# Ultrasound vs. Computed Tomography for Diagnosing Appendicitis: A Systematic Review

**Qaed Salem Alhamami[1], Arwa Hussain Al-alhareth[2], Abeer Hadi J Alkhadrh[3], Alhanouf Mofareh H Almakhalas[4], Shooq Mnbat Alshahrani[5], Amjad Mohammed Majrashi[6], Ghzlan Hadi Al-Ajbi[7], Taif Mahdi Abuhasirah[8]**

1 Associate Professor, Department of Internal Medicine, College of Medicine, Najran University, Najran, Saudi Arabia – Email: gsalhammami@nu.edu.sa
2 Medical Intern, Najran University, Najran, Saudi Arabia – Email: Arwahuss77@gmail.com
3 Medical Intern, Najran University, Najran City, Saudi Arabia - Email: Abeeralkhadrh@gmail.com
4 Medical Intern, Najran University, Najran City, Saudi Arabia - Email: alhanouf11.m@hotmail.com
5 Medical Intern, Najran University, Najran City, Saudi Arabia - Email: shooqalrasasmeh@gmail.com
6 Medical Intern, Najran University, Najran City, Saudi Arabia - Email: am.7crq@gmail.com
7 Medical Student, Najran University, Najran City, Saudi Arabia - Email: ghzlanaljby1@gmail.com
8 Medical Student, Najran University, Najran City, Saudi Arabia - Email: taif.5575@gmail.com

**Corresponding author: Arwa Hussain Al-alhareth**, **Email:** Arwahuss77@gmail.com

## ABSTRACT

**Background:** Acute appendicitis is the most common surgical emergency, yet clinical diagnosis remains challenging. Ultrasound (US) and computed tomography (CT) are primary imaging modalities, but optimal strategies are debated, particularly in equivocal adult cases.

**Objective:** To synthesise contemporary evidence comparing the diagnostic accuracy of US and CT for acute appendicitis, with histopathology as the reference standard.

**Methods:** Systematic review of PubMed, Scopus, Web of Science, and Cochrane CENTRAL (2021–2026). Included studies enrolled suspected appendicitis patients, evaluated both modalities, reported diagnostic metrics, and used histopathology. Quality assessed with QUADAS-2; narrative synthesis performed.

**Results:** Eight studies comprising 4,608 participants were included. CT demonstrated superior accuracy, with sensitivity 88–99.4% and specificity 71–80%, versus US sensitivity 40–85.7% and specificity 50–62.5%. CT was significantly better for overweight, obese, and elderly patients ($p<0.05$). Combining clinical scores with US achieved comparable performance in selected populations, with negative appendectomy rates as low as 2.9%. CT added limited value for high clinical scores (≥7) but was beneficial in equivocal cases (scores 4–6). **Conclusion:** CT offers superior accuracy overall, but a risk-stratified approach is optimal. US with clinical scoring is effective first-line, with CT reserved for inconclusive or complex cases. Standardised US protocols and operator training are needed to optimise pathways.

**Keywords:** Acute appendicitis; Ultrasound; Computed tomography; Diagnostic accuracy; Systematic review; Sensitivity; Specificity.

## INTRODUCTION

Acute appendicitis is the most common surgical emergency worldwide, with a lifetime risk of 7–12% and peak incidence in the second and third decades of life [1]. Despite advances in diagnostic imaging, the clinical diagnosis remains challenging, particularly in atypical presentations, contributing to negative appendectomy rates of 10–30% [2]. The diagnostic dilemma arises from the need to accurately identify patients requiring surgery while avoiding unnecessary operations and their associated morbidity [3].Clinical evaluation alone, including history, physical examination, and laboratory findings, has demonstrated limited diagnostic accuracy [4]. The Alvarado score, introduced in 1986, and the more recent Appendicitis Inflammatory Response (AIR) score have improved clinical risk stratification but remain insufficient for definitive diagnosis without imaging confirmation [5]. This has driven the widespread adoption of cross-sectional imaging to enhance preoperative diagnostic accuracy [6].

Ultrasound (US) is recommended as the first-line imaging modality, particularly in children and pregnant women, due to its non-invasive nature, low cost, and absence of ionizing radiation [7].

