Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients with Suspected Appendicitis

Approved by ACEP Board of Directors February 1, 2023

From the American College of Emergency Physicians Clinical Policies Subcommittee (Writing Committee) on Appendicitis

Deborah B. Diercks, MD, MSc (Subcommittee Chair, Committee Chair)
Eric J. Adkins, MD
Nicholas Harrison, MD
Peter E. Sokolove, MD
Heemun Kwok, MD, MS (Methodologist)
Stephen J. Wolf, MD (Chair 2018 - 2021)

Members of the American College of Emergency Physicians Clinical Policies Committee (Oversight Committee):
Deborah B. Diercks, MD, MSc (Co-Chair 2021-2022, Chair 2022-2023)
John D. Anderson, MD
Richard Byyny, MD, MSc (Methodologist)
Christopher R. Carpenter, MD
Benjamin Friedman, MD (Methodologists)
Seth R. Gemme, MD
Charles J. Gerardo, MD, MHS
Steven A. Godwin, MD
Sigrid A. Hahn, MD, MPH
Benjamin W. Hatten, MD, MPH
Jason S. Haukoos, MD, MSc (Methodologist)
Amy Kaji, MD, MPH, PhD (Methodologist)
Heemun Kwok, MD, MS (Methodologist)
Bruce M. Lo, MD, MBA, RDMS
Sharon E. Mace, MD

Maggie Moran, MD (EMRA Representative 2022-2023)
Susan B. Promes, MD, MBA
Kaushal H. Shah, MD
Richard D. Shih, MD
Scott M. Silvers, MD
Andrea Slivinski, RN, DNP (ENA Representative 2021-2023)
Michael D. Smith, MD, MBA
Molly E. W. Thiessen, MD
Christian A. Tomaszewski, MD, MS, MBA
Stacy Trent, MD MPH (Methodologist)
Jonathan H. Valente, MD
Stephen P. Wall, MD, MSc, MAEd (Methodologist)
Lauren M. Westafer, DO
Yanling Yu, PhD (Washington Advocates for Patient Safety)
Stephen V. Cantrill, MD (Liaison with Quality and Patient Safety Committee)
John T. Finnell, MD (Board Liaison 2020-2023)
Travis Schulz, MLS, AHIP, Staff Liaison, Clinical Policies Committee
Kaeli Vandertulip, MBA, MSLS, AHIP, Staff Liaison, Clinical Policies Committee and Subcommittee on Appendicitis

<BEGIN ABSTRACT>

ABSTRACT

This clinical policy from the American College of Emergency Physicians is a revision of the 2010 “Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients with Suspected Appendicitis.” A writing subcommittee conducted a systematic review of the literature to derive evidence-based recommendations to answer the following clinical questions: 1) in ED patients with possible acute appendicitis, can a clinical prediction rule be used to identify patients for whom no advanced imaging is required? 2) in ED patients with suspected acute appendicitis, is the diagnostic accuracy of ultrasound comparable with computed tomography or magnetic resonance imaging for the diagnosis of acute appendicitis? 3) In ED patients who are undergoing computed tomography of the abdomen and pelvis for suspected acute appendicitis, does the addition of contrast improve diagnostic accuracy? Evidence was graded, and recommendations were made based on the strength of the available data.

<END ABSTRACT>

<BEGIN ARTICLE>

<H1>**INTRODUCTION**

Abdominal pain is a high-volume, high-risk chief complaint. In 2016, patients with abdominal pain composed 8.6% of emergency department visits. Almost 200,000 patients are diagnosed with appendicitis each year.[1] Missed diagnosis of appendicitis remains an area at high risk of litigation.[2] Among children, appendicitis is the fifth most common cause of malpractice litigation against emergency physicians.[3] The diagnosis of appendicitis can be challenging even in the most experienced of clinical hands.

Despite the increasing use of computed tomography (CT) in patients with possible appendicitis, such widespread use may be unnecessary. Traditional teaching suggests that clinical indicators (eg, signs, symptoms, and laboratory tests) exist that may be used to identify patients with acute appendicitis. It has been suggested that such indicators may be used to facilitate the early identification of ED patients who have acute appendicitis. Of particular interest to the emergency physician is the identification of patients who are so unlikely to have appendicitis that they do not warrant imaging to confirm the diagnosis. Similarly, patients with high clinical suspicion of appendicitis may be referred to a surgeon with minimal or no testing.[4]

Once the decision is made to use imaging, performing a CT may or may not involve the use of contrast. If contrast is used, does it increase diagnostic performance in a clinically meaningful way? In children, some clinicians use ultrasound before or in lieu of CT to diagnose appendicitis. Although ultrasound does not involve ionizing radiation or the risks associated with contrast, the accuracy of either a positive or negative ultrasound result merits discussion. More recently, magnetic resonance imaging (MRI) has been suggested as an alternative imaging modality in patients with suspected appendicitis because it also does not involve ionizing radiation. Understanding the differences in diagnostic accuracy of ultrasound, CT, and MRI can inform decisions about choosing the imaging modality.

This policy is an update of the 2010 American College of Emergency Physicians (ACEP) "Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients with Suspected Appendicitis."[5] All the previous critical questions from the 2010 policy were updated in this version with some expansion with different comparators. The prior questions were the following: (1) can clinical findings be used to guide decisionmaking in the risk stratification of patients with possible appendicitis? (2) in adult patients with suspected acute appendicitis who are undergoing a CT scan, what is the role of contrast? (3) in children with

suspected acute appendicitis who undergo diagnostic imaging, what are the roles of CT and ultrasound in diagnosing acute appendicitis?

<H1>**METHODOLOGY**

This ACEP clinical policy was developed by emergency physicians with input from medical librarians and a patient safety advocate. It is based on a systematic review and critical, descriptive analysis of the medical literature and is reported in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[6]

<H2>Search and Study Selection

This clinical policy is based on a systematic review with a critical analysis of the medical literature meeting the inclusion criteria. Searches of PubMed, SCOPUS, Embase, Web of Science, and the Cochrane Database of Systematic Reviews were performed by a librarian. Search terms and strategies were peer reviewed by a second librarian. All searches were limited to human studies published in English. Specific key words/phrases, years used in the searches, dates of searches, and study selection are identified under each critical question. In addition, relevant articles from the bibliographies of included studies and more recent articles identified by committee members and reviewers were included.

Two subcommittee members independently read the identified abstracts to assess them for possible inclusion. Of those identified for potential inclusion, each full-length text was reviewed for eligibility. Those identified as eligible were subsequently forwarded to the committee's methodology group (emergency physicians with specific research methodological expertise) for methodological grading using a Class of Evidence framework (Appendix E1, available at http://www.annemergmed.com.).

<H2>Assessment of Risk of Bias and Determination of Classes of Evidence

Each study identified as eligible by the subcommittee was independently graded by 2 methodologists. Grading was done with respect to the specific critical questions; thus, the Class of Evidence for any one study may vary according to the question for which it is being considered. For example, an article that is graded an "X" because of "inapplicability" for one

critical question may be considered perfectly relevant for another question and graded I to III. As such, it was possible for a single article to receive a different Class of Evidence grade when addressing a different critical question.

Design 1 represents the strongest possible study design to answer the critical question, which relates to whether the focus was therapeutic, diagnostic, prognostic, or meta-analysis. Subsequent design types (ie, design 2 and design 3) represent weaker study designs, respectively. Articles are then graded on dimensions related to the study's methodological features and execution, including but not limited to randomization processes, blinding, allocation concealment, methods of data collection, outcome measures and their assessment, selection and misclassification biases, sample size, generalizability, data management, analyses, congruence of results and conclusions, and potential for conflicts of interest.

Using a predetermined process that combines the study's design, methodological quality, and applicability to the critical question, 2 methodologists independently assigned a preliminary Class of Evidence grade for each article. Articles with concordant grades from both methodologists received that grade as their final grade. Any discordance in the preliminary grades was adjudicated through discussion, which involved at least 1 additional methodologist, resulting in a final Class of Evidence assignment (ie, class I, class II, class III, or class X) (Appendix E2, available at http://www.annemergmed.com). Studies identified with significant methodologic limitations and/or ultimately determined to not be applicable to the critical question received a Class of Evidence grade "X" and were not used in formulating recommendations for this policy. However, the content in these articles may have been used to formulate the background and to inform expert consensus in the absence of evidence. Question-specific Classes of Evidence grading may be found in the Evidentiary Table included at the end of this policy.

<H2>Translation of Classes of Evidence to Recommendation Levels

Based on the strength of evidence for each critical question, the subcommittee drafted the recommendations and supporting text, synthesizing the evidence using the following guidelines:

Level A recommendations. Generally accepted principles for patient care that reflect a high degree of scientific certainty (eg, based on evidence from 1 or more Class of Evidence I, or multiple Class of Evidence II studies that demonstrate consistent effects or estimates).

Level B recommendations. Recommendations for patient care that may identify a particular strategy or range of strategies that reflect moderate scientific certainty (eg, based on evidence from 1 or more Class of Evidence II studies or multiple Class of Evidence III studies that demonstrate consistent effects or estimates).

Level C recommendations. Recommendations for patient care that are based on evidence from Class of Evidence III studies or, in the absence of adequate published literature, based on expert consensus. In instances where consensus recommendations are made, "consensus" is placed in parentheses at the end of the recommendation.

There are certain circumstances in which the recommendations stemming from a body of evidence should not be rated as highly as the individual studies on which they are based. Factors such as consistency of results, the uncertainty of effect magnitude, and publication bias, among others, might lead to a downgrading of recommendations. When possible, clinically oriented statistics (eg, likelihood ratios [LRs], number needed to treat) are presented to help the reader better understand how the results may be applied to the individual patient. This can assist the clinician in applying the recommendations to most patients but allow adjustment when applying to patients with extremes of risk (Appendix E3, `available at http://www.annemergmed.com`.).

<H2>Evaluation and Review of Recommendations

Once drafted, the policy was distributed for internal review (by members of the entire committee), followed by an external expert review and an open comment period for all ACEP membership. Comments were received during a 60-day open comment period, with notices of the comment period sent electronically to ACEP members, published in *EM Today*, posted on the ACEP website, and sent to other pertinent physician organizations. The responses were used to further refine and enhance this clinical policy, although responses did not imply endorsement. Clinical policies are scheduled for revision every 3 years; however, interim reviews are conducted when technology, methodology, or the practice environment changes significantly.

<H2>Application of the Policy

This policy is not intended to be a complete manual on the evaluation and management of patients with suspected appendicitis but rather a focused examination of critical questions that

have particular relevance to the current practice of emergency medicine. The potential benefits and harms of implementing recommendations are briefly summarized within each critical question.

