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

This DRAFT is EMBARGOED – Not for Distribution

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

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

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 emergency department patients with possible acute appendicitis, can a clinical prediction rule be used to identify patients for whom no advanced imaging is required? 2) In emergency department patients with suspected acute appendicitis, is the diagnostic accuracy of ultrasound comparable to CT or MRI for the diagnosis of acute appendicitis? 3) In emergency department patients who are undergoing CT 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.

INTRODUCTION

Abdominal pain is a high-volume, high-risk chief complaint. In 2016, patients with abdominal pain composed 8.6% of emergency department (ED) visits and almost 200,000 patients have the diagnosis of appendicitis each year.[1] Missed diagnosis of appendicitis remains an area at high risk for 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 utilization of computed tomography (CT) in patients with possible appendicitis, such widespread use may be unnecessary. Traditional teaching suggests that clinical indicators (eg, signs, symptoms, laboratory tests) exist that may be utilized 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 medicine physician is the identification of patients who are so unlikely to have appendicitis that do not warrant imaging to confirm the diagnosis.

Similarly, patients with high clinical suspicion for appendicitis may be referred to a surgeon with minimal or no testing.[4]

Once the decision is made to image, 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 patient with suspected appendicitis as it also does not involve ionizing radiation. Understanding the differences in diagnostic accuracy of ultrasound, CT, and MRI can inform decisions about imaging.

This policy is an update of the 2010 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 decision-making in the risk stratification of patients with possible appendicitis? (2) In adult patients with suspected acute appendicitis who are undergoing a computed tomography scan, what is the role of contrast? (3) In children with suspected acute appendicitis who undergo diagnostic imaging, what are the roles of computed tomography and ultrasound in diagnosing acute appendicitis?

METHODOLOGY

This ACEP clinical policy is based on a systematic review and critical descriptive analysis of the medical literature and is reported in accordance with PRISMA guidelines.[6]

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 A).

Assessment of Risk of Bias and Determination of Classes of Evidence

Each study identified as eligible by the subcommittee was independently graded by two 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" due to "inapplicability" for one critical question may be considered perfectly relevant for another question and graded I – 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, or prognostic, or a meta-analysis. Subsequent design types (ie, Design 2 and Design 3) represent respectively weaker study designs. 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, two 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 one additional methodologist, resulting in a final Class of Evidence assignment (ie, Class I, Class II, Class III, or Class X) (Appendix B). 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, 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.

<u>Translation of Classes of Evidence to Recommendation Levels</u>

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 one 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 one 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, 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 C).

Evaluation and Review of Recommendations

Once drafted, the policy was distributed for internal review (by members of the entire committee) followed by 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 Web site, and sent to other pertinent physician organizations. The responses were used to further refine and enhance this clinical policy, although responses do 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.

Application of the Policy

This policy is not intended to be a complete manual on the evaluation and management of adult patients with acute heart failure syndromes but rather a focused examination of critical questions that have particular relevance to the current practice of emergency medicine. 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 for which medical literature exists 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, non-traumatic 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 pregnant patients.

CRITICAL QUESTIONS

1. In emergency department 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.

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, due to insufficient data, do not use clinical prediction rules to identify patients for whom no advanced imaging is required.

<u>Potential Benefit of Implementing the Recommendations:</u>

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

<u>Potential Harm of Implementing the Recommendations:</u>

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

<u>Key words/phrases for literature searches:</u>
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 10-11, 2020.

<u>Study Selection:</u> One hundred and twenty articles were identified in searches. Eighteen articles were selected from the search results as potentially addressing this question and were candidates for further review. After grading for methodological rigor, 4 articles were selected from the search results for further review with zero Class I studies, 0 Class II studies, and 4 Class III study included for this critical question (Appendix D).

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 review of the initial set of 18 articles, only 4 met criteria for inclusion. All 4 articles were level III evidence. Gonzalez del Castillo et al[7] compared a prospective observational cohort of younger patients ages 2-20 years old using the APPY1 test to risk stratify the patients. The APPY1 test evaluates for C- reactive protein and calprotectin levels that gets combined with a WBC result. Patients were also broken out using Alvarado score into low, intermediate, or high-risk cohorts as part of a secondary data analysis. An 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) for the diagnosis of appendicitis. Saucier et al[8] evaluated the pediatric appendicitis score (PAS) in patients 136 patients aged 3-17 years old with suspected appendicitis. In patients with a low PAS the prevalence of appendicitis was 0 (95% CI 0.0-0.08). Fleishman et al[9] performed a prospective study of children (3-18 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 88-100), specificity 0.41 (95% CI 0. 31-0.50), positive LR 1.6 (95% CI 1.4-1.9), negative LR 0.06 (95% CI 0.02-0.30). Mandeville et al[10] performed a prospective study in children (4-17 years old) 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. All studies did not 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.