However, its diagnostic performance is operator-dependent and limited by patient body habitus, bowel gas, and anatomical variations, resulting in frequent equivocal or non-diagnostic findings [8]. Conversely, computed tomography (CT) offers superior spatial resolution and diagnostic accuracy, with reported sensitivity and specificity exceeding 90% in numerous meta-analyses [9]. A recent systematic review reported pooled CT sensitivity of 97.2% and specificity of 95.6%, compared to 82.1% and 85.9% for US, respectively [3]. Despite this superiority, CT use is constrained by radiation exposure concerns, higher costs, and reduced accessibility in resource-limited settings [3].

Recent evidence suggests that combining clinical scoring systems with ultrasound may achieve comparable diagnostic accuracy to CT in selected patient populations,



Received: 14/04/2026
Accepted: 16/05/2026

potentially reducing unnecessary radiation exposure [7]. However, controversy persists regarding the optimal imaging strategy, particularly in adult patients with equivocal clinical presentations. This systematic review aims to synthesise contemporary evidence comparing the diagnostic accuracy of ultrasound and computed tomography for acute appendicitis, with histopathological examination as the reference standard, to provide evidence-based guidance for clinical practice.

## Methodology

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [10]. The research question was formulated using the PICO framework, addressing patients with suspected acute appendicitis (P), ultrasound as index test (I), computed tomography as comparator (C), and diagnostic accuracy with histopathology as reference standard (O).

### Eligibility Criteria

Studies were included if they: (1) enrolled patients with clinically suspected acute appendicitis; (2) evaluated both ultrasound and computed tomography; (3) reported diagnostic accuracy metrics (sensitivity, specificity, predictive values, or accuracy); and (4) used histopathological examination as the reference standard.

Only original primary research (prospective or retrospective cohort studies, cross-sectional studies, or randomized controlled trials) published between 2021 and 2026 were considered.

**Exclusion criteria included:** review articles, case reports, conference abstracts, editorials, studies without direct comparison of both modalities, studies of appendiceal neoplasms, and studies lacking sufficient data to construct two-by-two contingency tables.

### Information Sources and Search Strategy

A comprehensive literature search was conducted across PubMed/MEDLINE, Scopus, Web of Science, and Cochrane CENTRAL for publications from 2021 to 2026. The search strategy combined controlled vocabulary and free-text terms for: ("acute appendicitis" OR "appendicitis") AND ("ultrasound" OR "sonography" OR "US") AND ("computed tomography" OR "CT" OR "computerized tomography") AND ("diagnostic accuracy" OR "sensitivity" OR "specificity"). Reference lists of included studies and relevant reviews were manually screened. No language restrictions were applied.

### Study Selection Process

All identified records were exported to Rayyan, a web-based systematic review software, for deduplication and screening [11]. Two independent reviewers (D.A. and M.A.) performed title/abstract screening followed by full-text review against eligibility criteria. Disagreements were resolved through discussion or consultation with a third reviewer. The selection process was documented using a PRISMA flow diagram, with exclusion reasons recorded at the full-text stage.

### Data Collection and Data Extraction

A standardized data extraction form was developed in Microsoft Excel and pilot-tested on two studies. Two reviewers independently extracted data, including: bibliographic information, study characteristics (design, setting, sample size), participant demographics, imaging protocols, reference standard details, and diagnostic accuracy outcomes (sensitivity, specificity, predictive values, accuracy, and negative appendectomy rates). Two-by-two contingency tables were constructed where data permitted. Discrepancies were resolved by consensus, and missing data were recorded as "not recorded" (NR).

### Risk of Bias and Quality Assessment

Methodological quality was independently assessed by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [12]. This instrument evaluates risk of bias across four domains: patient selection, index test, reference standard, and flow and timing. Each domain was rated as "low," "high," or "unclear" risk of bias. Applicability concerns were assessed for the first three domains. Disagreements were resolved through discussion.

### Data Synthesis

Due to substantial clinical and methodological heterogeneity across studies, a narrative synthesis was conducted rather than meta-analysis. Diagnostic accuracy metrics were summarized descriptively in structured tables. Sensitivity, specificity, positive and negative predictive values, and overall accuracy were compared between ultrasound and CT across studies. Subgroup analyses were performed where data permitted (e.g., by body mass index, age, and appendicitis severity). Negative appendectomy rates were extracted as an important clinical outcome measure.