It is the goal of the Clinical Policies Committee to provide evidence-based recommendations when the scientific literature provides sufficient quality information to inform recommendations for a critical question. When the medical literature does not contain adequate empirical data to inform a critical question, the members of the Clinical Policies Committee believe that it is equally important to alert emergency physicians to this fact.

This clinical policy is not intended to represent a legal standard of care for emergency physicians. Recommendations offered in this policy are not intended to represent the only diagnostic or management options available to the emergency physician. ACEP recognizes the importance of the individual physician's judgment and patient preferences. This guideline provides clinical strategies based on medical literature to inform the critical questions addressed in this policy. ACEP funded this clinical policy.

Scope of Application. This guideline is intended for physicians working in hospital-based EDs.

Inclusion Criteria. This guideline is intended for patients presenting to the ED with acute, nontraumatic abdominal pain and possible or suspected appendicitis.

Exclusion Criteria. This guideline is not intended to address the care of patients with trauma-related abdominal pain or patients who are pregnant.

<H1>**CRITICAL QUESTIONS**

1. In ED patients with possible acute appendicitis, can a clinical prediction rule be used to identify patients for whom no advanced imaging is required?

Patient Management Recommendations

Level A recommendations. None specified.

Level B recommendations. In pediatric patients, clinical prediction rules can be used to risk stratify for possible acute appendicitis. However, do not use clinical prediction rules alone to identify patients who do not warrant advanced imaging for the diagnosis of appendicitis.

Level C recommendations. In adult patients, because of insufficient data, do not use clinical prediction rules to identify patients for whom no advanced imaging is required.

<H2>Potential Benefit of Implementing the Recommendations:

- Reduction of CT imaging, radiation exposure, cost, and ED length of stay

<H2>Potential Harm of Implementing the Recommendations:

- Possible missed diagnosis of appendicitis in a patient presenting with low-risk symptoms, atypical presentations, or early in the disease course.

<H3>Key words/phrases for literature searches: appendicitis, ruptured appendicitis, perforated appendicitis, clinical decision support systems, clinical decision rules, clinical prediction rules, clinical prediction tools, computer assisted tomography, x-ray computed tomography, CT scans, ultrasonic tomography, medical imaging, ultrasonography, diagnostic ultrasound, ultrasound imaging, ultrasonic imaging, ultrasonic diagnosis, ultrasonographic imaging, sonography, medical sonography, diagnostic imaging, echography, computer echotomography, emergency, emergency health service, hospital emergency service, emergency ward, emergency medicine, emergency care, emergency treatment, emergency department, emergency room, emergency service, emergency services, and variations and combinations of the key words/phrases. Searches included January 2009 to search dates of May 10 to 11, 2020.

<H2>Study Selection: One hundred twenty-three articles were identified in searches. Twenty-one articles were selected from the search results as potentially addressing this question and were candidates for further review. After grading for methodological rigor, 6 articles were selected from the search results for further review, with 0 class I studies, 0 class II studies, and 6 class III studies included for this critical question (Appendix E4, available at http://www.annemergmed.com.).

The ability to accurately identify or exclude acute appendicitis using a clinical prediction rule without advanced imaging represents one of the holy grails in emergency medicine. After a review of the initial set of 21 articles, only 6 met the criteria for inclusion. All 6 articles were

level III evidence. Gonzalez del Castillo et al[7] compared a prospective observational cohort of younger patients aged 2 to 20 years using the APPY1 test to risk stratify the patients. The APPY1 test evaluates for C-reactive protein and calprotectin levels that get combined with a white blood cell count result. Patients were also broken out using the Alvarado score into low, intermediate, or high-risk cohorts as part of secondary data analysis. An Alvarado score of more than 4 had sensitivity 0.92 (95% CI, 0.85 to 0.96), specificity 0.45 (95% CI, 0.38 to 0.52), positive LR 1.7 (95% CI, 1.5 to 1.9), and negative LR 0.2 (95% CI, 0.1 to 0.3) for the diagnosis of appendicitis. Saucier et al[8] evaluated the pediatric appendicitis score (PAS) in patients 136 patients aged 3 to 17 years with suspected appendicitis. In patients with a low PAS the prevalence of appendicitis was 0 (95% CI, 0.0 to 0.08). Fleischman et al[9] performed a prospective study of children (3 to 18 years old) with suspected appendicitis and were categorized as low, intermediate, or high risk according to a previously derived score. Classification as intermediate or high risk by score had sensitivity 0.97 (95% CI, 88 to 100), specificity 0.41 (95% CI, 0. 31 to 0.50), positive LR 1.6 (95% CI, 1.4 to 1.9), negative LR 0.06 (95% CI, 0.02 to 0.30). Mandeville et al[10] performed a prospective study in children (4 to 17 years) with suspected appendicitis and evaluated the Alvarado and PAS scores. The overall prevalence of appendicitis in this cohort was 54%. The authors report the Cohen's kappa coefficients for interrater reliability of score calculation between 2 providers to be 0.67 for Alvarado and 0.59 for PAS. This suggests moderate agreement between providers. Cotton et al[11] prospectively validated the Pediatric Appendicitis Risk Calculator (pARC) in 2089 patients aged 5 to 20.9 years with a mean age of 12.4. Appendicitis was confirmed in 353 (16.9%) patients. In patients with a pARC score of less than 5 (very low risk), the prevalence of disease was 1.4 (0.5% to 2.3%) and a sensitivity of 100%. In those with a low score or very low score (14 or more), the negative LR was 0.08 (96%CI, 0.05 to 0.12), positive LR was 5.65 (95% CI, 5.07 to 6.31). The overall pARC score had an area under the curve of 0.89 (95% CI, 0.87 to 0.92), and the PAS score had an area under the curve of 0.8 (CI%, 0.77 to 0.82). The authors conclude that the pARC score had a higher sensitivity than PAS at any specificity. Kharbanda et al[12] enrolled 2625 children with suspected appendicitis and a mean age of 10.8 (SD, 3.8 years). A total of 1,018 (38.7%) were diagnosed with appendicitis. The primary outcome was the performance of a clinical prediction rule to identify children who are at low risk of appendicitis. The authors refined their rule to include the following parameters, absolute neutrophil count (ANC) of 6.75×

10^3/uL or more and absence of maximal tenderness in the right lower quadrant (RLQ) or ANC of 6.75×10^3/uL and absence of maximal tenderness in the RLQ but no abdominal pain with walking, coughing or jumping. This rule had a negative LR of 0.08 (95% CI, 0.05 to 0.13) and a positive LR of 1.29 (95% CI, 1.25 to 1.32) with a negative predictive value of 95.3% (95% CI, 92.3 to 97.0).

No studies have adequate LR to rule in or rule out appendicitis by using a risk score alone. It is important to note that no studies of adult patients met the methodology criteria for this clinical policy.

<H2>Summary

The diagnosis of appendicitis remains a clinical challenge for even the most experienced emergency physician. The Alvarado score is a well-known clinical scoring system from a retrospective study of patients with abdominal pain discussed in the prior guideline from 2010 in the *Annals of Emergency Medicine*.[5] It is often used by emergency physicians to assist in the detection of appendicitis and determine the need for a CT scan. This score's low diagnostic accuracy and agreement make it insufficient to use alone to identify pediatric and adolescent patients who do not need additional imaging. There is insufficient data to support the use of the Alvarado score in adult patients. In pediatric patients, PAS and the pARC score can aid in the identification of patients at low risk of appendicitis but should not be used alone to identify patients who do not warrant advanced imaging.

<H2>Future Research

Develop a prospectively validated clinical prediction rule that is reproducible across institutions to identify patients at high risk who do not need further imaging but likely have appendicitis. There is a similar need for the prediction rule to identify patients at low risk of appendicitis who can be treated conservatively without advanced imaging.

2. In ED patients with suspected acute appendicitis, is the diagnostic accuracy of ultrasound comparable to CT or MRI for the diagnosis of acute appendicitis?

Patient Management Recommendations

Level A recommendations. None specified.

Level B recommendations. In pediatric patients with suspected acute appendicitis, if readily available and reliable, use RLQ ultrasound to diagnose appendicitis.

An unequivocally* positive RLQ ultrasound with complete visualization of a dilated appendix has comparable accuracy to a positive CT or MRI in pediatric patients.

Level C recommendations. In adult patients with suspected acute appendicitis, an unequivocally* positive RLQ ultrasound has comparable accuracy to a positive CT or MRI for ruling in appendicitis.

*A nonvisualized or partially visualized appendix should be considered equivocal. Reasonable options for pediatric patients with an equivocal ultrasound and residual suspicion of acute appendicitis include MRI, CT, surgical consult, and/or observation, depending on local resources and patient preferences with shared decisionmaking.

<H2>Potential Benefit of Implementing the Recommendations:

- Lower rates of abdominal/pelvic CT for appendicitis evaluation, which in turn would lessen the risks of ionizing radiation.
- Faster throughput for ED patients when ultrasound results are unequivocal (see text for a description of the characteristics defining an unequivocal examination versus an equivocal/nondiagnostic [ND] examination).
- Enhanced patient engagement through shared decisionmaking.

<H2>Potential Harm of Implementing the Recommendations:

- Prolonged ED patient throughput when ultrasound is equivocal/ND.
- Increased resource usage when ultrasound is ordered, and results as ND, in patients already at a very low pretest probability for acute appendicitis (ie, those unlikely to need any imaging in the first place). For instance, in a patient with a very low pretest probability, an equivocal ultrasound may lead to CT, MRI, hospital observation, or surgical consult, which are unnecessary based on the patient's pretest odds of acute appendicitis.

- Reduced diagnostic accuracy when a point-of-care ultrasound (POCUS), rather than radiology-performed ultrasound, is used by clinicians lacking experience in POCUS for acute appendicitis.

<H3>Key words/phrases for literature searches: appendicitis, ruptured appendicitis, perforated appendicitis, computer assisted tomography, x-ray computed tomography, CT scans, ultrasonic tomography, medical imaging, ultrasonography, diagnostic ultrasound, ultrasound imaging, ultrasonic imaging, ultrasonic diagnosis, ultrasonographic imaging, sonography, medical sonography, diagnostic imaging, echography, computer echotomography, steady-state free precession MRI, magnetic resonance imaging, magnetization transfer contrast imaging, MRI Scan, fMRI, functional MRI, functional magnetic resonance imaging, emergency, emergency health service, hospital emergency service, emergency ward, emergency medicine, emergency care, emergency treatment, emergency department, emergency room, emergency service, emergency services, and variations and combinations of the key words/phrases. Searches included January 2009 to search dates of May 10 to 11, 2020.