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 detection of appendicitis and determine need for CT scan. These scores low diagnostic accuracy and agreement make them insufficient to use alone to identify pediatric and adolescent patients that do not need additional imaging. There is insufficient data to support the use of the Alvarado score in adult patients.

Future Research

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

2. In emergency department 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.

Level B recommendations. In pediatric patients with suspected acute appendicitis, if readily available and reliable, use right lower quadrant (RLQ) ultrasound (US) to diagnose appendicitis.

An unequivocally* positive RLQ US 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 US has comparable accuracy to a positive CT or MRI for ruling in appendicitis.

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

<u>Potential Benefit of Implementing the Recommendations:</u>

- 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 exam vs. an equivocal/non-diagnostic exam)
- Patient-centering of care when diagnostic options are at equipoise for pediatric patients (e.g. US vs. CT or MRI, serial exam or observation after non-diagnostic US vs. follow-up CT or MRI)

<u>Potential Harm of Implementing the Recommendations:</u>

- Prolonged ED patient throughput when US is equivocal/non-diagnostic
- Increased resource utilization when US is ordered, and results as non-diagnostic, 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 very low pretest probability an equivocal US 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 point of care US (POCUS), rather than radiology-performed US, is used by clinicians lacking experience in POCUS for acute appendicitis.

<u>Key words/phrases for literature searches:</u> 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 10-11, 2020.

Study Selection: Two hundred and 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 zero Class I studies, 2 Class II studies, and 11 Class III studies included for this critical question.

Diagnosis of acute appendicitis in the emergency department (ED) is typically accomplished with one of three medical imaging modalities: computed tomography (CT), magnetic resonance imaging (MRI), and/or ultrasound (US). US represents an attractive alternative to CT and MRI. Image acquisition is fast, US is generally more available than MRI, and requires no ionizing radiation like CT. US may also reduce costs compared to CT and can be performed as a bedside a point of care (POCUS) exam by trained practitioners. [11, 12] Because of these advantages, an US-first approach to pediatric appendicitis diagnosis has been previously recommended by the American College of Radiology[13] and the previous version of this ACEP Clinical Policy.[5] Utilizing an US first approach requires skilled sonographers who are able to clearly report when the appendix has been fully visualized. The role for US 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 BMI.[14] This Critical Question sought to evaluate whether its diagnostic accuracy of US was comparable to CT and/or MRI in suspected acute appendicitis in both pediatric and adult ED patients.

Characteristics of the search and included studies

Two hundred and eighty-eight articles were retrieved in the search for this Critical Question. On full text screening, 94 were of these were determined to be ED-based studies where the diagnostic test characteristics (e.g. sensitivity, Specificity, positive likelihood ratio {positive LR}, negative likelihood ratio {negative LR}) of US for suspected acute appendicitis was reported and/or could be calculated from the reported results. After methodologist review, 2 studies were graded as class II, 11 graded as class III, and 81 graded as class X (Appendix D). Two class III studies were meta-analyses,[11,15] in which four other class III studies[16, 17, 18, 19] were included, leaving an effective total of 7 unique class III studies. One class II study[20] was included in a class III meta-analysis[15] for its results on MRI, but not for its results on US.

Prevalence of acute appendicitis in the primary research reports ranged from 32%[21] to 54%.[20] In one class III meta-analysis assessing CT, MRI, and US separately in adult and pediatric patients,[15] prevalence ranged from 26% (pediatric CT) to 80% (adult ultrasound). Each imaging modality, for both adults and children, was assessed by at least one included article.

<u>CT and MRI Diagnostic Accuracy</u>

Diagnostic test characteristics for studies evaluating CT and MRI in suspected acute appendicitis, including both adults and children, are summarized in Table 1. A primary limitation of most studies on CT and MRI in this population is that US was often performed first, with CT or MRI as a second study. This had the potential to introduce incorporation bias in those studies where CT or MRI interpreters were unblinded to US results, spectrum bias, and partial verification or differential verification bias for studies where the indication to obtain CT or MRI was a non-diagnostic ultrasound exam. Nevertheless, sensitivity and specificity for CT in the included studies were similar to previously published values of 94% and 95%, respectively.[11] Likewise, MRI studies included had similar accuracy to prior reports (sensitivity .97, specificity .96).[21]

Table 1 - Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) for Appendicitis Diagnosis

