



Published in final edited form as:

*J Magn Reson Imaging*. 2016 June ; 43(6): 1346–1354. doi:10.1002/jmri.25115.

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

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

<sup>1</sup>Department of Emergency Medicine, University of Wisconsin-Madison, Madison, WI

<sup>2</sup>Department of Radiology, University of Wisconsin-Madison, Madison, WI

<sup>3</sup>Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI

<sup>4</sup>Department of Medicine, University of Wisconsin-Madison, Madison, WI

<sup>5</sup>Department of Medical Physics, University of Wisconsin-Madison, Madison, WI

<sup>6</sup>Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI

### Abstract

**Purpose**—To perform a systematic review and meta-analysis of all published studies since 2005 that evaluate the accuracy of 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 multi-center studies. A total of 838 subjects were enrolled; 406 (48%) were women. All studies routinely used unenhanced MR images, though two used intravenous contrast-enhancement and three used diffusion-weighted imaging. Using a bivariate random-effects model the summary sensitivity was 96.6% (95% 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 CT.

### Keywords

MRI; Appendicitis; Meta-Analysis; Appendectomy; Diffusion Weighted Imaging

---

## Introduction

In the United States in 2005, 38.8 million patients were seen in emergency departments (EDs) for abdominal pain.<sup>1</sup> Appendicitis is a frequent cause of such visits, leading to 250,000 appendectomies performed annually.<sup>2</sup> Diagnosing appendicitis using only clinical findings is inaccurate in as many as 30% of cases, and may lead to unnecessary surgery.<sup>3</sup> Conversely, a missed diagnosis of appendicitis carries significant morbidity. Though 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.<sup>4,5</sup> 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.<sup>6</sup> Emergency physician performed ultrasound has been reported to have sensitivity 44–67%, specificity 85–98% and accuracy 67%, though this specifically relates to clinicians with limited training in sonography.<sup>7,8</sup> 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%.<sup>9</sup> Further, though the test accuracy of ultrasound in the case of pediatric abdominal pain concerning for 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 don't use the technology often.<sup>10</sup> Finally, nearly half of cases using ultrasound to diagnose appendicitis are either negative or non-diagnostic.<sup>11</sup> 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.<sup>12</sup> 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.<sup>13</sup> As a result, CT has become widely adopted as the primary imaging modality for detecting appendicitis in the United States, particularly for adults.<sup>14</sup> 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.<sup>15,16</sup> In a five-year period (1999–2004), CT use increased from 51 to 76% for those eventually diagnosed with appendicitis in one study<sup>17</sup> while another study found that the use of CT in patients presenting to the ED with abdominal pain doubled over a five-year period (2001–2005) to 22.5%.<sup>18</sup> Though 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).<sup>19</sup> 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 approximately 10 mSv, corresponding to an estimated excess risk of radiation-induced cancer of 1:2000.<sup>20</sup> Though “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, though still superior to ultrasound (sensitivity 82.5%, specificity 82%).<sup>21</sup> 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.<sup>22,23</sup>

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.<sup>24</sup>

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.<sup>25</sup> 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, multi-center 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 is to perform a systematic review and meta-analysis of the use of MRI to diagnose appendicitis in the general population, i.e. 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 and colleagues as well as the Cochrane Collaboration.<sup>26,27</sup> Given that our meta-analysis does not qualify as human subjects research, it was exempt from 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 while 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, though 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.<sup>28</sup> 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.<sup>29</sup> In addition, it has been suggested that this approach is preferred to the original bivariate model presented by Reitsma when cell counts are low.<sup>30</sup>

The bivariate generalized linear mixed model produces summary estimates of sensitivity and specificity with 95% confidence intervals. 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.<sup>26,31</sup> Data analysis was performed using the 'mada' package in R (R version 3.2.2) and RevMan 5.1.<sup>32</sup>

## 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 1.5 Tesla (T) MRI scanners while 3 studies used scanners with lower field strengths (1.0T, 0.5T) or a combination of scanners that 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 (Figure 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, ten specifically enrolled children and fourteen enrolled only pregnant women, and therefore were excluded from this analysis. Of the twelve remaining articles, three were reports derived from the same study population. Ultimately, ten articles met all inclusion criteria.