<H2>Study Selection: Two hundred eighty-eight articles were identified in searches. Ninety-four articles were selected from the search results as potentially addressing this question and were candidates for further review. After grading for methodological rigor, 13 articles were selected from the search results for further review, with 0 class I studies, 2 class II studies, and 11 class III studies included for this critical question.

Diagnosis of acute appendicitis in the ED is typically accomplished with 1 of 3 medical imaging modalities: CT, MRI, and/or ultrasound. Ultrasound represents an attractive alternative to CT and MRI. Image acquisition is fast, ultrasound is generally more available than MRI, and requires no ionizing radiation like CT. Ultrasound imaging may also reduce costs compared with CT and can be performed as a bedside POCUS examination by trained practitioners. [13,14] Because of these advantages, an ultrasound-first approach to pediatric appendicitis diagnosis has been previously recommended by the American College of Radiology and the previous version of this ACEP clinical policy.[5,15] Using an ultrasound-first approach requires skilled sonographers

who are able to clearly report when the appendix has been fully visualized. The role of ultrasound imaging in adults with suspected acute appendicitis is less well-defined. In adult patients, there is a concern for false-negative studies, especially in women, older patients, and those patients with an elevated body mass index (BMI).[16] This critical question sought to evaluate whether the diagnostic accuracy of the ultrasound imaging was comparable with CT and/or MRI in suspected acute appendicitis in both pediatric and adult ED patients.

<H2>Characteristics of the Search and Included Studies

Two hundred eighty-eight articles were retrieved in the search for this critical question. On full-text screening, 94 of these were determined to be ED-based studies in which the diagnostic test characteristics (eg, sensitivity, specificity, positive LR, and negative LR) of ultrasound for suspected acute appendicitis were reported and/or could be calculated from the reported results. After the methodologist review, 2 studies were graded as class II, 11 were graded as class III, and 81 were graded as class X (Appendix E4). Two class III studies were meta-analyses,[13,17] in which 4 other class III studies[18, 19, 20, 21] were included, leaving an effective total of 7 unique class III studies. One class II study was included in a class III meta-analysis for its results on MRI but not for its results on ultrasound imaging.[17,22]

The prevalence of acute appendicitis in the primary research reports ranged from 32% to 54%.[22,23] In one class III meta-analysis assessing CT, MRI, and ultrasound separately in adult and pediatric patients,[17] prevalence ranged from 26% (pediatric CT) to 80% (adult ultrasound). Each imaging modality, for both adults and children, was assessed by at least 1 included article.

<H2>CT and MRI Diagnostic Accuracy

Diagnostic test characteristics for studies evaluating CT and MRI in suspected acute appendicitis, including both adults and children, are summarized in Table 1.[24-26] A primary limitation of most studies on CT and MRI in this population is that ultrasound imaging was often performed first, with CT or MRI as a second study. This had the potential to introduce incorporation bias in those studies in which CT or MRI interpreters were unblinded to ultrasound imaging results, spectrum bias, and partial verification or differential verification bias for studies in which the indication to obtain CT or MRI was a ND ultrasound examination. Nevertheless, sensitivity and specificity for CT in the included studies were similar to previously published

values of 94% and 95%, respectively.[13] Likewise, the MRI studies included had similar accuracy to prior reports (sensitivity 97% and specificity 96%).[23]

<H2>Ultrasound Diagnostic Accuracy Overall

Table 2 summarizes test characteristics for ultrasound studies.[27] The value of a positive test was high across nearly all studies. A positive (unequivocal) test was defined as complete visualization of a dilated appendix except in one class II and one class III study.[19,26] In the former, nonvisualization of the appendix with inflammatory signs was considered positive; in the latter, positive studies were subclassified by certainty of interpretation (probable versus equivocal). Nine pediatric studies showed a positive LR of 10 or more. Those pediatric studies with a positive LR of less than 10 included 1 small class II study,[22] 1 class III meta-analysis which exclusively studied POCUS,[13] and a small class III POCUS study within that same meta-analysis.[18] A recent class III meta-analysis including 548 pediatric patients[17] showed test characteristics similar to CT and MRI (sensitivity 91%, specificity 95%, positive LR 18, and negative LR 0.09).

Only 2 class III studies reported results on ultrasound for suspected acute appendicitis in adults.[17,18] Both had reasonably strong specificities (92%[18] and 95%[17]) and positive LRs (7.2[18] and 17[17]), comparable with CT and MRI. Neither had comparable sensitivity (Table 2) to CT or MRI (Table 1). The dearth of adult studies prevents strong recommendations regarding ultrasound in this patient population, but the 2 class III studies available would at least suggest a positive ultrasound in adults may be similarly interpreted as a positive result in children.

<H2>Equivocal Examinations

One of the most significant limitations of ultrasound imaging for suspected acute appendicitis is the high rate of ND/equivocal examinations. The most common and challenging type of ND examination is when no part of the appendix is visualized by the sonographer. In other ND examinations, the appendix may be only partially visualized or described with an indeterminate impression by the responsible clinician (ie, radiologist or, for the POCUS scan, the performing physician). The rate of ND examinations varied markedly between studies, likely reflecting differences in the practice environment and expertise with ultrasound imaging for acute appendicitis, ranging from 10% to 81%. Equivocal examinations present a serious challenge to the clinician and a point of potential confusion because quoted diagnostic statistics for ultrasound imaging may be calculated with different methods for reporting and summarizing

ND studies. Diagnostic accuracy differed markedly between studies in relation to the way ND examinations were included in calculations (Table 2 and Table 3), particularly sensitivity and negative LR. Multiple diagnostic strategies, which are beyond the scope of this question, are available to follow-up and evaluate a nonvisualized examination.

The most common way included studies treated ND examinations was to count anything other than an unequivocally positive study (a dilated appendix that is completely visualized) as a negative (4 studies, 2,362 patients). In this methodology, examinations resulting in nonvisualization of the appendix, partial visualization with or without dilation, and nondilated appendices with secondary signs (eg, inflammation) were counted the same as an unequivocally negative examination (complete visualization of a nondilated appendix without any secondary signs of acute appendicitis). Five studies did not report how the ND were counted or used other methods in reporting ND results. Specificity and positive LR remained high regardless of the handling of ND examinations (Table 3). This likely reflects the fact that counting any ND examination as negative was a particularly common practice and strengthens the confidence in the value of a positive US result.

<H2>Ultrasound, CT, and MRI by Pretest vs Posttest Probability

When ordering an imaging test for appendicitis, providers often have some estimation of the risk for the diagnosis. The Figure demonstrates posttest probability for each of the 3 modalities (ultrasound, CT, and MRI) at varying pretest probabilities (15%, 30%, and 50%). For each, the study-size weighted mean sensitivity and specificity were used to calculate an average positive and negative LR. The ultrasound was divided into those studies reporting ND examinations as negative, and those excluding ND examinations. In general, regardless of the reporting of ND examinations, posttest probability after a positive ultrasound was similar to probability after a positive CT or MRI, at any pretest probability. Posttest probabilities after a negative CT or MRI, or an unequivocally negative ultrasound, were similarly low for pretest probabilities of 15% and 30%. At a high pretest probability of 50%, posttest probability after negative CT or MRI approaches 5% and 2% to 3% by an unequivocally negative ultrasound. By contrast, among studies considering an ND ultrasound as “negative,” a negative result yielded a more than 5% posttest probability for acute appendicitis even when the pretest probability was

low (15%). Therefore, independent of a clinician's pretest probability, the results of the unequivocally negative ultrasound are comparable with CT or MRI.

One class III study[13] derived test-treatment thresholds for pediatric acute appendicitis based on published complication rates of appendectomy and risk of ionizing radiation from CT or MRI (ie, 0 in the latter). They calculated that a test with a positive LR of 5.8 and higher would meet the treatment threshold for ruling in acute appendicitis without further testing and a negative LR of 0.03 or less for ruling out acute appendicitis. Every class II or III ultrasound study except one[18] showed a positive LR of more than 5.8 in both adults and children. The lone study with a positive LR of less than 5.7 was included in another class III study as part of a meta-analysis,[13] for which the overall positive LR was 9.2. Both of the ultrasound studies, excluding ND examinations, had negative LR of less than 0.03 (Tables 2 and 3). One additional class III ultrasound study involving a re-evaluations pathway in the case of ND examination showed a negative LR of 0.03.[21] All other ultrasound studies, 3 of 5 CT studies and 3 of 5 MRI studies had a negative LR of more than 0.03.

<H3>Reevaluation and serial examination after ND ultrasound

Patients with ND ultrasounds may not warrant immediate CT or MRI imaging. One class III study evaluated a wait-and-reassess pathway for pediatric patients with an ND ultrasound in the ED.[21] Patients with an ND ultrasound (42%) were reassessed by clinical examination. Based on clinician discretion of the reexamination, most remaining patients were discharged from the ED (73/123), whereas those with ongoing clinical suspicion for acute appendicitis received a surgical consult. Among the latter group, 80% received a second ultrasound at a mean of 9.2 hours after the initial scan. The overall pathway had excellent negative and positive predictive value comparable with CT and MRI (sensitivity 97%, specificity 91%, positive LR 11, and negative LR 0.03) without requiring either. Notably, the pathway had far superior performance to either ultrasound alone when ND examinations were considered negative. This study suggests that observation, consultation, and reassessment may be reasonable alternatives to immediate CT or MRI in the case of an ND initial ultrasound.

<H2>Summary

Ultrasound imaging is useful for ruling in acute appendicitis and, when positive, is typically the only test needed before surgical consultation. This fact, along with its lack of ionizing radiation, as well as likely broader availability for most emergency providers compared

with MRI, should make it the initial first test of choice for pediatric patients. Although its role in adults is less clear, it may be a reasonable first test in select situations given a similarly high positive predictive value. The greatest limitation of ultrasound is the large amount of ND results, the rate of which varies widely between studies and settings. Negative predictive performance of ultrasound varies far more than MRI or CT, but in pediatric patients, this variation in performance appears closely related to whether or not ND examinations are counted as negative or excluded. An unequivocal negative ultrasound (visualization of a compressible tubular structure from tip to the cecum of less than 6 mm in diameter without secondary signs of inflammation) in a pediatric patient may be comparable with a negative CT or MRI based on low certainty of evidence (3 class III studies). For ND ultrasound examinations in children, a strategy of watchful waiting, including clinical reevaluation, surgical consultation, hospital observation, and/or serial ultrasound examination may be a reasonable alternative to immediate MRI or CT. Shared decisionmaking of the relative risks and benefits, and an assessment of local resources (eg, rapid MRI availability), is likely reasonable to guide such a decision.