Study	**Class**	**Age Group**	**Prevalence (n total)**	**Imaging Protocol Features of Note**	**Sensitivity (%)**	**Specificity (%)**	**LR Positive**	**LR Negative**
CT								
Abo 2011[22]	III	Pediatric	43% (128)	29 did not receive US. 99 had US and CT, with CT performed second in the majority of cases.	96 (86-99)	97 (90-100)	35.2 (9-138)	0.04 (0.01-0.15)
Eng 2018[15]	III	Pediatric	26% (1498)	Meta-analysis, includes Kaiser 2002	96 (93-98)	95 (93-96)	18 (14-23)	0.04 (0.02-0.07)
Kaiser 2002[23]	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 2018[15]	III	Adult	29% (1027)	Meta-analysis	90 (85-93)	94 (91-95)	14 (11-18)	0.11 (0.08-0.15)
Repplinger[21] 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 = 2851, 4 Studies. Eng 2018 includes Kaiser 2002)					94	94	16.7	0.06
MRI								
Orth 2014[24]	II	Pediatric	37% (81)	All patients had MRI and US, with blinded interpretations	93 (78-99)	94 (84-99)	15 (5.2-46)	0.07 (0.02-0.28)
Thieme 2014[20]	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 2018[15]	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 2018[15]	III	Adult	52% (223)	NR	90 (85-94)	94 (91-96)	15 (7.1-30)	0.04 (0.01-0.10)
Repplinger 2018[21]	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 = 789, 4 Studies. Eng 2018 includes Thieme 2014)					95	92	11.6	0.06

NR = Not Reported, LR = Likelihood Ratio

US Diagnostic Accuracy Overall

Table 2 summarizes test characteristics for US studies. 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[24] and one class III study.[17] In the former, non-visualization of the appendix with inflammatory signs was considered positive; in the latter, positive studies were subclassified by certainty of interpretation (probable vs. equivocal). Nine pediatric studies showed a positive LR ≥10. Those pediatric studies with a positive LR<10 included one small class II study,[20] one class III meta-analysis which exclusively studied point-of-care US (POCUS),[11] and a small class III POCUS study within that same meta-analysis.[16] A recent class III meta-analysis including 548 pediatric patients[15] showed test characteristics similar to CT and MRI (sensitivity .91, specificity .95, positive LR 18, negative LR 0.09).

Only 2 class III studies reported results on US for suspected acute appendicitis in adults.[15,16] Both had reasonably strong specificities (92%,[16] 95%[15]) and positive LRs (7.2,[16] 17[15]), comparable to CT and MRI. Neither had comparable sensitivity (Table 2) to CT or MRI (Table 1). The dearth of adult studies prevents strong recommendations regarding US in this patient population, but the two class III studies available would at least suggest a positive ultrasound in adults may be similarly interpreted to a positive result in children.

Table 2 - Ultrasound for Appendicitis Diagnosis

Study	**Class**	**Prevalence (n total)**	**Non-Diagnostic US %**	**How Were Non-Diagnostic (ND) Exams Considered?**	**Sensitivity (%)**	**Specificity (%)**	**LR Positive**	**LR Negative**
				Pediatric Ultrasound				
Orth 2014[24]	II	37% (81)	NR	Non-visualized, 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 2014[20]	II	54% (104)	42%	ND = Negative	76 (63-86)	89 (76-96)	6.9 (3.1-16)	0.27 (0.17-0.43)
Abo 2011[22]	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 2017[11] *Fox 2008[16]* *Sivitz 2014[18]*	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 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 2018[15]	III	27% (548)	NR	NR	91 (84-96)	95 (92-97)	18 (12-28)	0.09 (0.06-0.16)
Mittal 2013[25] *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 2015[19] *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 2018[26]	III	NR (766)	10%	ND = negative	69 (NR)	94 (NR)	11.5 (NR)	0.33 (NR)
van Atta 2015[17] *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 2002[23]	III	41% (600)	NR	ND results not allowed. Radiologist must report positive or negative only, even if confidence in diagnosis was low or appendix non-visualized.	80 (75-85)	94 (91-96)	13 (8.8-20)	0.21 (0.17-0.27)
Adult Ultrasound								
Fox 2008[16]	III	NR (83)	NR	ND = negative	59 (42-74)	92 (81-97)	7.2 (2.7-19.2)	0.64 (NR)
Eng 2018[15]	III	80% (169)	NR	NR	83 (70-91)	95 (92-97)	17 (3.8-72)	0.18 (0.12-0.26)

NR = Not Reported, LR = Likelihood Ratio

Equivocal exams

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

Table 3 - Comparison of Pediatric US Test Characteristics by Method of Counting Non-Diagnostic Exams

How Were Non-Diagnostic (ND) Exams 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	2362	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	2202	85%	93%	12.2	0.16
Any Method	9 Studies‡ - 7 Class III - 2 Class II	4187	78%	95%	14.4	0.23

*Mittal 2013[25] reported 2 analyses: Non-diagnostic (ND) as negative, and ND exams excluded.
†van Atta 2015[17] reported 2 analyses: Non-diagnostic (ND) as “likely positive” or “likely negative”, and ND exams excluded.