### Characteristics of Included Studies

A total of 838 patients were enrolled in the ten 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,<sup>33,34</sup> Turkey,<sup>35–37</sup> Denmark,<sup>38</sup> Germany,<sup>39</sup> Japan,<sup>40</sup> China,<sup>41</sup> and the United States.<sup>42</sup> All studies used clinical

follow-up as their reference standard for the purposes of calculating the test 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 (Figure 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.<sup>38</sup> Cobben and colleagues used a 1.0T Siemens scanner.<sup>34</sup> The study from Japan used a Philips 0.5T Gyroscan.<sup>40</sup> Almost all studies used unenhanced imaging protocols, though 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 Figure 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 ten studies' sensitivity and specificity on a 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 Chabanaova and colleagues (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 characteristic 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, though 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, though these effects were driven primarily by the results of Chabanova and colleagues, 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 multi-variate 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 have performed a meta-analysis of all studies using MRI to diagnose appendicitis since 2005 in the general population. Of the 10 studies that met eligibility criteria, the summary ROC curve demonstrated that MRI is highly accurate for the diagnosis of appendicitis, mimicking previous reports of the accuracy of CT.<sup>12</sup>

In the past decade, there has been increasing awareness of the potential harms associated with 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.<sup>18</sup> 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, though 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 of this study were limited by the relatively low number of patients included and types of scanners/sequences in use during that time period.<sup>43</sup> In addition, the quality assessment tool and data reporting standards have changed significantly since that publication.<sup>26,28</sup> The other meta-analysis by Blumenfeld and colleagues addressed a somewhat different question: how well does MRI diagnose appendicitis in the pregnant population?<sup>44</sup> 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 (e.g. – 95% confidence intervals).

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.<sup>14,45,46</sup> Importantly, though 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.<sup>14,46</sup> Moreover, we explored potential sources of heterogeneity and found that of the observable characteristics in the included studies, none had a significant impact on the overall sensitivity and specificity, though it is possible other non-observed characteristics might contribute to study differences.

These findings strongly suggest that MRI is a reasonable alternative to CT for the diagnosis of appendicitis in hospitals with appropriate access to this technology. In particular, use of MRI instead of CT would avoid exposing patients to ionizing radiation, which may increase a patient's lifetime risk of developing cancer, particularly in younger patients. While MRI is not suited for patients with certain contraindications (e.g. metallic implants, claustrophobia, etc), it may be well-suited for other patient populations including those at risk for contrast-induced nephropathy from iodinated contrast material, those who have a history of iodinated contrast reactions, or those at risk for radiation-induced malignancies, especially young patients.

There are a number of limitations with this meta-analysis. First, the included studies reported a relatively high prevalence of appendicitis when compared with the general population presenting to the ED with acute abdominal pain concerning for appendicitis. This is likely due to some of the studies using scheduled appendectomy as an eligibility criterion, i.e. patients with a very high pre-test probability of having appendicitis. This could lead to spectrum bias and limits the external validity of the summary test characteristics. In addition, recruitment using convenience sampling for most of the studies and clinical follow-up to ascertain possible false negative results was not always well described. This could lead to both spectrum and information bias. Further, the MRI protocols used at the different sites were not uniform. In particular, most studies used unenhanced imaging sequences, although some allowed for contrast-enhancement if the initial images were non-diagnostic, and others utilized diffusion-weighted imaging.

Regardless, the combined data are compelling. With the exception of one study that showed a specificity of 61%,<sup>38</sup> the remaining articles included for this analysis were very consistent. Future study in this field should focus on enrolling a more clinically relevant cohort (i.e. – moderate pre-test probability of disease) to reflect the population where imaging is particularly helpful. In addition, using a single reference standard, such as CT, would greatly assist in directly comparing the test performance of each imaging modality as well as clearly elucidating the inter-test reliability of MRI and CT.



## Conclusion

Magnetic resonance imaging is an effective alternative to CT for the diagnosis of acute appendicitis with pooled sensitivity and specificity similar to CT. However, limitations regarding subject recruitment and reference standards limit the strength of this conclusion. A large, prospective study utilizing a single imaging reference standard would further justify routine use of MRI to diagnose acute appendicitis.