<H2>Future Research

Future research should focus on reducing the rate of equivocal ultrasound examinations, increasing interoperator reliability, standardization of result reporting for both radiology-performed ultrasound and POCUS, and further examination of specific decision pathways integrating ultrasound that may enhance diagnostic performance and decrease the need for CT and/or MRI. To the latter point, further elaboration of the use of serial examination, observation, combination with clinical decision tools, and/or serial ultrasound testing could be significantly useful to provide stronger evidence to inform shared decisionmaking with equivocal ultrasound scans. Additional high-quality literature addressing the role of ultrasound in adult patients is likely to be beneficial as well.

3. In ED patients who are undergoing CT of the abdomen and pelvis for suspected acute appendicitis, does the addition of contrast improve diagnostic accuracy?

Patient Management Recommendations

Level A recommendations. None specified.

Level B recommendations. In adult and pediatric ED patients undergoing CT for suspected acute appendicitis, use intravenous contrast when feasible. The addition of oral or rectal contrast does not improve diagnostic accuracy.

Level C recommendations. In adult ED patients undergoing CT for suspected acute appendicitis, noncontrast CT scans may be used for the evaluation of acute appendicitis with minimal reduction in sensitivity.

<H2>Potential Benefit of Implementing the Recommendations:

- The use of intravenous contrast alone when obtaining a CT for patients with suspected appendicitis will result in sufficient diagnostic accuracy and improved ED throughput.

<H2>Potential Harm of Implementing the Recommendations:

- The use of intravenous contrast is dependent on adequate intravenous access. This may result in additional discomfort to patients. In addition, there is a small risk of anaphylactoid reaction when using intravenous contrast.
- The use of noncontrast CT scans may result in additional imaging if patients present again with recurrent symptoms.

<H3>Key words/phrases for literature searches: appendicitis, ruptured appendicitis, perforated appendicitis, diagnosis, diagnostic accuracy, accuracy, computer assisted tomography, x-ray computed tomography, CT scans, contrast media, contrast agent, contrast materials, radiocontrast media, radiocontrast agent, radiopaque media, IV contrast, intravenous contrast, oral contrast, rectal contrast, emergency, emergency health service, hospital emergency service, emergency ward, emergency medicine, emergency care, emergency treatment, emergency department, emergency room, emergency service, emergency services, and variations and combinations of the key words/phrases. Searches included January 2009 to search dates of May 10 to 11, 2020.

<H2>Study Selection: Two hundred twenty articles were identified in searches. Twenty-eight articles were selected from the search results as potentially addressing this question and were

candidates for further review. After grading for methodological rigor, 0 class I studies, 1 class II study, and 8 class III studies were included for this critical question.

<H2>Summary

CT imaging is frequently used when evaluating patients with suspected appendicitis. A review of the literature notes similar diagnostic accuracy of CT imaging for appendicitis for both adult and pediatric patients who receive intravenous or intravenous and oral contrast. In adult patients, CT performed with intravenous contrast should be considered comparable with CT without intravenous contrast.

<H2>Background

CT of the abdomen and pelvis imaging is frequently used in the evaluation of patients with suspected appendicitis. Radiology protocols for CT of the abdomen and pelvis often include the use of enteric or intravenous contrast. There is still debate regarding the diagnostic advantage of using contrast. The previously published clinical policy on the evaluation and management of patients with suspected appendicitis summarized the potential benefit of enteric contrast, which includes better differentiation of the appendix from surrounding structures, particularly in those patients with low BMI. In addition, this prior policy suggested that intravenous and enteric contrast help identify conditions other than appendicitis that may result in abdominal pain.[5] However, over the last decade, there have been significant advancements in CT imaging technology (eg, increased use of multirow detector CT and reduced slice width), resulting in improved image quality. This may affect the diagnostic advantage of enteric or intravenous contrast previously identified. The 2018 American College of Radiology Appropriateness Criteria for Adults and Children reports that CT abdomen and pelvis with intravenous contrast or without intravenous contrast may both be appropriate, further highlighting the uncertainty in this area.[29] However, this document does not comment on the use of enteric contrast.[29] With this critical question, we set out to review the recent literature on the role of contrast in the evaluation of appendicitis.

In 2012, in a class II study by Kepner et al,[30] 227 adult patients were randomized to receive intravenous contrast or oral contrast. Imaging was performed using a now somewhat older generation 16-slice scanner. The diagnosis of appendicitis was based on a combination of

CT findings, and clinical follow-up. If patients were admitted or had appendicitis, they had follow-up through electronic medical record review. The discharged patients were followed by telephone calls. A total of 80 patients have a CT diagnosis of appendicitis. The authors report that for intravenous contrast alone, the sensitivity was 100% (95% CI, 89.3 to 100), and specificity was 98.6% (95% CI, 91.6 to 99), resulting in a positive LR of 72 (95% CI, 10.3 to 504) and negative LR 0.00. For intravenous and oral contrast, the sensitivity was 100% (95% CI, 87.4 to 100), specificity 94.9 (95% CI, 86.9 to 98.4), and positive LR of 25 (95% CI, 8.24 to 75.8). There was no statistically significant difference between the use of intravenous and intravenous with oral contrast leading the authors to report that there was similar diagnostic performance. One difference that was noted, however, was that patients receiving intravenous contrast alone were discharged faster. Two other class III studies directly evaluated the role of contrast. Anderson et al,[31] used a 64-slice multidetector CT (MDCT) on a convenience sample of 303 adult patients, and Keyzer et al,[32] used a 4-slice MDCT in 131 adult patients. Both studies showed no difference in diagnostic accuracy, with the former demonstrating a positive LR of 34 (95% CI, 13.04 to 89.9) and negative LR of 0.00 for intravenous and a positive LR of 35 (95% CI, 13.3 to 91.9) with a negative LR 0.00 for intravenous and oral contrast. In another class III study by Jacobs et al,[33] 228 patients with suspected appendicitis underwent both a focused CT of the RLQ with oral contrast and a CT with both oral and intravenous contrast. They reported that the sensitivity of oral contrast was 76% and specificity 94% and for both the oral and intravenous contrast, the sensitivity was 91% and specificity 95%. Specific to pediatric patients, a 2018 class III study by Farrell et al[34] retrospectively compared pediatric cohorts receiving intravenous contrast alone versus oral contrast. A total of 558 64-MDCT scans met the inclusion criteria. Appendicitis was diagnosed in 22.4% of patients. The authors reported similar sensitivities of 93.8% (95% CI,84.8 to 98.3) and 94.6% (95% CI, 84.9 to 98.9) and specificities of 98.5% (95% CI, 95.8 to 99.7) and 98.3% (95% CI, 95.7 to 99.5) regardless of the administration of oral contrast.

A search of the medical literature identified 2 class III meta-analyses and 2 class III studies that addressed the use of rectal contrast or noncontrast CT diagnostic accuracy. A class III meta-analysis by Hlibczuk et al[35] included 7 studies with adult patients who had noncontrast CT for the evaluation of appendicitis. They reported a pooled sensitivity of 92.7% (95% CI, 89.5 to 95%) and specificity of 96.1% (95% CI, 94.2 to 97.5%). In another class III meta-analysis,

Rud et al[36] reported the pooled sensitivities for unenhanced CT 91% (95% CI, 87 to 93%), oral contrast only 89% (95% CI, 81 to 94%), intravenous contrast 96% (95% CI, 92 to 98), intravenous with oral contrast 96% (95% CI, 93 to 98), and rectal contrast 96% (95% CI, 92 to 98). There were no differences in pooled specificity estimates. Both of these meta-analyses included studies that were low quality, included a high risk of bias, and had a high prevalence of appendicitis. In a class III study, Seo et al[37] reported no difference in the sensitivity and specificity between low radiation dose noncontrast CT and standard radiation dose intravenous contrast CT in a 200-patient study. This study is limited by the confounder of different radiation doses. Chiu et al[38] evaluated the sensitivity of noncontrast CT to intravenous contrast CT in 100 patients with suspected appendicitis. In this cohort, with 44 of 100 patients diagnosed with appendicitis, they reported noncontrast CT had a lower sensitivity than intravenous contrast CT (91% versus 100%; *P*=.04) and greater specificity (100% versus 95%; *P*=.04) for the diagnosis of appendicitis. In a class X study by Hershko et al,[39] 232 adult patients with suspected appendicitis were randomized to receive a noncontrast, rectal contrast, or dual contrast (oral and intravenous) CT. They noted positive LR of 8.9, 12.3, and 8.2 and negative LR of 0.1, 0.05, and 0.0 in no contrast, rectal contrast, and dual contrast CTs, respectively. In another class X study by Ozdemir et al,[40] 293 patients older than 16 years with abdominal pain underwent noncontrast enhanced imaging using a 16-MDCT. They noted a sensitivity of 70.1% and specificity of 76.0% for a correct diagnosis in a noncontrast CT. It is important to note that the noncontrast studies have included only adult patients.

<H2>Future Research

Studies that look at the diagnostic accuracy of the noncontrast CT stratified by BMI would further clarify the need for contrast in patients presenting with suspected appendicitis.

<END ARTICLE>

REFERENCES

1. Rui P, Kang K, Ashman JJ. National Hospital Ambulatory Medical Care Survey: 2016 emergency department summary tables. 2016. Accessed April 5, 2022. https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2016_ed_web_tables.pdf

2. Mahajan P, Basu T, Pai CW, et al. Factors associated with potentially missed diagnosis of appendicitis in the emergency department. *JAMA Netw Open*. 2020;3:e200612.

3. Wong KE, Parikh PD, Miller KC, Zonfrillo MR. Emergency department and urgent care medical malpractice claims 2001-15. *West J Emerg Med*. 2021;22:333-338.

4. Bundy DG, Byerley JS, Liles EA, et al. Does this child have appendicitis? *JAMA*. 2007;298:438-451.

5. Howell JM, Eddy OL, Lukens TW, et al. Clinical policy: critical issues in the evaluation and management of emergency department patients with suspected appendicitis. *Ann Emerg Med*. 2010;55:71-116.

6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

7. González del Castillo J, Ayuso FJ, Trenchs V, et al. Diagnostic accuracy of the APPY1 Test in patients aged 2–20 years with suspected acute appendicitis presenting to emergency departments. *Emerg Med J*. 2016;33:853-859.

8. Saucier A, Huang EY, Emeremni CA, et al. Prospective evaluation of a clinical pathway for suspected appendicitis. *Pediatrics*. 2014;133:e88-e95.