‡Studies included in Eng 2018[16] or Bennabas 2017[11] are only counted once, as part of each meta-analysis. Eng 2018[11] includes Schuh 2015.[19] Bennabas 2017[11] includes Fox 2008[16] and Sivitz 2014.[18]

<u>LR = Likelihood Ratio</u>

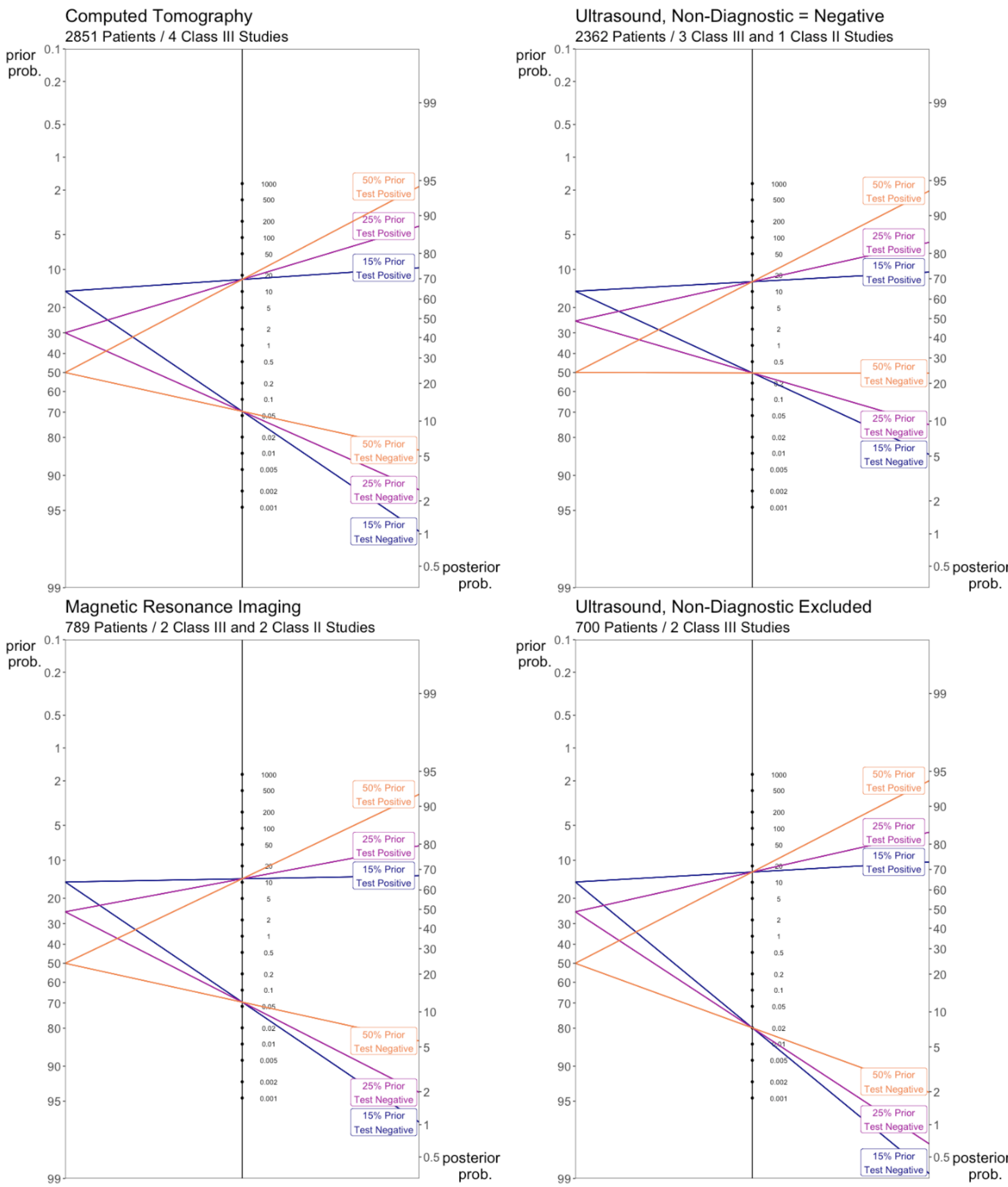
The most common way included studies treated ND exams was to count anything other than an unequivocally positive study (a dilated appendix which is completely visualized) as a negative (4 studies, 2362 patients). In this methodology exams resulting in non-visualization of the appendix, partial visualization with or without dilation, and non-dilated appendices with secondary signs (e.g. inflammation) were counted the same as an unequivocally-negative exam (complete visualization of a non-dilated appendix without any secondary signs of acute appendicitis). Five studies did not report how the ND were counted or utilized other methods in reporting ND results. Specificity and positive LR remained high regardless of the handling of ND exams (Table 3). This likely reflects the fact that counting any ND exam as negative was a particularly common practice and strengthens the confidence in the value of a positive US result.