## Acknowledgements

The project described was supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427 and KL2TR000428. Additional support was provided by the National Institute on Aging, grant K23AG038352, the National Institute on Drug Abuse, grant K23DA032306, the National Institute of Mental Health, grant T21MH18029, and the National Institute for Diabetes and Digestive and Kidney Diseases, grant K24DK102595. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## References

1. Nawar EW, Niska RW, Xu J. National Hospital Ambulatory Medical Care Survey: 2005 emergency department summary. *Adv Data*. 2007; (386):1–32. [PubMed: 17703794]
2. Mason RJ. Surgery for appendicitis: is it necessary? *Surg Infect*. 2008; 9(4):481–488.
3. Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology*. 2000; 215(2):337–348. [PubMed: 10796905]
4. Sun JS, Noh HW, Min YG, et al. Receiver operating characteristic analysis of the diagnostic performance of a computed tomographic examination and the Alvarado score for diagnosing acute appendicitis: emphasis on age and sex of the patients. *J Comput Assist Tomogr*. 2008; 32(3):386–391. [PubMed: 18520542]
5. Mán E, Simonka Z, Varga A, Ráosi F, Lázár G. Impact of the Alvarado score on the diagnosis of acute appendicitis: comparing clinical judgment, Alvarado score, and a new modified score in suspected appendicitis: a prospective, randomized clinical trial. *Surg Endosc*. 2014; 28(8):2398–2405. [PubMed: 24705731]
6. Pickuth D, Heywang-Köbrunner SH, Spielmann RP. Suspected acute appendicitis: is ultrasonography or computed tomography the preferred imaging technique? *Eur J Surg Acta Chir*. 2000; 166(4):315–319.
7. Fathi M, Hasani SA, Zare MA, Daadpey M, Hojati Firoozabadi N, Lotfi D. Diagnostic accuracy of emergency physician performed graded compression ultrasound study in acute appendicitis: a prospective study. *J Ultrasound*. 2015; 18(1):57–62. [PubMed: 25767641]
8. Mallin M, Craven P, Ockerse P, et al. Diagnosis of appendicitis by bedside ultrasound in the ED. *Am J Emerg Med*. 2015; 33(3):430–432. [PubMed: 25559314]
9. D'Souza N, D'Souza C, Grant D, Royston E, Farouk M. The value of ultrasonography in the diagnosis of appendicitis. *Int J Surg Lond Engl*. 2015; 13:165–169.
10. Mittal MK, Dayan PS, Macias CG, et al. Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2013; 20(7):697–702.
11. Atema JJ, Gans SL, Van Randen A, et al. Comparison of Imaging Strategies with Conditional versus Immediate Contrast-Enhanced Computed Tomography in Patients with Clinical Suspicion of Acute Appendicitis. *Eur Radiol*. 2015; 25(8):2445–2452. [PubMed: 25903701]
12. van Randen A, Laméris W, van Es HW, et al. A comparison of the accuracy of ultrasound and computed tomography in common diagnoses causing acute abdominal pain. *Eur Radiol*. 2011; 21(7):1535–1545. [PubMed: 21365197]
13. 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. [PubMed: 9428814]