## RESULTS

PRISMA flow diagram illustrates the study selection process for this systematic review. An initial database search identified 402 records, of which 187 duplicates were removed, leaving 215 records for screening. Following title and abstract screening, 96 records were excluded, and 119 reports were sought for retrieval. Of these, 62 reports could not be retrieved, leaving 57 full-text reports assessed for eligibility. At the full-text stage, 49 reports were excluded for the following reasons: wrong outcome (n=20), wrong population (n=17), and abstract-only publications (n=12). Ultimately, 8 studies met all eligibility criteria and were included in the final systematic review.

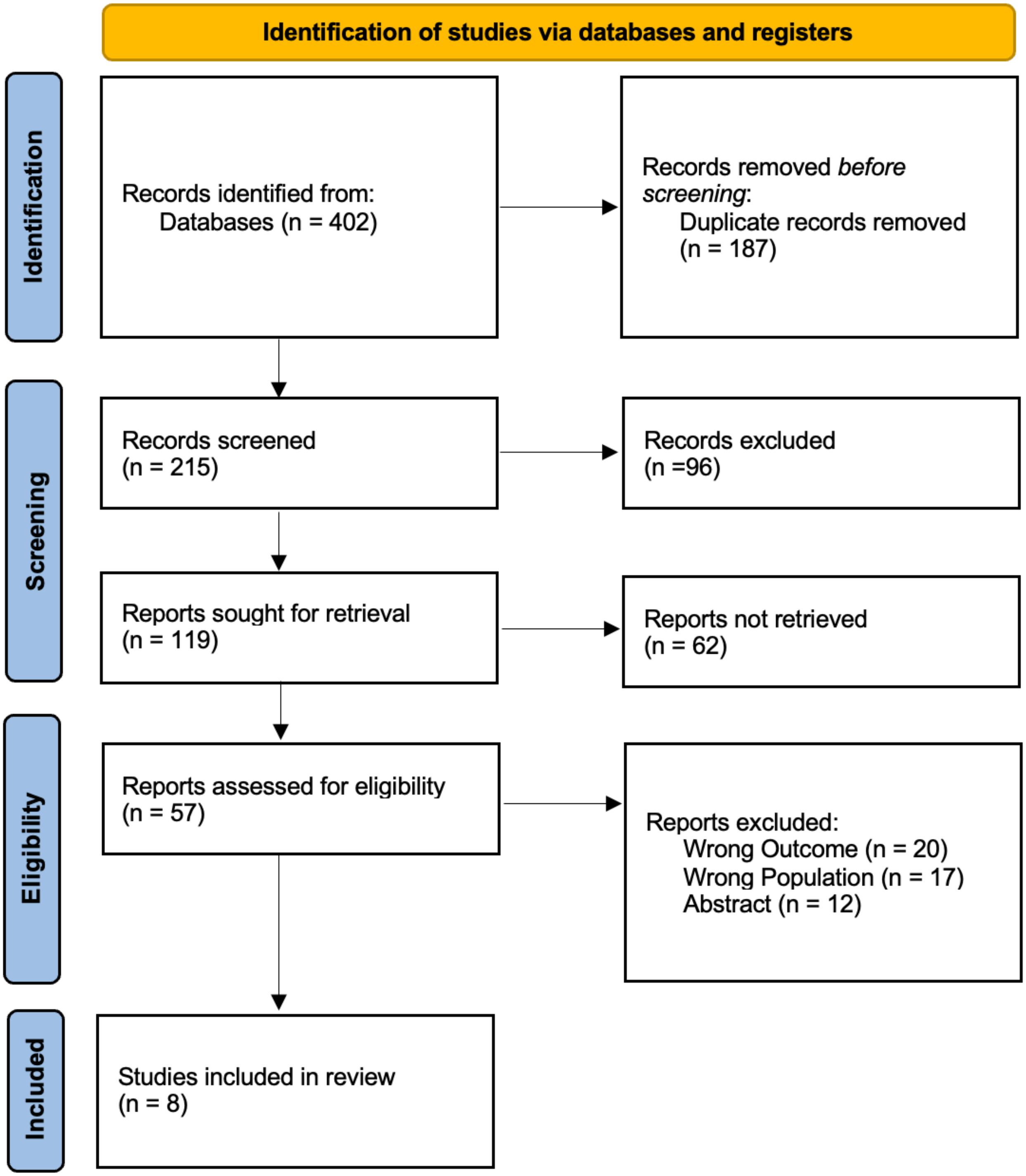


**Figure 1:** PRISMA Flow Diagram of Study Selection.

Table 1 summarizes the methodological and demographic characteristics of the eight included studies [13-20]. The studies included four retrospective cohort studies [14–18], one prospective study [13], one randomized controlled trial [19], and one international multicenter prospective observational study [20]. Sample sizes ranged from 41 patients in the single-center study by **Zhao *et al.*** [18] to 3,123 patients in the multicenter investigation by **Coccolini *et al.*** [20].