<u>US, CT, and MRI by Pretest vs. Post-test Probability</u>

When ordering an imaging test for appendicitis, providers often have some estimation of the risk for the diagnosis. Figure 1 demonstrates post-test probability for each of the 3 modalities (US, CT, 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 LR and negative LR. US was divided into those studies reporting ND exams as negative, and those excluding ND exams. In general, regardless of the reporting of ND exams, post-test probability after a positive US was similar to probability after positive CT or MRI, at any pretest probability. Post-test probabilities after a negative CT or MRI, or an unequivocally-negative US, were similarly low for pretest probabilities of 15% and 30%. At a high pretest probability of 50%, post-test probability after negative CT or MRI approaches 5%, and 2-3% by an unequivocally-negative US. By contrast, among studies considering a ND US as “negative,” a negative result yielded a >5% post-test probability for acute appendicitis even when pretest probability was low (15%). Therefore independent of a clinicians pre-test probability the results of unequivocally negative US are comparable to CT or MRI.

One class III study[11] derived test-treatment thresholds for pediatric acute appendicitis based on published complication rates of appendectomy and risk of ionizing radiation from CT or MRI (i.e. zero in the latter). They calculated that a test with positive LR ≥ 5.8 would meet the treatment threshold for ruling in acute appendicitis without further testing, and negative LR ≤0.03 for ruling out acute appendicitis. Every class II or III US study except one[16] showed a positive LR >5.8, in both adults and children. The lone study with positive LR<5.7 was included in another class III study as part of a meta-analysis,[11] for which the overall positive LR was 9.2. Both of the US studies excluding ND exams had negative LR < 0.03 (Tables 2 and 3). One additional class III US study involving a re-evaluations pathway in the case of ND exam showed a negative LR of 0.03.[19] All other US studies, 3 of 5 CT studies, and 3 of 5 MRI studies had a negative LR > 0.03.

__Figure 1 -__ Fagan nomograms for various acute appendicitis imagin strategies at low (15%), moderate (30%), and higl pretest probability



<u>Reevaluation and serial examination after non-diagnostic US</u>

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

<u>Summary</u>

US is useful for ruling in acute appendicitis, and when positive is typically the only test needed prior to surgical consultation. This fact, along with its lack of ionizing radiation, as well as likely broader availability for most emergency providers compared to MRI, should make it the initial first test of choice for pediatric patients. While 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 US is a large amount of ND results, the rate of which varies widely between studies and settings. Negative predictive performance of US varies far more than MRI or CT but in pediatric patients, this variation in performance appears closely related with whether or not ND exams are counted as negative or excluded. An unequivocal negative US (visualization of a compressible tubular structure from tip to cecum <6 mm in diameter without secondary signs of inflammation) in a pediatric patient may be comparable to a negative CT or MRI based on low certainty of evidence (3 class III studies). For non-diagnostic US exams in children, a strategy of watchful waiting including clinical reevaluation, surgical consultation, hospital observation, and/or serial US exam may be a reasonable alternative to immediate MRI or CT. Shared decision making of the relative risks and benefit is, as well as an assessment of local resources (e.g. rapid MRI availability), is likely reasonable to guide such a decision.

<u>Future Research</u>

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

3. In emergency department 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.

Level B recommendations. In adult and pediatric ED patients undergoing CT for suspected acute appendicitis, use IV 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, non-contrast CT scans may be used for the evaluation of acute appendicitis with minimal reduction in sensitivity.

<u>Potential Benefit of Implementing the Recommendations:</u>

- 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.

<u>Potential Harm of Implementing the Recommendations:</u> (Sokolove, Dierks)

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

<u>Key words/phrases for literature searches:</u>
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 10-11, 2020.

Study Selection: Two hundred and twenty articles were identified in searches. Nine articles were selected from the search results as potentially addressing this question and were candidates for further review. After grading for methodological rigor with zero Class I studies, 1 Class II study, and 8 Class III studies included for this critical question.

Summary

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

Background

Computed tomography of the abdomen and pelvis (CTAP) imaging is frequently used to in the evaluation of patients with suspected appendicitis. Radiology protocols for CTAP often include the use of enteric or intravenous (IV) 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, in particularly in those patients with low body mass. In addition, this prior policy suggested that IV 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 (e.g., increased use of multi-row detector CT and reduced slice width) resulting in improved image quality. This may impact the diagnostic advantage of enteric of IV contrast previously identified. The 2018 American College of Radiology Appropriateness Criteria for adult and children[27] reports that CT abdomen and pelvis with IV contrast or without IV contrast may both be appropriate, further highlighting the uncertainty in this area. However, this document does not comment on the use of enteric contrast.[27] With this critical question, we set out to review the recent literature on the role of contrast in the evaluation for appendicitis.