14. 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. [PubMed: 18682583]
15. Brenner DJ, Hall EJ. Cancer risks from CT scans: now we have data, what next? *Radiology*. 2012; 265(2):330–331. [PubMed: 22915598]
16. Berrington de González A, Mahesh M, Kim K-P, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med*. 2009; 169(22):2071–2077. [PubMed: 20008689]
17. Otero HJ, Ondategui-Parra S, Erturk SM, Ochoa RE, Gonzalez-Beicos A, Ros PR. Imaging utilization in the management of appendicitis and its impact on hospital charges. *Emerg Radiol*. 2008; 15(1):23–28. [PubMed: 17972120]
18. Pines JM. Trends in the rates of radiography use and important diagnoses in emergency department patients with abdominal pain. *Med Care*. 2009; 47(7):782–786. [PubMed: 19536032]
19. Kotagal M, Richards MK, Chapman T, et al. Improving ultrasound quality to reduce computed tomography use in pediatric appendicitis: the Safe and Sound campaign. *Am J Surg*. 2015; 209(5):896–900. discussion 900. [PubMed: 25771132]
20. Dixon AK, Dendy P. Spiral CT: how much does radiation dose matter? *Lancet*. 1998; 352(9134):1082–1083. [PubMed: 9798579]
21. Karabulut N, Kiroglu Y, Herek D, Kocak TB, Erdur B. Feasibility of low-dose unenhanced multi-detector CT in patients with suspected acute appendicitis: comparison with sonography. *Clin Imaging*. 2014; 38(3):296–301. [PubMed: 24513314]
22. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med*. 2012; 366(17):1596–1605. [PubMed: 22533576]
23. Soyer P, Dohan A, Eveno C, et al. Pitfalls and mimickers at 64-section helical CT that cause negative appendectomy: an analysis from 1057 appendectomies. *Clin Imaging*. 2013; 37(5):895–901. [PubMed: 23845254]
24. Ginde AA, Foianini A, Renner DM, Valley M, Camargo CA Jr. Availability and quality of computed tomography and magnetic resonance imaging equipment in U.S. emergency departments. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2008; 15(8):780–783.
25. Ramalingam V, LeBedis C, Kelly JR, Uyeda J, Soto JA, Anderson SW. Evaluation of a sequential multi-modality imaging algorithm for the diagnosis of acute appendicitis in the pregnant female. *Emerg Radiol*. 2015; 22(2):125–132. [PubMed: 25148766]
26. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM. Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008; 149(12):889–897. [PubMed: 19075208]
27. Deeks, JJ.; Higgins, JP.; Altman, DG. *Cochrane Handbook for Systematic Reviews of Interventions*. 2011. Chapter 9: Analysing data and undertaking meta-analyses. 5.1.0 ed. <http://handbook.cochrane.org/>
28. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011; 155(8):529–536. [PubMed: 22007046]
29. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol*. 2006; 59(12):1331–1332. author reply 1332–1333. [PubMed: 17098577]
30. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005; 58(10):982–990. [PubMed: 16168343]
31. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostat Oxf Engl*. 2007; 8(2):239–251.
32. Doebler, P. [Accessed September 4, 2015] Mada: Meta-Analysis of Diagnostic Accuracy. 2015. <https://cran.r-project.org/web/packages/mada/index.html>.
33. Leeuwenburgh MMN, 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. [PubMed: 23481162]
34. 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. [PubMed: 19137303]
  35. 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 Ank Turk*. 2013; 19(2):106–110.
  36. İnci E, Hocaoglu E, Aydın S, et al. Efficiency of unenhanced MRI in the diagnosis of acute appendicitis: Comparison with Alvarado scoring system and histopathological results. *Eur J Radiol*. 2011; 80(2):253–258. [PubMed: 20655156]
  37. İnci E, Kilickesmez O, Hocaoglu E, Aydın S, Bayramoglu S, Cimilli T. Utility of diffusion-weighted imaging in the diagnosis of acute appendicitis. *Eur Radiol*. 2011; 21(4):768–775. [PubMed: 20924585]
  38. Chabanova E, Balslev I, Achiam M, et al. Unenhanced MR Imaging in adults with clinically suspected acute appendicitis. *Eur J Radiol*. 2011; 79(2):206–210. [PubMed: 20347539]
  39. Heverhagen JT, Pfestroff K, Heverhagen AE, Klose KJ, Kessler K, Sitter H. Diagnostic accuracy of magnetic resonance imaging: a prospective evaluation of patients with suspected appendicitis (diamond). *J Magn Reson Imaging JMRI*. 2012; 35(3):617–623. [PubMed: 22033948]
  40. 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 JMRI*. 2005; 21(2):156–165. [PubMed: 15666398]
  41. 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. [PubMed: 23289088]
  42. 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. [PubMed: 18649091]
  43. Barger RL Jr, Nandalur KR. Diagnostic performance of magnetic resonance imaging in the detection of appendicitis in adults: a meta-analysis. *Acad Radiol*. 2010; 17(10):1211–1216. [PubMed: 20634107]
  44. Blumenfeld YJ, Wong AE, Jafari A, Barth RA, El-Sayed YY. MR imaging in cases of antenatal suspected appendicitis--a meta-analysis. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet*. 2011; 24(3):485–488.
  45. 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. [PubMed: 17268692]
  46. Hlibczuk V, Dattaro JA, Jin Z, Falzon L, Brown MD. Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review. *Ann Emerg Med*. 2010; 55(1):51–59. e1. [PubMed: 19733421]