Geographic settings varied, with studies conducted in Belgium [15], Saudi Arabia [16], and multiple international centers [20], while the locations of others were not explicitly stated [13,14,17-19]. The study populations predominantly comprised adults with suspected acute appendicitis, with mean ages ranging from 24 to 38 years as reported by **Coccolini *et al.*** [20]. Notably, **Anan *et al.*** [13] uniquely integrated the Alvarado score with focused ultrasound as a double-check triage tool before CT, while **Lu *et al.*** [17] provided valuable subgroup analyses based on body mass index and age.

Table 2 presents the core diagnostic accuracy metrics and primary outcomes. Across all studies, CT consistently demonstrated superior sensitivity, specificity, and overall accuracy compared to ultrasound when measured against histopathology as the reference standard.

CT sensitivity ranged from 88% in **Zhao *et al.*** [18] to 99.4% in **Haentjens *et al.*** [15], whereas ultrasound sensitivity varied widely from 40% in **Ahmed *et al.*** [14] to 85.7% in **Raffa *et al.*** [16]. Similarly, CT specificity (71–80%) consistently exceeded that of ultrasound (50–62.5%).

The diagnostic accuracy of CT reached 100% in **Anan *et al.*** [13], while ultrasound accuracy peaked at 77.78% in **Raffa *et al.*** [16]. The negative appendectomy rate (NAR) was reported in four studies, with the lowest rates observed in the ultrasound group (2.9%) of **Coccolini *et al.*** [20] and the CT group (4.5%) of **Haentjens *et al.*** [15].

Key subgroup findings revealed that CT provides significantly better diagnostic accuracy for overweight, obese, and elderly patients [17], and that ultrasound combined with clinical scores (Alvarado/AIR) outperforms CT or upfront surgery, particularly in doubtful or equivocal cases [19,20].

Table 3 presents the risk of bias assessment using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) tool, which is the Cochrane-recommended framework for evaluating methodological quality in diagnostic test accuracy reviews. The overall risk of bias was judged as unclear for most studies due to insufficient reporting of critical methodological details, particularly regarding whether patients were enrolled consecutively or randomly, whether the index tests (US and CT) were interpreted blindly without knowledge of the reference standard, and whether the time interval between imaging and histopathological confirmation was appropriate.