9. Fleischman RJ, Devine MK, Yagapen MA, et al. Evaluation of a novel pediatric appendicitis pathway using high- and low-risk scoring systems. *Pediatr Emerg Care*. 2013;29:1060-1065.

10. Mandeville K, Pottker T, Bulloch B, et al. Using appendicitis scores in the pediatric ED. *Am J Emerg Med.* 2011;29:972-977.

11. Cotton DM, Vinson DR, Vazquez-Benitez G, et al. Validation of the pediatric appendicitis risk calculator (pARC) in a community emergency department setting. *Ann Emerg Med.* 2019;74:471-480.

12. Kharbanda AB, Dudley NC, Bajaj L, et al. Validation and refinement of a prediction rule to identify children at low risk for acute appendicitis. *Arch Pediatr Adolesc Med*. 2012;166:738-744. Published correction appears in Arch Pediatr Adolesc Med. 2012;166:901

13. Benabbas R, Hanna M, Shah J, et al. Diagnostic accuracy of history, physical examination, laboratory tests, and point-of-care ultrasound for pediatric acute appendicitis in the emergency department: a systematic review and meta-analysis. *Acad Emerg Med*. 2017;24:523-551.

14. Kharbanda AB, Christensen, EW, Dudley NC, et al. Economic analysis of diagnostic imaging in pediatric patients with suspected appendicitis. *Acad Emerg Med.* 2018;25:785-794.

15. Koberlein GC, Trout AT, Rigsby CK, et al. ACR appropriateness criteria suspected appendicitis-child. *J Am Coll Radiol.* 2019;16:S252-S263.

16. Atwood R, Blair S, Fisk M, et al. NSQIP based predictors of false negative and indeterminate ultrasounds in adults with appendicitis. *J Surg Res*. 2021;261:326-333.

17. Eng KA, Abadeh A, Ligocki C, et al. Acute appendicitis: a meta-analysis of the diagnostic accuracy of US, CT, and MRI as second-line imaging tests after an initial US. *Radiology*. 2018;288:717-727.

18. Fox JC, Solley M, Anderson CL, et al. Prospective evaluation of emergency physician performed bedside ultrasound to detect acute appendicitis. *Eur J Emerg Med*. 2008;15:80-85.

19. van Atta AJ, Baskin HJ, Maves CK, et al. Implementing an ultrasound-based protocol for diagnosing appendicitis while maintaining diagnostic accuracy. *Pediatr Radiol*. 2015;45:678-685.

20. Sivitz AB, Cohen SG, Tejani C. Evaluation of acute appendicitis by pediatric emergency physician sonography. *Ann Emerg Med*. 2014;64:358-364.e4.

21. Schuh S, Chan K, Langer JC, et al. Properties of serial ultrasound clinical diagnostic pathway in suspected appendicitis and related computed tomography use. *Acad Emerg Med*. 2015;22:406-414.

22. Thieme ME, Leeuwenburgh MM, Valdehueza ZD, et al. Diagnostic accuracy and patient acceptance of MRI in children with suspected appendicitis. *Eur Radiol*. 2014;24:630-637.

23. Repplinger MD, Pickhardt PJ, Robbins JB, et al. Prospective comparison of the diagnostic accuracy of MR imaging versus CT for acute appendicitis. *Radiology.* 2018;288:467-475.

24. Abo A, Shannon M, Taylor G, et al. The influence of body mass index on the accuracy of ultrasound and computed tomography in diagnosing appendicitis in children*. Pediatr Emerg Care*. 2011;27:731-736.

25. Kaiser S, Frenckner B, Jorulf HK. Suspected appendicitis in children: US and CT--a prospective randomized study. *Radiol.* 2002;223:633-638.

26. Orth RC, Guillerman RP, Zhang W, et al. Prospective comparison of MR imaging and US for the diagnosis of pediatric appendicitis. *Radiol.* 2014; 272:233-240.

27. 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.* 2013;20:697-702.

28. Sola R Jr, Theut SB, Sinclair KA, et al. Standardized reporting of appendicitis-related findings improves reliability of ultrasound in diagnosing appendicitis in children. *J Pediatr Surg*. 2018;53:984-987.

29. Garcia EM, Camacho MA, Karolyi DR, et al. Right lower quadrant pain-suspected appendicitis. ACR Appropriateness Criteria, 2018. Accessed April 5, 2022. https://www.acr.org/Clinical-Resources/ACR-Appropriateness-Criteria

30. Kepner AM, Bacasnot JV, Stahlman BA. Intravenous contrast alone vs intravenous and oral contrast computed tomography for the diagnosis of appendicitis in adult ED patients. *Am J Emerg Med.* 2012;30:1765-1773.

31. Anderson SW, Soto JA, Lucey BC, et al. Abdominal 64-MDCT for suspected appendicitis: the use of oral and IV contrast material versus IV contrast material only. *AJR Am J Roentgenol*. 2009;193:1282-1288.

32. Keyzer C, Cullus P, Tack D, et al. MDCT for suspected acute appendicitis in adults: impact of oral and IV contrast media at standard-dose and simulated low-dose techniques. *AJR Am J Roentgenol.* 2009;193:1272-1281.

33. Jacobs JE, Birnbaum BA, Macari M, et al. Acute appendicitis: comparison of helical CT diagnosis focused technique with oral contrast material versus nonfocused technique with oral and intravenous contrast material. *Radiol.* 2001;220:683-690.

34. Farrell CR, Bezinque AD, Tucker JM, et al. Acute appendicitis in childhood: oral contrast does not improve CT diagnosis. *Emerg Radiol.* 2018;25:257-263.

35. Hlibczuk V, Dattaro JA, Jin Z, et al. Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review. *Ann Emerg Med.* 2010;55:51-59.e1.

36. Rud B, Vejborg TS, Rappeport ED, et al. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev.* 2019;2019:CD009977.

37. Seo H, Lee KH, Kim HJ, et al. Diagnosis of acute appendicitis with sliding slab ray-sum interpretation of low-dose unenhanced CT and standard-dose I.V. contrast-enhanced CT scans. *AJR Am J Roentgenol*. 2009;193:96-105.

38. Chiu YH, Chen JD, Wang SH, et al. Whether intravenous contrast is necessary for CT diagnosis of acute appendicitis in adult ED patients?. *Acad Radiol*. 2013;20:73-78.

39. Hershko DD, Awad N, Fischer D, et al. Focused helical CT using rectal contrast material only as the preferred technique for the diagnosis of suspected acute appendicitis: a prospective, randomized, controlled study comparing three different techniques. *Dis Colon Rectum*. 2007;50:1223-1229.

40. Özdemir O, Metin Y, Tasci F, et al. Added value of diffusion-weighted MR imaging to non-enhanced CT in the evaluation of acute abdominopelvic pain. *Biomed Res India*. 2017;28:7735-7743.

Figure. Fagan nomograms for various acute appendicitis imaging strategies at low (15%), moderate (30%), and high pretest probability

Table 1. CT and MRI for appendicitis diagnosis.

Study	Class	Age Group	Prevalence (n total)	Imaging Protocol Features of Note	Sensitivity (%)	Specificity (%)	LR Positive	LR Negative
CT								
Abo et al[24] 2011	III	Pediatric	43% (128)	Twenty-nine did not receive US. 99 had US and CT, with CT performed second in most cases.	96 (86-99)	97 (90-100)	35.2 (9-138)	0.04 (0.01-0.15)
Eng et al[17] 2018	III	Pediatric	26% (1,498)	Meta-analysis, includes Kaiser 2002	96 (93-98)	95 (93-96)	18 (14-23)	0.04 (0.02-0.07)
Kaiser et al[25] 2002	III	Pediatric	43% (317)	CT always performed after US. Radiologist unblinded to US at time of CT interpretation	97 (93-99)	93 (89-97)	15 (8.5-25)	0.03 (0.01-0.08)
Eng et al[17] 2018	III	Adult	29% (1,027)	Meta-analysis	90 (85-93)	94 (91-95)	14 (11-18)	0.11 (0.08-0.15)
Repplinger et al[23] 2018	III	Pediatric and adult (age >12)	32% (198)	All patients had CT and MRI, but clinically-indicated CT was the impetus for enrollment	98 (90-100)	90 (83-94)	9.4 (5.9-16)	0.02 (0.00-0.06)
CT means, weighted by study N (total N=2,851, 4 studies. Eng 2018 includes Kaiser 2002)					94	94	16.7	0.06
MRI								
Orth et al[26]	II	Pediatric	37% (81)	All patients had MRI and US, with	93 (78-99)	94 (84-99)	15 (5.2-46)	0.07 (0.02-

2014				blinded interpretations				0.28)
Thieme et al[22] 2014	II	Pediatric	54% (104)	All patients had MRI after US	100 (92-100)	89 (76-96)	9.1 (3.9-18)	0.00 (0.00-0.16)
Eng et al[17] 2018	III	Pediatric	27% (287)	Meta-analysis, includes Theime 2014	97 (86-100)	97 (92-99)	34 (15-75)	0.03 (0.01-0.10)
Eng et al[17] 2018	III	Adult	52% (223)	NR	90 (85-94)	94 (91-96)	15 (7.1-30)	0.04 (0.01-0.10)
Repplinger et al[23] 2018	III	Pediatric and Adult (Age >12)	32% (198)	All patients had CT and MRI, but clinician-ordered CT was required for enrollment.	97 (88-99)	81 (74-87)	5.2 (3.7-7.7)	0.04 (0.00-0.11)
MRI means, weighted by study N (total N=7,894 studies. Eng 2018 includes Thieme 2014)					95	92	11.6	0.06

NR, Not Reported

Table 2. Ultrasound for appendicitis diagnosis.