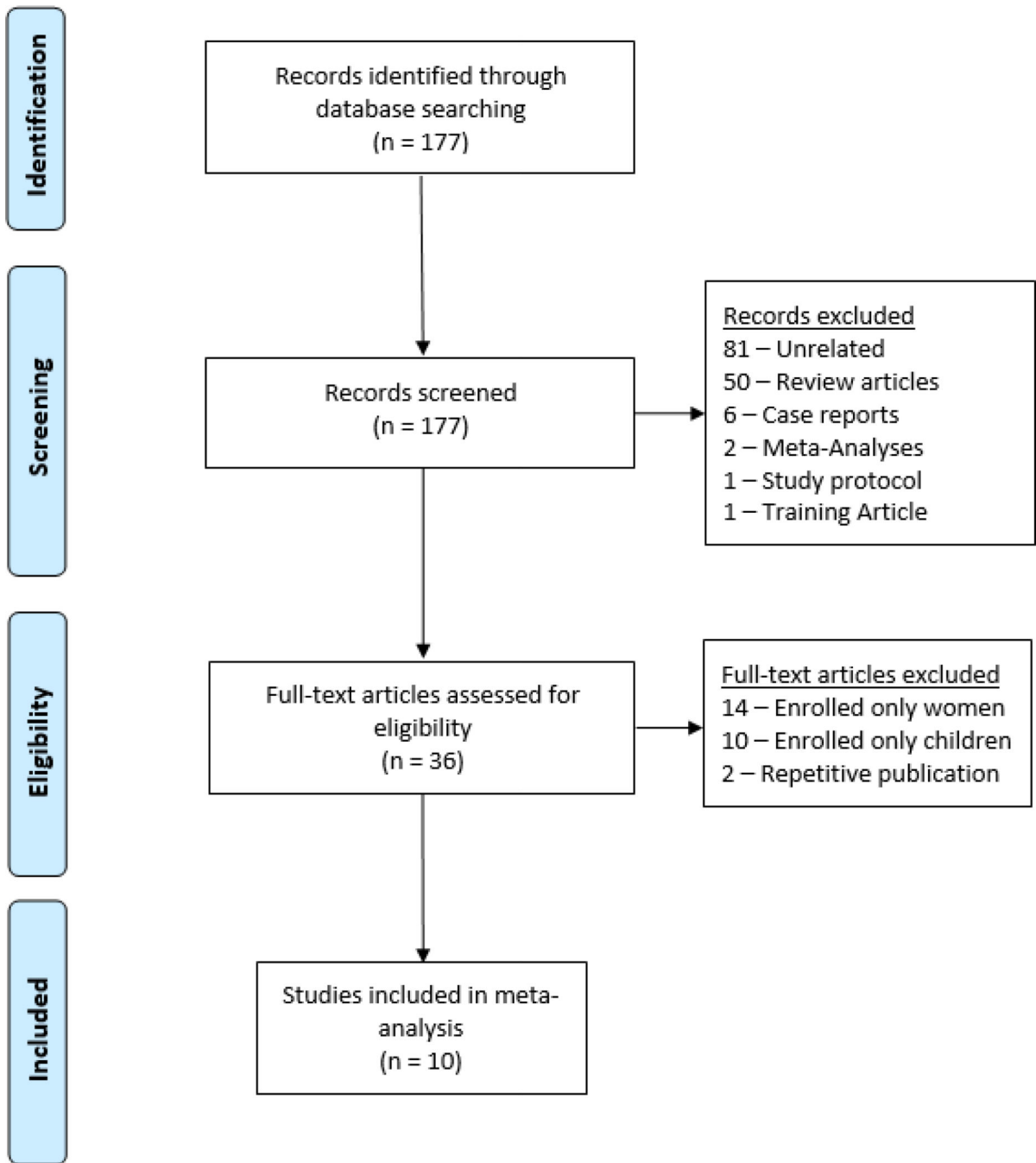


Figure 1. Diagram of article selection

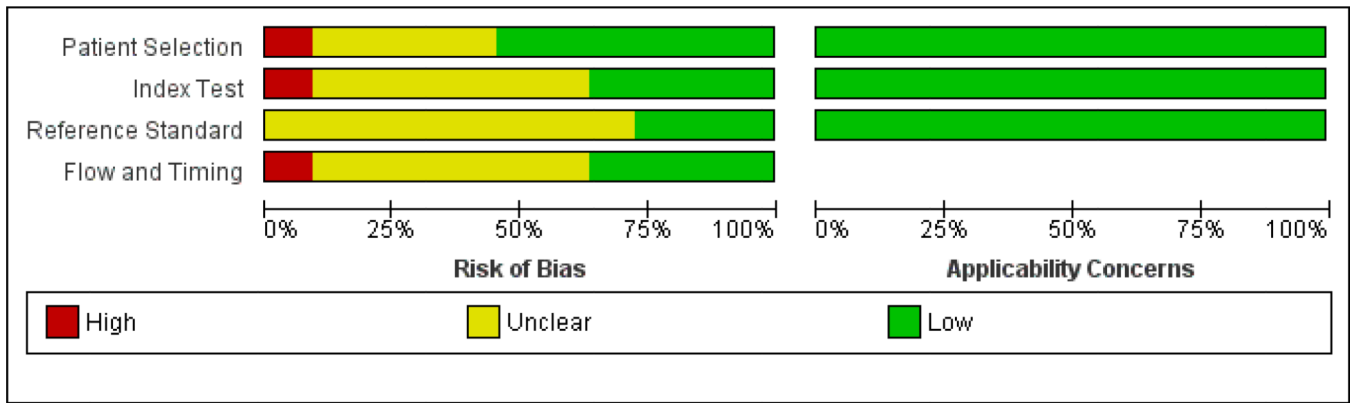