In 2012, a Class II study by Kepner et al,[28] 227 adult patients were randomized to receive IV contrast or oral contrast. Imaging was done 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 discharge patients were followed by phone call. A total of 80 patients has a CT diagnosis of appendicitis. The authors report that for IV contrast alone the sensitivity was 100% (95% CI 89.3-100) and specificity was 98.6 (95% CI 91.6-99) resulting in a positive LR of 72 (CI 10.3-504) and negative LR 0.00. For IV and oral contrast the sensitivity was 100% (95% CI 87.4-100), specificity 94.9 (95% CI 86.9-98.4), positive LR of 25 (95%CI 8.24-75.8). There was no statistically significant difference between the use of IV and IV with oral contrast leading the authors to report that there was similar diagnostic performance. One difference that was noted, however, was that patients receiving IV contrast alone were discharged faster. Two other Class III studies directly evaluated the role of contrast. Anderson et al,[29] using a 64-slice MDCT on a convenience sample of 303 adult patients, and Keyzer et al,[30] using a 4 slide MDCT in 131 adult patients. Both studies showed not difference in diagnostic accuracy, with the former demonstrating a positive LR of 34 (CI 13.04-89.9) and negative LR 0.00 for IV and positive LR 35 (95%CI 13.3-91.9) with negative LR 0.00 for IV and oral contrast. In another Class III study by Jacobs et al,[31] 228 patients with suspected appendicitis underwent both a focused CT of the right lower quadrant with oral contrast and a CT with both oral and IV contrast. They reported that the sensitivity of oral contrast was 0.76 and specificity 0.94 and for both the oral and IV contrast the sensitivity was 0.91 and specificity 0.95. Specific to pediatric patients, a 2018 Class III study by Farrell et al[32] retrospectively compared pediatric cohorts receiving IV contrast alone versus oral contrast. A total of 558 64-MDCT scans met inclusion criteria. Appendicitis was diagnosed in 22.4% of patients. The authors reported similar sensitivities of 93.8% (95%CI 84.8-98.3) and 94.6% (95% CI 84.9-98.9) and specificities of 98.5% (%CI 95.8-99.7) and 98.3% (95% CI 95.7-99.5) 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 non-contrast CT diagnostic accuracy. A Class III meta-analysis by Hlibczuk et al[33] included 7 studies with adult patients who had non-contrast CT for the evaluation of appendicitis. He reported a pooled sensitivity of 92.7% (95% CI 89.5-95%) and specificity of 96.1% (95% CI 94.2-97.5%). In another Class III meta-analysis, Rud et al[34] reported the pooled sensitivities for unenhanced CT 91% (95% CI 87-93%), oral contrast only 89% (95% CI 81-94%), IV contrast 96% (95% CI 92-98), IV with oral contrast 96% (95% CI 93-98), and rectal contrast 96% (95% CI 92-98). There were no differences in pooled specificity estimates. Both of these meta-analyses included studies that were low quality, included high risk of bias, and had high prevalence of appendicitis. In a Class III study, Seo et al[35] reported no difference in the sensitivity and specificity between low radiation dose non-contrast CT and standard radiation dose IV contrast CT in a 200-patient study. This study is limited by the confounder of different radiation doses. Chiu et al[36] evaluated the sensitivity of non-contrast CT to IV contrast CT in 100 patients with suspected appendicitis. In this cohort, with 44/100 patients diagnosed with appendicitis, he reported non-contrast CT had a lower sensitivity than IV contrast CT (91% versus 100%, p=0.04), and greater specificity (100% versus 95%, p=0.04) for the diagnosis of appendicitis. In Class X study by Hershko et al,[37] 232 adult patients with suspected appendicitis were randomized to receive a non-contrast, rectal contrast, or dual contrast (oral and IV). The noted positive LR of 8.9, 12.3, 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,[38] 293 patients >16 yo with abdominal pain underwent non-contrast enhanced imaging using a 16-MDCT. They noted a sensitivity of 70.1% and specificity of 76.0% for a correct diagnosis in a non-contrast CT. It is important to note that the non-contrast studies have included only adult patients.

<u>Future Research</u>

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

REFERENCES

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

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(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(2):333-338.

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

5. Howell JM, Edyy 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, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi: 10.1136/bmj.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, Pershad J. Prospective evaluation of a clinical pathway for suspected appendicitis. *Pediatrics*. 2014;133(1):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(10):1060-1065.

10. Mandeville K, Pottker T, Bulloch B, Liu J. Using appendicitis scores in the pediatric ED. *Am J Emerg Med.* 2011;29(9):972-977.

11. 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.

12. 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.

13. Koberlein GC, Trout AT, Rigsby CK, et al. ACR Appropriateness Criteria® Suspected Appendicitis-Child. *J Am Coll Radiol.* 2019; 16: S252–S263.

14. Atwood R, Blair S, Fisk M,et al. (2021). NSQIP based predictors of false negative and indeterminate ultrasounds in adults with appendicitis. *J Surg Res*. 2021;261:326-333. https://doi.org/10.1016/j.jss.2020.10.021

15. 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(3):717-727.

16. Fox JC, Solley M, Anderson CL, Zlidenny A, Lahham S, Maasumi K. Prospective evaluation of emergency physician performed bedside ultrasound to detect acute appendicitis. *Eur J Emerg Med*. 2008;15(2):80-85.