Study	Class	Prevalence (n Total)	ND US %	How Were ND Examinations Considered?	Sensitivity (%)	Specificity (%)	LR Positive	LR Negative
Pediatric Ultrasound								
Orth et al[26] 2014	II	37% (81)	NR	Nonvisualized inflammation present = positive No inflammation, partial or no visualization = negative	86 (69-96)	100 (93-100)	∞ (5.6-∞)	0.14 (0.07-0.35)
Thieme et al[22] 2014	II	54% (104)	42%	ND = negative	76 (63-86)	89 (76-96)	6.9 (3.1-16)	0.27 (0.17-0.43)
Abo et al[24] 2011	III	37% (147)	81%	ND = negative	38 (26-52)	97 (90-99)	11.7 (3.7-37.0)	0.64 (0.52-0.79)
Benabbas et al[13] 2017 *Fox et al[18] 2008* *Sivitz et al[20] 2014*	III *III* *III*	35% (461) *54% (42)* *33% (264)*	NR *NR* *30%*	- *3 studies:* ND = negative - *1 study:* ND = positive or Negative based on Likert scale of 1-5 of how well visualized the appendix was.	86 (79-91) *74 (52-90)* *85 (75-95)*	91 (87-94) *90 (81-95)* *93 (85-100)*	9.2 (6.4-13.3) *4.7 (1.6-13.6)* *11.7 (6.9-19.8)*	0.17 (0.09-0.30) *0.31 (0.15-0.63)* *0.16 (0.10-0.27)*
Eng et al[17] 2018	III	27% (548)	NR	NR	91 (84-96)	95 (92-97)	18 (12-28)	0.09 (0.06-0.16)

Mittal et al[27] 2013 *ND Excluded*	III	33% (968) *NR (469)*	52% *NA*	ND = negative (primary analysis) *ND = excluded (secondary analysis)*	73 (59-86) *98 (95-100)*	97 (96-98) 92 (87-97)	24.5 (15.6-38.3) 11.8 (NR)	0.28 (0.24-0.34) 0.02 (NR)
Schuh et al[21] 2015 *Initial US* *Second US*	III	38% (294) *38% (294)* *43% (40)*	6% *42%* *43%*	If initial US was ND (n=123), patient was observed. If clinical suspicion remained on reevaluation, a second US and surgical consultation were obtained (n=40), where ND = negative.	97 (94-100) *80 (71-87)* *70 (44-89)*	91 (87-95) *95 (90-97)* *96 (76-100)*	11 (6.8-17) *27 (12-61)* *17 (2.3-134)*	0.03 (0.01-0.09) *0.21 (0.14-0.30)* *0.31 (0.15-0.65)*
Sola Jr et al[28] 2018	III	NR (766)	10%	ND = negative	69 (NR)	94 (NR)	11.5 (NR)	0.33 (NR)
van Atta et al[19] 2015 *Unequivocal only*	III	34% (512) *55% (231)*	55% *NA*	4 category results based on interpretation = positive vs negative, and certainty = probable vs unequivocal.	87 (81-91) *99 (96-100)*	94 (91-96) *97 (92-99)*	15 (9.8-23) *34 (11-104)*	0.14 (0.09-0.21) *0.01 (0.00-0.06)*
Kaiser et al[25] 2002	III	41% (600)	NR	ND results not allowed. Radiologists must report positive or negative only, even if confidence in diagnosis was low or	80 (75-85)	94 (91-96)	13 (8.8-20)	0.21 (0.17-0.27)

				appendix nonvisualized.				
Adult Ultrasound								
Fox et al[18] 2008	III	NR (83)	NR	ND = negative	59 (42-74)	92 (81-97)	7.2 (2.7-19.2)	0.64 (NR)
Eng et al[17] 2018	III	80% (169)	NR	NR	83 (70-91)	95 (92-97)	17 (3.8-72)	0.18 (0.12-0.26)

NR = Not Reported.

Table 3. Comparison of pediatric ultrasound test characteristics by the method of counting ND examinations

How Were ND Examinations Considered?	**Number of Studies (Classes)**	**N Total**	**Sensitivity**	**Specificity**	**LR Positive**	**LR Negative**
			Mean, Weighted by Study N			
ND = negative	4 Studies* - 3 class III - 1 class II	2,362	70%	95%	15.2	0.31
ND excluded	2 Studies*† - 2 class III	700	98%	94%	15.5	0.02
Method other than above	5 Studies*†‡ - 4 class III - 1 class II	2,202	85%	93%	12.2	0.16
Any method	9 Studies‡ - 7 class III - 2 class II	4,187	78%	95%	14.4	0.23

*Mittal et al[27] 2013 reported 2 analyses: ND as negative, and ND examinations excluded.

†van Atta et al[19] 2015 reported 2 analyses: ND as "likely positive" or "likely negative," and ND examinations excluded.

‡Studies included in Eng et al[18] 2018 or Bennabas et al[13] 2017 are only counted once as part of each meta-analysis. Eng et al[13] 2018 includes Schuh et al[21] 2015. Bennabas et al[13] 2017includes Fox et al[18] 2008 and Sivitz et al[20] 2014.

Appendix A. Literature classification schema.*

Design/ Class	**Therapy**†	**Diagnosis**‡	**Prognosis**§
1	Randomized, controlled trial or meta-analysis of randomized trials	Prospective cohort using a criterion standard or meta-analysis of prospective studies	Population prospective cohort or meta-analysis of prospective studies
2	Nonrandomized trial	Retrospective observational	Retrospective cohort Case control
3	Case series	Case series	Case series

*Some designs (eg, surveys) will not fit this schema and should be assessed individually.

†Objective is to measure therapeutic efficacy comparing interventions.

‡Objective is to determine the sensitivity and specificity of diagnostic tests.

§Objective is to predict outcome, including mortality and morbidity.

Appendix B. Approach to downgrading strength of evidence.

	Design/Class		
Downgrading	**1**	**2**	**3**
None	I	II	III
1 level	II	III	X
2 levels	III	X	X
Fatally flawed	X	X	X

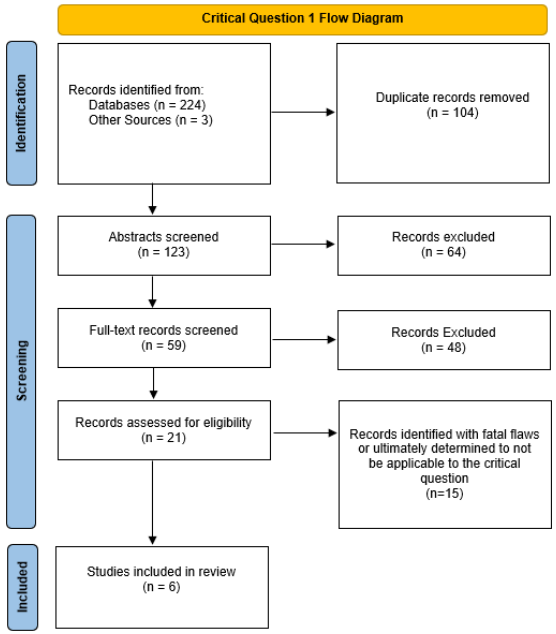
Appendix C. Likelihood ratios and number needed to treat.*

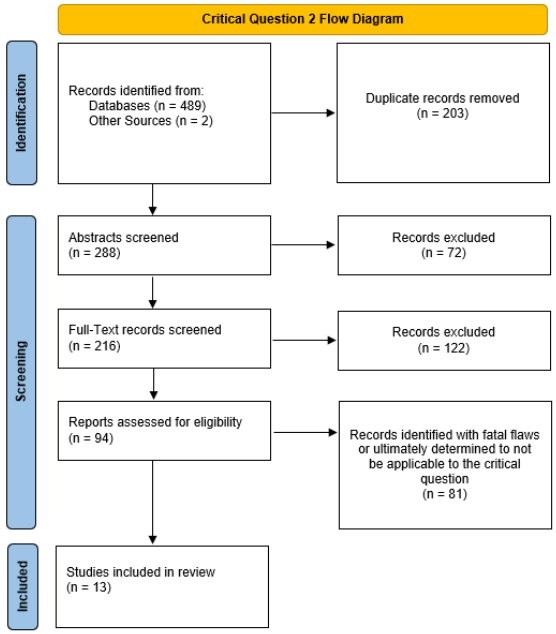
LR (+)	**LR (-)**	
1.0	1.0	Does not change pretest probability
1-5	0.5-1	Minimally changes pretest probability
10	0.1	May be diagnostic if the result is concordant with pretest probability
20	0.05	Usually diagnostic
100	0.01	Almost always diagnostic even in the setting of low or high pretest probability

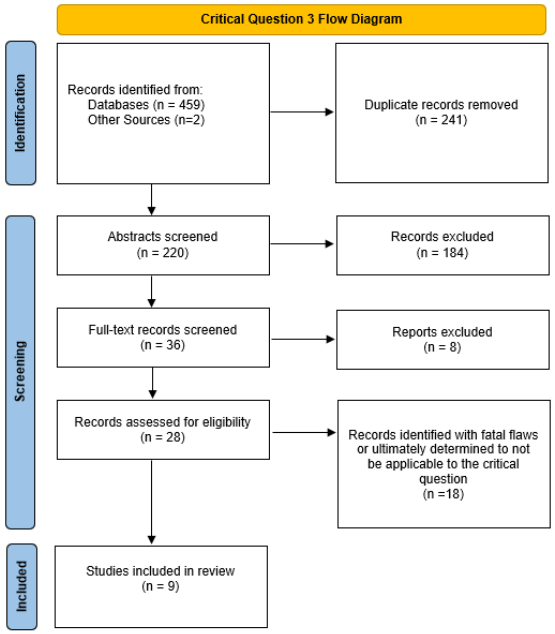
LR, likelihood ratio.

*Number needed to treat (NNT): number of patients who need to be treated to achieve 1 additional good outcome; NNT=1/absolute risk reduction×100, where absolute risk reduction is the risk difference between 2 event rates (ie, experimental and control groups).

Appendix D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagrams.[6]




Critical Question 2 Flow Diagram
Identification
Records identified from:
Databases (n = 489)
Other Sources (n = 2)
Duplicate records removed
(n = 203)
Screening
Abstracts screened
(n = 288)
Records excluded
(n = 72)
Full-Text records screened
(n = 216)
Records excluded
(n = 122)
Reports assessed for eligibility
(n = 94)
Records identified with fatal flaws or ultimately determined to not be applicable to the critical question
(n = 81)
Included
Studies included in review
(n = 13)


Critical Question 3 Flow Diagram
Identification
Screening
Included
Records identified from:
Databases (n = 459)
Other Sources (n=2)
Duplicate records removed
(n = 241)
Abstracts screened
(n = 220)
Records excluded
(n = 184)
Full-text records screened
(n = 36)
Reports excluded
(n = 8)
Records assessed for eligibility
(n = 28)
Records identified with fatal flaws or ultimately determined to not be applicable to the critical question
(n =18)
Studies included in review
(n = 9)

Evidentiary Table.