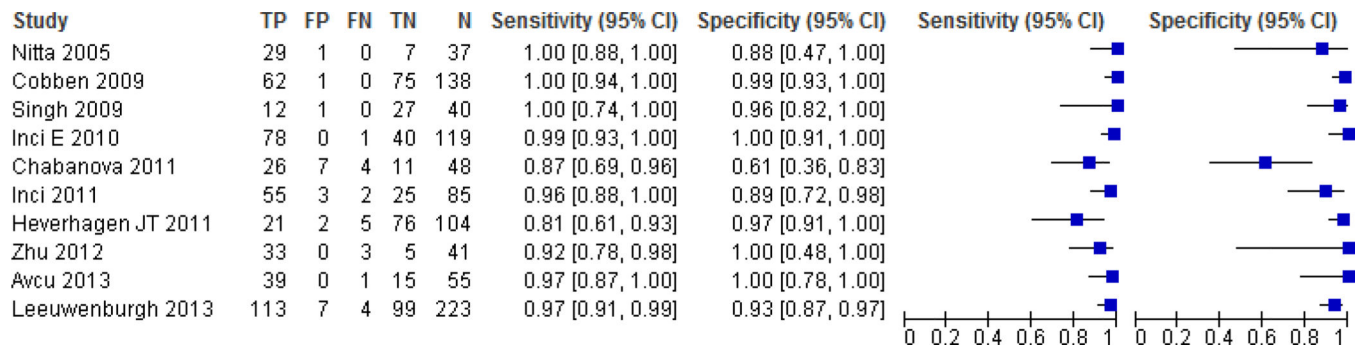
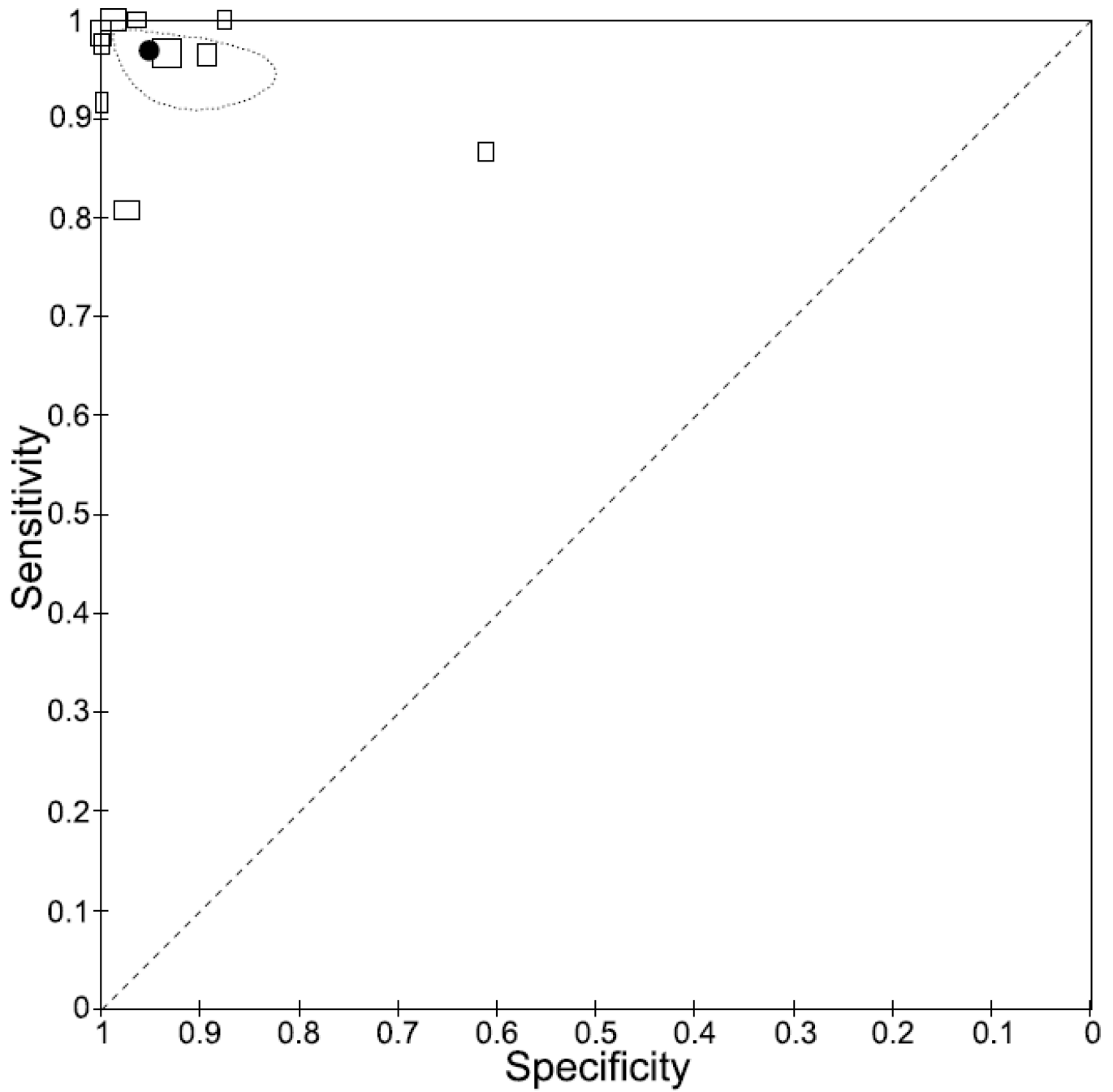