17. 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.

18. Sivitz AB, Cohen SG, Tejani C. Evaluation of acute appendicitis by pediatric emergency physician sonography. *Ann Emerg Med*. 2014;64(4):358-364.e4.

19. 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.

20. 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.

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

22. 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.

23. Kaiser S, Frenckner B, Jorulf HK. Suspected appendicitis in children: US and CT--a prospective randomized study. *Radiol.* 2002;223(3):633-638.

24. 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.

25. 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.

26 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.

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

28. 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.

29. 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.

30. 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.

31. 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(3):683-690.

32. Farrell CR, Bezinque AD, Tucker JM, Michiels EA, Betz BW. Acute appendicitis in childhood: oral contrast does not improve CT diagnosis. *Emerg Radiol.* 2018;25(3):257-263.

33. 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.

34. Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jørgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev.* 2019;2019(11):CD009977. Published 2019 Nov 19.

35. 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(1):96-105.

36. 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(1):73-78.

37. 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.

38. 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.

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).

Evidentiary Table.

Study & Year Published	Class of Evidence	Setting & Study Design	Methods & Outcome Measures	Results	Limitations & Comments
Gonzalez Del Castillo et al (2016)[7]	III for Q1	Prospective cohort study at 4 academic medical centers in Spain from June to December 2014	Pediatric patients (2-20 years 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 two 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), negative LR 0.3 (95% CI 0.2-0.5)	All patients had appendectomy or telephone follow-up
Saucier et al (2014)[8]	III for Q1	Prospective cohort study at a single academic urban hospital	Pediatric (3-17 years of age) with suspected appendicitis; Pediatric Appendicitis Score calculated by treating provider and incorporated into clinical pathway; Criterion standard was surgical pathology and one-day telephone follow-up	N = 196 patients with appendicitis prevalence of 33%; PPV for 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 zero; 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)	PAS guided imaging and consultation decisions, which may cause workup bias; limited telephone follow-up

Fleischman et al (2013)[9]	III for Q1	Prospective cohort in a single academic center	Children (3-18 years of age) with suspected appendicitis; patients categorized as low, intermediate or high risk according to previously derived score; physician judgement stratified patients as very low, low, intermediate, or high risk; criterion standard was surgical pathology, chart review, and two-week telephone follow-up	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 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)	Small sample size
Mandeville et al (2011)[10]	III for Q1	Prospective cohort; single center, urban, academic center	Children (4-17 years 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 two-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% CI 0.69-0.82) and specificity 0.72 (95% CI 0.65-0.80); AUC 0.77 (95% CI 0.72-0.83)	High prevalence of appendicitis may result in spectrum bias
Abo et al (2011)[22]	III for Q2	Prospective cohort; single center, urban, academic center	Children (2-20 years) with suspected appendicitis; US and CT at discretion of treating providers;	N = 176 with appendicitis prevalence of 42%; 147 patients had US, 128 had CT, and 99 had both.	Imaging interpretation not blinded to clinical data; CT generally used as second-line test

			interpretation by treating radiologist; appendicitis diagnosis determined by surgical pathology, chart review and 1-week phone follow-up	If non-diagnostic 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)	
Benabbas et al (2017)[11]	III for Q2	Meta-analysis of prospective studies	Included studies of pediatric (<21 years) ED patients with suspected appendicitis; Random effects models to estimate pooled test characteristics	ED POCUS (N=4 studies): Pooled sensitivity 0.86 (95% CI 0.79–0.91), specificity 0.91 (95% CI 0.87–0.94), positive LR 9.2 (95% CI 6.4–13), negative LR 0.17 (95% CI 0.09–0.30)	Most studies at high risk of differential verification bias
Eng et al (2018)[15]	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 non-diagnostic ultrasound; quality assessed by three 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);	Unclear how these results apply to first-line imaging choice.

				Adult CT: sensitivity 0.90 (95% CI: 0.85-0.93), specificity 0.94 (95% CI 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 (2013)[25]	III for Q2	Retrospective cohort study of multicenter, academic center	Children (3-18 years) 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 phone follow-up.	N = 2635 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 (2014)[24]	II for Q2	Prospective cohort study in single academic center	Pediatric (3- 17 years) 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 (2018)[21]	III for Q2	Prospective cohort study in single academic center	Pediatric (> 12 years) 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 three fellowship-trained abdominal radiologists blinded to clinical data; Appendicitis diagnosis determined by surgical pathology, chart review, and one-month phone follow-up	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 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.	1224 of 1551 eligible patients were not included.
Schuh et al (2015)[19]	III for Q2	Prospective cohort study in single academic center	Pediatric (4-17 years) 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 equivocal, an additional interval US was performed at discretion of providers; appendicitis diagnosis determined by surgical pathology, chart	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 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),	

			review, and 1-month phone follow-up	specificity 0.96 (95% CI 0.76-1.0), and 0.43 (95% CI 0.27-0.59) equivocal rate.	
Sola et al (2018)[26]	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; confidence intervals (or raw data) for sensitivity and specificity were not provided.
Thieme et al (2014)[20]	II for Q2	Prospective cohort study in single academic center	Pediatric (4-18 years) ED patients with suspected appendicitis; patients received US and MR within 2h; appendicitis diagnosis by review of hospital and outpatient medical records	N = 104 with appendicitis prevalence 56%. 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).	Small study with high prevalence of appendicitis.
van Atta et al (2015)[17]	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 (2008)[16]	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 determined by surgical pathology, chart review and	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.

