). In the case of scanner field strength, values were dichotomized as either using only a 1.5T scanner ($n = 7$) or not ($n = 3$). We present results partitioned by these study characteristics in Table 3. Overall, these differences appear small, with the largest discrepancy occurring when comparing specificity for studies using only 1.5T scanners versus those that used other field strengths, having a pooled specificity of 93.9% (95 CI: 89.8–96.4) and those not 87.8% (95 CI: 48.9–98.2).

Next, to better quantify heterogeneity, we performed separate meta-regressions using each study characteristic as a covariate in the meta-analysis; we then calculated the $P$-value of the covariate’s effect on sensitivity and specificity. For example, in the case of the variable of “proportion of females enrolled,” the $P$-value for sensitivity was 0.475 and specificity was 0.141. The full results are displayed in Table 3. None of the characteristics had a significant effect on the sensitivity and specificity, although this could partially be attributed to the small samples sizes. Notably, the effect of magnetic field strength and proportion of women in each study on test specificity approached significance, although these effects were driven primarily by the results of Chabanova et al, which we had already determined were outlier data. Ultimately, it was determined that the best estimate would not adjust for sources of heterogeneity since they appeared minimal. Finally, in multivariate regression, which simultaneously assesses all observable potential sources of heterogeneity that were previously assessed individually, no $P$-values were significant.

**Discussion**

In this work we performed a meta-analysis of all studies using MRI to diagnose appendicitis since 2005 in the

<table>
<thead>
<tr>
<th>Table 2. Pooled Test Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>All studies</td>
</tr>
<tr>
<td>$n = 10$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Outlier removed</td>
</tr>
<tr>
<td>$n = 9$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
</tr>
<tr>
<td>0.97 (0.92–0.99)</td>
</tr>
<tr>
<td>0.97 (0.95–0.99)</td>
</tr>
<tr>
<td>Specificity</td>
</tr>
<tr>
<td>0.96 (0.89–0.98)</td>
</tr>
<tr>
<td>0.96 (0.92–0.98)</td>
</tr>
<tr>
<td>LR+</td>
</tr>
<tr>
<td>20 (8–49)</td>
</tr>
<tr>
<td>24 (12–47)</td>
</tr>
<tr>
<td>LR−</td>
</tr>
<tr>
<td>0.03 (0.02–0.07)</td>
</tr>
<tr>
<td>0.03 (0.01–0.06)</td>
</tr>
<tr>
<td>PPV</td>
</tr>
<tr>
<td>0.96 (0.92–0.99)</td>
</tr>
<tr>
<td>0.97 (0.94–0.98)</td>
</tr>
<tr>
<td>NPV</td>
</tr>
<tr>
<td>0.96 (0.91–0.98)</td>
</tr>
<tr>
<td>0.97 (0.93–0.98)</td>
</tr>
</tbody>
</table>

Since there was one outlier in the articles included in this analysis, results are presented both with and without the data from that outlier included. The results are reported as point estimates with 95% confidence intervals. LR+ = positive likelihood ratio, LR− = negative likelihood ratio, PPV = positive predictive value, NPV = negative predictive value.
The summary ROC curve demonstrated that MRI is highly accurate for the diagnosis of appendicitis, mimicking previous reports of the accuracy of CT. In the past decade, there has been increasing awareness of the potential harms associated with the use of ionizing radiation from CT, despite its very high accuracy for imaging acute pathology in the abdomen. Previous reports have shown a strong trend of increasing use of CT for the evaluation of patients presenting to the ED with abdominal pain, without a corresponding increase in the number of cases of surgical emergencies identified. While the reason for this increase has not been clearly identified, it does prompt the question of whether an alternative effective imaging modality that does not expose patients to ionizing radiation is available. Recently, MRI has emerged as a possible alternative, although an adequately powered, prospective study comparing MRI with a single reference imaging standard has not yet been reported. Meta-analyses can help to address this type of knowledge gap.

Prior to our analysis, there were only two published meta-analyses on this topic, both of which have limitations. Barger and Nandalur performed a meta-analysis of eight studies comprising 363 adult patients from 1995–2009, but conclusions from this study were limited by the relatively low number of patients included and types of scanners/sequences in use during that time period. In addition, the quality assessment tool and data reporting standards have changed significantly since that publication. The other meta-analysis by Blumenfeld et al addressed a somewhat different question: How well does MRI diagnose appendicitis in the pregnant population? The quality of data included in that meta-analysis was notably low, comprising five retrospective case series. Moreover, several important items were not present in that publication, including a quality assessment for the articles included in the analysis, a description of the methods for data abstraction, forest plots for individual and pooled test characteristics, and describing the degree of uncertainty around point estimates (eg, 95% CI).

Our results build upon the two previously published meta-analyses by including significantly more patients, incorporating the most up-to-date MRI protocols (7 of the 10 studies were published since the last meta-analysis was performed), as well as following Cochrane Review methodology. Our calculated pooled sensitivity and specificity are similar to that previously reported in several meta-analyses of CT to diagnose appendicitis. Importantly, although the prevalence of appendicitis in the included studies may seem higher than those encountered in general practice, our results are actually in line with what has been previously reported in meta-analyses of CT and ultrasound. Moreover, we explored potential sources of
utilizing a single imaging reference standard would further justify routine use of MRI to diagnose acute appendicitis.

**Acknowledgment**

Contract grant sponsor: National Center for Advancing Translational Sciences (NCATS); contract grant numbers: UL1TR000427; KL2TR000428; Contract grant sponsor: National Institute on Aging; contract grant number: K23AG038352; Contract grant sponsor: National Institute on Drug Abuse; contract grant number: K23DA032306; Contract grant sponsor: National Institute of Mental Health; contract grant number: T21MH118029; Contract grant sponsor: National Institute for Diabetes and Digestive and Kidney Diseases; contract grant number: K24DK102595

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**References**