Figure 2. QUADAS-2 Assessment findings



**Figure 3. Sensitivity and specificity of included studies**

The first column includes the last name of the first author for each of the included studies as well as the year of publication, listed in chronological order. The next five columns report the number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), and total number of patients (N) for each of the studies. Sensitivity and specificity are depicted numerically and then graphically as forest plots.



**Figure 4. Summary receiver operating curve plot**  
 The solid circle represents the summary point estimate for sensitivity and specificity; the ellipse shows the 95% confidence interval for this estimate. Boxes depict the sensitivity and specificity of individual studies included in this analysis. The size of the box is proportionate to the number of patients enrolled for each study.

**Table 1**

**List of articles meeting inclusion criteria**

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.

First author	Year	Location	Study Type	Study Dates	Reference Standard	Total Patients	Prevalence	Women (%)	Mean (range) age (yrs)	Scanner Type (Manufacturer)	MRI Sequences	Reviewer	Time (min)
Nitta	2005	Japan	Prospective, single-center	Unknown	Surgical pathology or clinical follow up	37	78%	19 (52)	37.1 (16-69)	0.5T Gyroscan (Philips)	T1 SE, T2 FSE, T2 with fat saturation	Three experienced radiologists, by consensus	Unknown
Cobben	2009	Netherlands	Prospective, single-center	1/05-10/06	Surgical pathology or clinical follow up	138	45%	80 (56)	29 (6-80)	1.0T (Siemens)	T1 SGRE, T2 FSE, T2 FSE with fat saturation	One of three GI radiologists (>5 years experience)	20
Singh	2009	USA	Retrospective, multi-center	2001-2007	Final diagnosis at hospital discharge	40	30%	Unknown	34 (11-69)	1.5T Excite Twinspeed (GE Medical Systems)	T2 SSFSE with fat saturation, T2 FSE with fat saturation, STIR, pre-gadolinium T1, post-gadolinium T1	Two experienced radiologists (5 & 10 years). A third radiologist (>20 yrs experience) if lack of consensus	Unknown
Inci	2011	Turkey	Prospective, single-center	7-month period	Surgical pathology or clinical follow up	85	67%	40 (47)	26.5 ± 11.3 (14-72)	1.5 T Avanto (Siemens)	T1 FSE, T2 FSE with and without fat saturation	Two abdominal radiologists (8 years and 5 years experience)	Unknown
Chabanova	2011	Denmark	Prospective, single-center	Unknown	Surgical pathology or clinical follow up	48	63%	29 (60)	37.1 (18-70)	0.23T and 0.6T Panorama (Philips), 1.5T Infinion (Philips), 1.5T Achieva (Philips)	T1 SE, T2 FSE, STIR	One radiologist (10 years of experience, 6 years experience in abdominal MRI) for	20