Study & Year Published	Class of Evidence	Setting & Study Design	Methods & Outcome Measures	Results	Limitations & Comments
Gonzalez Del Castillo et al[7] (2016)	III for Q1	Prospective cohort study at 4 academic medical centers in Spain from June to December 2014	Pediatric patients (2-20 y of age) with suspected appendicitis and abdominal pain <72 hours; study investigators recorded Alvarado Score elements blinded to diagnosis, but not imaging results; criterion standard was surgical pathology and telephone follow-up at 2 weeks	N=321 with prevalence of appendicitis 111/321 (35%); Alvarado Score >4 had sensitivity 0.92 (95% CI, 0.85-0.96) specificity 0.45 (95% CI, 0.38-0.52), positive LR 1.7 (95% CI, 1.5-1.9), and negative LR 0.2 (95% CI, 0.1-0.3); Alvarado Score >6 had sensitivity 0.76 (95% CI, 0.66-0.83) specificity 0.73 (95% CI, 0.66-0.79), positive LR 2.8 (95% CI, 2.2-3.6), and negative LR 0.3 (95% CI, 0.2-0.5)	All patients had appendectomy or telephone follow-up
Saucier et al[8] (2014)	III for Q1	Prospective cohort study at a single academic urban	Pediatric (3-17 y of age) with suspected appendicitis; Pediatric Appendicitis Score	N=196 patients with appendicitis prevalence of 33%; PPV for	PAS guided imaging and consultation decisions, which may

Commented [Editor1]:

		hospital	calculated by treating provider and incorporated into clinical pathway; Criterion standard was surgical pathology and one-day telephone follow-up	appendicitis by risk category: low risk (PAS 1-3) group 0 of 44 (0.0%), intermediate (PAS 4-7) risk 37 of 119 (31.1%), high (PAS 8-10) risk 28 of 33 (84.8%); Negative predictive value is 0; AUC 0.86 for PAS (95% CI, 0.81–0.91); PAS ≥6 had sensitivity 0.82 (95% CI, 0.70-0.90) and specificity 0.71 (95% CI, 0.62-0.79)	cause workup bias; limited telephone follow-up
Fleischman et al[9] (2013)	III for Q1	Prospective cohort in a single academic center	Children (3-18 y of age) with suspected appendicitis; patients categorized as low, intermediate or high risk according to previously derived score; physician judgment stratified patients as very low, low,	N=178 patients with appendicitis prevalence of 37%; classification as intermediate or high risk by score had sensitivity 0.97 (95% CI, 88-100), specificity 0.41 (95% CI, 0. 31-0.50), positive LR	Small sample size

			intermediate, or high risk; criterion standard was surgical pathology, chart review, and 2-week telephone follow-up	1.6 (95% CI, 1.4-1.9), negative LR 0.06 (95% CI, 0.02-0.30); classification as intermediate or high risk by physician judgment: sensitivity 1.0, specificity 0.50 (95% CI, not provided)	
Mandeville et al[10] (2011)	III for Q1	Prospective cohort; single center, urban, academic center	Children (4-17 y of age) with suspected appendicitis; Alvarado and Pediatric Appendicitis Scores recorded by treating physicians; 63% patients had scores recorded by 2 providers; Criterion standard was surgical pathology, chart review, and 2-week telephone follow-up	N=287 with appendicitis prevalence of 54%; Cohen's kappa coefficients for interrater reliability were 0.67 for Alvarado and 0.59 for PAS; PAS ≥ 6 had sensitivity 0.88 (95% CI, 0.83-0.93) and specificity 0.50 (95% CI, 0.42-0.59). AUC 0.78 (95% CI, 0.72-0.83); Alvarado score ≥ 7 had sensitivity 0.76 (95%	High prevalence of appendicitis may result in spectrum bias

				CI, 0.69-0.82) and specificity 0.72 (95% CI, 0.65-0.80); AUC 0.77 (95% CI, 0.72-0.83)	
Cotton et al[11] (2019)	III for Q1	Prospective cohort; 11 community EDs	Patients (5-20.9 y) with a chief complaint of RLQ pain. Physicians entered variables for the pARC score and PAS into a clinical decision support system. Criteria standard was diagnosis of appendicitis within 7 days of the index visit by hospital diagnosis and procedural code for appendectomy.	N=2,089 with a prevalence of appendicitis was 16.9 percent. pARC score <5 had sensitivity 100 (95% CI, 0.83-0.93) and prevalence of appendicitis of 1.4% (95% CI, 0.5-2.3%). A pARC score of low to very low (<=14) the negative LR 0.08 (96% CI, 0.05-0.12), positive LR 5.65 (95% CI, 5.07-6.31). Overall pARC score had a AUC of 0.89 (95% CI, 0.87-0.92) and the PAS score had an AUC 0.8 (95% CI, 0.77-	No information regarding if imaging was performed in addition to the score was provided. Patients aged <5 y were not included.

				0.82).	
Kharbanda et al[12] (2012)	III for Q1	Prospective cohort, 10, pediatric EDs, one sites data excluded	Children (3-18 y) being evaluated for suspected appendicitis with treating physician was ordering laboratories, imaging or surgical consultation. Criterion standard of appendicitis based on the attending pathologist's written report; for discharge patients telephone follow-up within 2 weeks or medical record review at enrolling facilities.	N=2,625 with appendicitis prevalence of 38.8%; Refined rule included the following parameters, ANC<= 6.75×10^3/uL and absence of maximal tenderness in the RLQ or ANC<= 6.75×10^3 /uL and absence of maximal tenderness in the RLQ but no abdominal pain with walking, coughing, or jumping. This rule had a negative LR.08 (95% CI, 0.05-0.13) and positive LR 1.29 (95% CI, 1.25-1.32) with a NPV 95.3% (95% CI, 92.3-97.0).	High prevalence of appendicitis may result in spectrum bias; limited telephone follow-up.
Abo et al[24] (2011)	III for Q2	Prospective cohort; single center,	Children (2-20 y) with suspected appendicitis;	N=176 with appendicitis prevalence of 42%; 147	Imaging interpretation not blinded to clinical

		urban, academic center	US and CT at discretion of treating providers; interpretation by treating radiologist; appendicitis diagnosis determined by surgical pathology, chart review and 1-week phone follow-up	patients had US, 128 had CT, and 99 had both. If nondiagnostic US was categorized as negative, US sensitivity 0.38 (95% CI, 0.26-0.52), specificity 0.97 (95% CI, 0.90-0.99), positive LR 11.7 (95% CI, 3.7-37), negative LR 0.64 (95% CI, 0.52-0.79); CT sensitivity 0.96 (95% CI, 0.86-0.99), specificity 0.97 (95% CI, 0.90-1.0), positive LR 35 (95% CI, 9-138), negative LR 0.04 (95% CI, 0.01-0.15)	data; CT generally used as second-line test
Benabbas et al[13] (2017)	III for Q2	Meta-analysis of prospective studies	Included studies of pediatric (<21 y) ED patients with suspected appendicitis; Random effects models to estimate pooled test	ED POCUS (N=4 studies): Pooled sensitivity 0.86 (95% CI, 0.79-0.91), specificity 0.91 (95% CI, 0.87-0.94),	Most studies at high risk of differential verification bias

			characteristics	positive LR 9.2 (95% CI, 6.4-13), negative LR 0.17 (95% CI, 0.09-0.30)	
Eng et al[17] (2018)	III for Q2	Meta-analysis of prospective and retrospective studies	Included studies of second-line US, CT, or MR in pediatric and adult patients who had an initial nondiagnostic ultrasound; quality assessed by 3 investigators; separate fixed effect models were used to estimate pooled sensitivity and specificity in pediatric and adult populations	37 studies were included; 9 studies and evaluated ultrasound, 30 studies evaluated CT, and 11 studies evaluated MR Pediatric US: sensitivity 0.91 (95% CI, 0.84-0.96), specificity 0.95 (95% CI, 0.92-0.97); Adult US: sensitivity 0.83 (95% CI, 0.70-0.91), specificity 0.91 (95% CI, 0.59-0.99); Pediatric CT: sensitivity 0.96 (95% CI, 0.93-0.98), specificity 0.95 (95% CI, 0.93-0.96); Adult CT: sensitivity 0.90 (95% CI, 0.85-0.93), specificity 0.94 (95% CI,	Unclear how these results apply to first-line imaging choice.

				0.91-0.95). Pediatric MR: sensitivity 0.97 (95% CI, 0.86-1.0%), specificity 0.97 (95% CI, 0.92-0.99%). Adult MR: sensitivity 0.90 (95% CI, 0.85-0.94), specificity 0.94 (95% CI, 0.91-0.96).	
Mittal et al[27] (2013)	III for Q2	Retrospective cohort study of multicenter, academic center	Children (3-18 y) with suspected appendicitis US ordered at discretion of treating provider and interpreted by treating radiologist. Appendicitis diagnosis determined by surgical pathology, chart review and 2-week telephone follow-up.	N = 2,635 with appendicitis prevalence of 39%. US performed in 965 (36.8%) patients. Sensitivity 0.73 (95% CI, 0.59-0.86%), specificity 0.97 (95% CI, 0.96-0.98), positive LR 25 (95% CI, 16-38), negative LR 0.28 (95% CI, 0.24-0.34)	Attrition not reported. Abstraction of US report was not blinded to patient outcome.

Orth et al[26] (2014)	II for Q2	Prospective cohort study in single academic center	Pediatric (3-17 y) patients with suspected appendicitis and US ordered; All patients had US and MR. US and MR interpretations were blinded to one another and clinical outcome. Appendicitis diagnosis determined by surgical pathology, chart review, and phone follow-up	N=81 with appendicitis prevalence of 37%. US sensitivity 0.86 (95% CI, 0.69-0.96), specificity 1.0 (95% CI, 0.93-1.0). MR sensitivity 0.93 (95% CI, 0.78-0.99), specificity 0.94 (95% CI, 0.84-0.99).	Small sample size. All patients received US and MR.
Repplinger et al[23] (2018)	III for Q2	Prospective cohort study in single academic center	Pediatric (>12 y) and adult patients with suspected appendicitis and CT ordered; All patients had CT with IV/oral contrast and MR; CT and MR interpreted on 5-point scale for likelihood of appendicitis by 3 fellowship-trained abdominal radiologists	N=198. Appendicitis prevalence was 32%. For likelihood of appendicitis categorized as possible to definite, sensitivity and specificity were 0.97 (95% CI, 0.88-0.99) and 0.81 (95% CI, 0.74-0.87) for MR	1,224 of 1,551 eligible patients were not included.