			phone follow-up 2 weeks – 3 months.		
Kaiser et al (2002)[23]	III for Q2	Prospective randomized clinical trial in single academic center	Patients at pediatric ED randomized to US v. 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 (2014)[18]	III for Q2	Prospective cohort study in single academic center	Pediatric patients with suspected appendicitis; US performed by pediatric emergency medicine physicians; appendicitis diagnosis determined by surgical pathology, chart review and phone follow-up	N = 254. Among 231 analyzed patients, prevalence of appendicitis 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).	9% patients lost to follow-up. Some patients received more than one ultrasound.
Chiu et al (2013)[36]	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	N=100 with appendicitis prevalence of 44%. Non-contrast CT had lower sensitivity than contrast CT	Convenience sample with relatively high prevalence of appendicitis could result in spectrum bias.

			interpretated by two study radiologists blinded to clinical data and original interpretation; diagnosis of appendicitis by pathology and 6-month chart review	(91% v. 100%, p=0.04) and greater specificity (100% v 95%, p=0.04)	
Anderson et al (2009)[29]	III for Q3	Randomized controlled trial in single academic center	Adults with acute abdominal pain randomized to CT with oral and IV contrast v. 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 two groups. Confidence not associated with BMI or intra-abdominal fat.	Study did not assess differences in sensitivity and specificity with the addition of oral contrast.
Kepner et al (2012)[28]	II for Q3	Randomized controlled trial in single academic center	Adults with suspected appendicitis randomized to CT with oral and IV contrast v. CT with IV contrast and no oral contrast; interpretation by 2 independent study radiologists blinded to original interpretation and clinical data; diagnosis of appendicitis by pathology, chart review and telephone follow-up	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 92-100%); IV and oral contrast: sensitivity 100% (95% CI 87-100%), specificity 95% (95% CI 87-98%)	CTs were interpretated study radiologists. Contemporaneous CT interpretation influenced clinical management and outcome assessment (workup bias) 16-slice CT scanner.
Keyzer et al (2009)[30]	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,	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	CTs were interpretated study radiologists. Small sample size. Contemporaneous CT interpretation influenced clinical management and

			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	either radiologist comparing 4 types of CT scans.	outcome assessment (workup bias) 4-slice CT scanner.
Seo et al (2009)[35]	III for Q3	Retrospective cohort in single academic center	Adult (≥15 years) patients with suspected appendicitis 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	N = 207 with appendicitis prevalence 34%; sensitivity and specificity were not significantly different for either radiologist comparing 2 types of CT scans.	Small sample size. Unable to separate potential effects of radiation dose and IV contrast.
Hlibczuk et al (2010)[33]	III for Q3	Meta-analysis of prospective and retrospective studies	Included studies of non-contrast CT for evaluation of appendicitis in adult (≥16 years), ED patients with at least two 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 (2019)[34]	III for Q3	Meta-analysis of prospective and retrospective studies	Included ED and non-ED based studies of CT for evaluation of appendicitis in adult (≥14 years) patients; random effects model to estimate pooled sensitivity and specificity for different types of contrast (oral, rectal and IV)	N = 64 studies included with median appendicitis 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).	Only 2/64 studies were assessed as low risk of bias in all four domains; relatively high prevalence of appendicitis; no study was considered high quality with differential verification a common threat to bias.

				Pooled specificity estimates were similar for different types of contrast, with point estimates ranging from 93-95%.	
Farrell et al (2018)[32]	III for Q3	Retrospective cohort study in single urban, academic center	Pediatric (0–17 years) ED patients with acute, non-traumatic abdominal pain who received CT with IV contrast. CT protocol changed from addition of oral contrast to non-contrast halfway during study period; surgical pathology and chart review for follow-up	N = 588 with appendicitis prevalence 22%. 270 patients in oral contrast group and 288 in non-contrast group; oral contrast (N=270): sensitivity 0.94 (95% CI 0.85–0.98) and specificity 0.99 (95% CI 0.96–1.0); non-contrast (N=288): sensitivity 0.95 (95% CI 0.85–0.99) and specificity 0.98 (95% CI 0.96-1.0).	No active follow-up and attrition not reported.
Jacobs et al (2001)[31]	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 and/or chart review	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.