First author	Year	Location	Study Type	Study Dates	Reference Standard	Total Patients	Prevalence	Women (%)	Mean (range) age (yrs)	Scanner Type (Manufacturer)	MRI Sequences	Reviewer	Time (min)
<b>Inci</b>	2011	Turkey	Prospective, single-center	11-month period	Surgical pathology or clinical follow up	119	66%	36	27 (17–72)	1.5 T Avanto (Siemens)	T1 FSE, T2 FSE with and without fat saturation, DWI	primary outcome, also a surgeon (3 years of experience in abdominal MRI), and a research fellow (1 year experience in abdominal MRI)	Unknown
<b>Heverhagen</b>	2012	Germany	Prospective, single-center	02/2008 – 10/2008	Surgical pathology or clinical follow up	52	25%	21 (40)	44.7 (18–88)	1.5 T Magnetom Sonata (Siemens)	STIR, T2 FSE, bSSFP, T1 fat-saturated SGRE (before and after IV contrast). IV butylscopolamin used to prevent peristalsis.	One radiologist (11 years of experience in MRI). A second radiologist with 4 years of experience reviewed cases for inter-rater variability.	14
<b>Zhu</b>	2012	China	Prospective, single-center	9/09–9/11	Surgical pathology	41	80%	23 (56)	41.5 ± 11.3	1.5T Achieva Nova Dual (Philips)	T2 FSE, bSSFP with fat saturation	One GI radiologist with >10 years experience.	2 min 42 sec

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

First author	Year	Location	Study Type	Study Dates	Reference Standard	Total Patients	Prevalence	Women (%)	Mean (range) age (yrs)	Scanner Type (Manufacturer)	MRI Sequences	Reviewer	Time (min)
Leeuwenburgh	2013	Netherlands	Prospective, multi-center	3/2010 – 9/2010	Expert panel reviewing surgery and clinical follow up	223	52%	132 (59)	35 (IQR 25–50)	1.5T Magnetom Avanto (Siemens)	T2 FSE with and without fat saturation, DWI	Two experienced radiologists (14 and 16 years experience) in consensus	15
Avcu	2013	Turkey	Prospective, single-center	03/2009 – 02/2010	Surgical pathology or clinical follow up	55	71%	26 (43)	35.6 ± 15.5 (17–83)	1.5T Magnetom Symphony (Siemens)	DWI, bSSFP, STIR	One radiologist (experience level not given)	Unknown

**Table 2****Pooled test characteristics**

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.

	All studies (n=10)	Outlier removed (n=9)
<b>Sensitivity</b>	0.97 (0.92–0.99)	0.97 (0.95–0.99)
<b>Specificity</b>	0.96 (0.89–0.98)	0.96 (0.92–0.98)
<b>LR+</b>	20 (8–49)	24 (12–47)
<b>LR-</b>	0.03 (0.02–0.07)	0.03 (0.01–0.06)
<b>PPV</b>	0.96 (0.92–0.99)	0.97 (0.94–0.98)
<b>NPV</b>	0.96 (0.91–0.98)	0.97 (0.93–0.98)

**Table 3**

**Investigation of heterogeneity**

These are the summary sensitivity and specificity with 95% confidence intervals for each covariate evaluated in this meta-regression. Subgroups were defined by whether the reported covariate was above or below the median value for all included studies. The only exception was magnetic field strength, which was dichotomized into studies that used only 1.5T MRI scanners and those that did not. The reported p-values indicate the effect of each covariate on the summary sensitivity and specificity when evaluated by pooled meta-regression.

	All	Percent Female		Median Age		Median Prevalence		Magnetic Field Strength (Tesla)					
		<51%	51%	p-value	<35.6	35.6	p-value	<64.5%	64.5%	p-value	<1.5T	1.5T	p-value
N	10	5	5		5	5		5	5		3	7	
Sensitivity	96.6 [92.3,98.5]	96.5 [89.0,98.9]	96.6 [89.6,98.9]	0.475	96.1 [89.8,98.6]	97.0 [89.1,99.2]	0.303	97.5 [91.3,99.4]	95.9 [87.9,98.7]	0.319	96.6 [77.7,99.6]	94.0 [88.5,97.0]	0.731
Specificity	95.9 [89.4,98.4]	97.1 [91.2,99.0]	92.2 [76.3,97.8]	0.141	95.2 [91.8,97.3]	94.1 [74.4,98.8]	0.422	95.1 [81.4,98.8]	96.1 [89.5,98.6]	0.291	87.8 [48.9,98.2]	93.9 [89.8,96.4]	0.209