			blinded to clinical data; Appendicitis diagnosis determined by surgical pathology, chart review, and one-month phone follow-up	imaging and 0.98 (95% CI, 0.90-1.0) and 0.90 (95% CI, 0.83-0.94) for CT, respectively. Positive LR 5.2 (95% CI, 3.7-7.7) and Negative LR 0.04 (95% CI, 0-0.11) for MR Positive LR 9.4 (95% CI, 5.9-16) and negative LR 0.02 (95% CI, 0.00-0.06) for CT.	
Schuh et al[21] (2015)	III for Q2	Prospective cohort study in single academic center	Pediatric (4-17 y) patients with suspected appendicitis, baseline pediatric appendicitis score ≥2, and need for imaging according to treating clinician; All patients received initial US. If initial US was	N=294 with appendicitis prevalence of 38%. 294 had initial US and 40 had interval US. Initial US had sensitivity 0.80 (95% CI, 0.71-0.87), specificity 0.95 (95% CI,	

			equivocal, an additional interval US was performed at discretion of providers; appendicitis diagnosis determined by surgical pathology, chart review, and 1-month phone follow-up	0.90-0.97), and 0.42 (95% CI, 0.36-0.48) equivocal rate. Interval US had sensitivity 0.70 (95% CI, 0.44-0.89), specificity 0.96 (95% CI, 0.76-1.0), and 0.43 (95% CI, 0.27-0.59) equivocal rate.	
Sola et al[28] (2018)	III for Q2	Prospective cohort study in single academic center	Patients at a pediatric ED with suspected appendicitis; use of US guided by Alvarado score; appendicitis diagnosis determined by surgical pathology, chart review, and 1-week phone follow-up	N=840 with appendicitis prevalence 28%. 766 had US; US sensitivity 0.69 and specificity 0.94.	Possible spectrum bias because use of US depended stratified by Alvarado score; CIs (or raw data) for sensitivity and specificity were not provided.
Thieme et al[22] (2014)	II for Q2	Prospective cohort study in single academic center	Pediatric (4-18 y) ED patients with suspected appendicitis; patients	N=104 with appendicitis prevalence 56%.	Small study with high prevalence of appendicitis.

			received US and MR within 2h; appendicitis diagnosis by review of hospital and outpatient medical records	US sensitivity 0.76 (95% CI, 0.63-0.86), specificity 0.89 (95% CI, 0.76-0.96). MR sensitivity 1.0 (95% CI, 0.92-1.0), specificity 0.89 (95% CI, 0.76-0.96).	
van Atta et al[19] (2015)	III for Q2	Prospective cohort study in single urban, academic center	Patients at a pediatric ED with suspected appendicitis; patients received US as first-line imaging; appendicitis diagnosis by review of hospital records. No telephone follow-up.	N=512 with appendicitis prevalence 34%; US sensitivity 0.86 (95% CI, 0.81-0.91), specificity 0.94 (95% CI, 0.91-0.96).	No active follow-up of patients who did not have surgery
Fox et al[18] (2008)	III for Q2	Prospective cohort study in single academic center	Patients (adult and pediatric) with suspected appendicitis and imaging (radiologist US or CT) ordered; bedside US performed by a study emergency physician but did not influence care; appendicitis diagnosis	N=132 with appendicitis prevalence 44%. US sensitivity 0.65 (95% CI, 0.52-0.76), specificity 0.90 (95% CI, 0.81-0.95).	Treating providers and radiologists blinded to bedside US result.

			determined by surgical pathology, chart review and phone follow-up 2 weeks–3 months.		
Kaiser et al[25] (2002)	III for Q2	Prospective randomized clinical trial in single academic center	Patients at pediatric ED randomized to US vs US and CT; in US and CT arm, US performed first; appendicitis diagnosis determined by surgical pathology, chart review and 6-month questionaire	N=600 with appendicitis prevalence 41% 283 patients in US only arm and 317 in US and CT arm. Total number who had US was 600. US sensitivity 0.80 (95% CI, 0.75-0.85), specificity 0.94 (95% CI, 0.91-0.96). CT sensitivity 0.94 (95% CI, 0.91-0.96), specificity 0.97 (95% CI, 0.92-0.99).	Results biased in favor of CT, because radiologist who interpreted CT was not blinded to US result.
Sivitz et al[20] (2014)	III for Q2	Prospective cohort study in single academic center	Pediatric patients with suspected appendicitis; US performed by pediatric	N=254. Among 231 analyzed patients, prevalence of appendicitis	9% patients lost to follow-up. Some patients received more

			emergency medicine physicians; appendicitis diagnosis determined by surgical pathology, chart review and phone follow-up	was 33%. 287 ultrasound examinations performed in 254 patients. Sensitivity 0.85 (95% CI, 0.75-0.95), specificity 0.93 (95% CI, 0.85-1.0), positive LR 11.7 (95% CI, 6.9-20), negative LR 0.16 (95% CI, 0.1-0.27).	than one ultrasound.
Chiu et al[38] (2013)	III for Q3	Retrospective cohort study in single academic center	Adult patients with suspected appendicitis received CTs both with and without IV contrast. Patients who received oral contrast were excluded; CTs interpretated by 2 study radiologists blinded to clinical data and original interpretation; diagnosis of	N=100 with appendicitis prevalence of 44%. Noncontrast CT had lower sensitivity than contrast CT (91% vs 100%, *P*=.04) and greater specificity (100% vs 95%, *P*=.04)	Convenience sample with relatively high prevalence of appendicitis could result in spectrum bias.

			appendicitis by pathology and 6-month chart review		
Anderson et al[31] (2009)	III for Q3	Randomized controlled trial in single academic center	Adults with acute abdominal pain randomized to CT with oral and IV contrast vs CT with IV contrast and no oral contrast; 2 study radiologists interpreted each CT with radiologist confidence measured by likelihood of appendicitis on 5-point scale; diagnosis of appendicitis by chart review	N=303 with appendicitis prevalence of 9%. No significant difference in distributions of radiologist confidence between the 2 groups. Confidence not associated with BMI or intraabdominal fat.	Study did not assess differences in sensitivity and specificity with the addition of oral contrast.
Kepner et al[30] (2012)	II for Q3	Randomized controlled trial in single academic center	Adults with suspected appendicitis randomized to CT with oral and IV contrast vs CT with IV contrast and no oral contrast; interpretation by 2 independent study radiologists blinded to	N=227 with appendicitis prevalence of 35%; interpretation was discrepant for 6 patients in each group; IV contrast: sensitivity 100% (95% CI, 89%-100%), specificity 99% (95% CI,	CTs were interpretated study radiologists. Contemporaneous CT interpretation influenced clinical management and outcome assessment (workup bias)

			original interpretation and clinical data; diagnosis of appendicitis by pathology, chart review and telephone follow-up	92%-100%); IV and oral contrast: sensitivity 100% (95% CI, 87%-100%), specificity 95% (95% CI, 87%-98%)	16-slice CT scanner.
Keyzer et al[32] (2009)	III for Q3	Randomized controlled trial in single academic center	Adults with suspected appendicitis. All patients had CTs with and without IV contrast; Arms: oral contrast and no oral contrast; 2 study radiologists, blinded to clinical data, interpretated 4 CTs for each patient: CT oral contrast, CT oral and IV contrast, CT no oral/no IV contrast, CT no oral/IV contrast; diagnosis of appendicitis by pathology, chart review and telephone follow-up	N=131 with appendicitis prevalence of 25% (20/66 in oral contrast group and 13/65 in no oral contrast group); sensitivity and specificity were not significantly different for either radiologist comparing 4 types of CT scans.	CTs were interpretated study radiologists. Small sample size. Contemporaneous CT interpretation influenced clinical management and outcome assessment (workup bias) 4-slice CT scanner.
Seo et al[37] (2009)	III for Q3	Retrospective cohort in single	Adult (≥15 y) patients with suspected appendicitis	N=207 with appendicitis prevalence 34%;	Small sample size.

		academic center	received low radiation dose, noncontrast CT and standard radiation dose, IV contrast CT; interpretation by 2 independent study radiologists blinded to original interpretation and clinical data; surgical pathology, chart review and telephone follow-up	sensitivity and specificity were not significantly different for either radiologist comparing 2 types of CT scans.	Unable to separate potential effects of radiation dose and IV contrast.
Hlibczuk et al[35] (2010)	III for Q3	Meta-analysis of prospective and retrospective studies	Included studies of noncontrast CT for evaluation of appendicitis in adult (≥16 y), ED patients with at least 2 weeks follow-up Random effects model to estimate pooled sensitivity and specificity	N=7 studies Pooled sensitivity was 92.7% (95% CI, 89.5%-95.0%) and specificity was 96.1% (95% CI, 94.2%-97.5%)	
Rud et al[36] (2019)	III for Q3	Meta-analysis of prospective and	Included ED and non–ED-based studies of CT for	N=64 studies included with median appendicitis	Only 2/64 studies were assessed as low risk of

		retrospective studies	evaluation of appendicitis in adult (≥14 y) patients; random effects model to estimate pooled sensitivity and specificity for different types of contrast (oral, rectal and IV)	prevalence of 0.43; Pooled sensitivity estimates: unenhanced CT 91% (95% CI, 87%-93%), oral contrast only 89% (95% CI, 81%-94%), IV contrast 96% (95% CI, 92-98), IV and oral contrast 96% (95% CI, 93-98), rectal contrast (95% CI, 92-98). Pooled specificity estimates were similar for different types of contrast, with point estimates ranging from 93%-95%.	bias in all 4 domains; relatively high prevalence of appendicitis; no study was considered high quality with differential verification a common threat to bias.
Farrell et al[34] (2018)	III for Q3	Retrospective cohort study in single urban, academic center	Pediatric (0-17 y) ED patients with acute, nontraumatic abdominal pain who received CT with	N=588 with appendicitis prevalence 22%. 270 patients in oral contrast group and 288 in	No active follow-up and attrition not reported.

			IV contrast. CT protocol changed from addition of oral contrast to noncontrast halfway during study period; surgical pathology and chart review for follow-up	noncontrast group; oral contrast (N=270): sensitivity 0.94 (95% CI, 0.85-0.98) and specificity 0.99 (95% CI, 0.96-1.0); noncontrast (N=288): sensitivity 0.95 (95% CI, 0.85-0.99) and specificity 0.98 (95% CI, 0.96-1.0).	
Jacobs et al[33] (2001)	III for Q3	Prospective cohort study in single urban, academic center	Patients with RLQ pain and suspected appendicitis with CT ordered; all patients received 2 CT scans: (1) Focused (over RLQ) CT with oral contrast and (2) CT abdomen with oral and IV contrast; both CTs per patient were interpretated by 3 study radiologists blinded to clinical data; diagnoses were established by surgical	N=228 with appendicitis prevalence 22%. 8% patients were lost to follow-up, leaving 210 for analysis; focused CT with oral contrast only: mean sensitivity 0.76, mean specificity 0.94, AUC 0.92; CT with oral and IV contrast: mean sensitivity 0.91, mean specificity 0.95, AUC 0.96.	Chart review methods to establish diagnosis were not described.

			and/or chart review		