**Table 1: Characteristics of the Included Studies**

| Study (Author, Year [Ref.]) | Country / Setting | Study Design | Sample Size (N) | Study Population / Subgroups | Key Demographic / Clinical Notes |
|---|---|---|---|---|---|
| **Anan *et al.* (2025)** [13] | NR | Prospective | 164 (31 double-positive controls, 133 received CT) | Patients (both sexes) presenting with acute appendicitis | Assessed via Alvarado score + focused US; CT reserved for non-double-positive cases. |
| **Ahmed *et al.* (2022)** [14] | NR (UK implied) | Retrospective Audit | 60 (18 had US, 46 had CT, 4 had both) | Adult patients presenting to A&E with suspected appendicitis | CT met RCR standards (>90%); US failed the 70% sensitivity standard. |
| **Haentjens *et al.* (2023)** [15] | Sint-Andries Hospital, Tielt, Belgium | Retrospective | 304 (109 US, 153 CT, 42 CT after US) | Patients treated surgically for appendicitis (Mean age: 31 years) | Overall prevalence of appendicitis was 95.1%. CRP and LPC were significantly higher in the appendicitis group. |
| **Raffa *et al.* (2022)** [16] | King Abdulaziz University Hospital, Jeddah, Saudi Arabia | Retrospective | 351 (83 underwent appendectomy; 18 had US, 62 had CT) | Patients with a clinical picture of AA and no trauma history | Pathology revealed 64 AA cases and 19 normal appendixes. |
| **Lu *et al.* (2025)** [17] | NR (China implied) | Retrospective | 279 confirmed AA cases (64 simple, 127 suppurative, 88 gangrenous) | Patients diagnosed with AA who underwent emergency surgery | Subgroup analysis for normal BMI, overweight, obese, and elderly patients. |
| **Zhao *et al.* (2025)** [18] | NR | Retrospective | 41 (24 CT, 17 US, 7 had both) | Patients with suspected AA | CT: 4 misdiagnosed (17%) and 8 equivocal (33%); US: 6 misdiagnosed (35%) and 2 equivocal (12%). |
| **Noori *et al.* (2023)** [19] | NR (Iraq/Iran implied) | Randomized Controlled Trial (RCT) | 286 (211 underwent appendectomy; 192 had CT) | Adult patients with suspected acute appendicitis (equivocal cases included) | Prevalence of AA was 89.1% (188 patients); NAR was 10.9%. 165 simple AA, 23 perforated AA. |
| **Coccolini *et al.* (2024)** [20] | International Multicenter | Prospective Observational | 3123 (899 no radiology, 1490 US, 734 CT) | Patients operated for suspected AA with definitive histopathology | Sex ratio favored males ($p<0.001$). Mean age: 24 (no radiology), 28 (US), 38 (CT-scan) years ($p<0.001$). |

*NR: Not reported. AA: Acute Appendicitis; US: Ultrasound; CT: Computed Tomography; NAR: Negative Appendectomy Rate; RCR: Royal College of Radiologists; CRP: C-Reactive Protein; LPC: Leukocyte Particle Count.*

**Table 2: Diagnostic Performance and Primary Outcomes**

| Study (Author, Year [Ref.]) | Reference Standard | US Sensitivity / Specificity / Accuracy | CT Sensitivity / Specificity / Accuracy | Negative Appendectomy Rate (NAR) | Key Findings & Remarks |
|---|---|---|---|---|---|
| **Anan *et al.* (2025)** [13] | Histopathology | **Sens:** NR<br>**Spec:** NR<br>**Acc:** 96.4% (US+ / Alv-); 94.3% (US- / Alv+); 84.4% (double-negative). | **Sens:** NR<br>**Spec:** NR<br>**Acc:** 100% | NR | The double-check test (US + Alvarado) is a cheap positive tool. CT is highly accurate and should be preserved for patients negative on the double-check test. |
| **Ahmed *et al.* (2022)** [14] | Confirmed diagnosis (Pathology implied) | **Sens:** 40%<br>**Spec:** NR<br>**Acc:** NR | **Sens:** 93%<br>**Spec:** NR<br>**Acc:** NR | 2% | CT standards (>90%) were met, but US sensitivity (40%) was significantly below the recommended RCR standard (>70%). Poor operator dependency suggested. |
| **Haentjens *et al.* (2023)** [15] | Histopathology | **Sens:** 74.8%<br>**Spec:** 62.5%<br>**Acc:** NR | **Sens:** 99.4%<br>**Spec:** 80.0%<br>**Acc:** NR | US group: 4.67%<br>CT group: 4.50%<br>CT after US: 7.1% | CT shows superior sensitivity and specificity. Combining radiology, clinical examination, and lab values achieves low NAR. |
| **Raffa *et al.* (2022)** [16] | Histopathology | **Sens:** 85.71%<br>**Spec:** 50.00%<br>**Acc:** 77.78% | **Sens:** 92.00%<br>**Spec:** 75.00%<br>**Acc:** 88.71% | 22.9% (19/83 normal appendixes) | CT showed significant agreement with pathology (P<0.001), whereas ultrasound did not (P=0.130). |
| **Lu *et al.* (2025)** [17] | Histopathology | **Sens:** NR<br>**Spec:** NR<br>**Acc:** 74.9% (overall derived*) | **Sens:** NR<br>**Spec:** NR<br>**Acc:** 82.8% (overall derived*) | NR | No significant overall difference (P>0.05). However, CT provides significantly better diagnostic accuracy for overweight, obese, and elderly patients (P<0.05). |
| **Zhao *et al.* (2025)** [18] | NR (Surgery implied) | **Sens:** 73%<br>**Spec:** 50%<br>**Acc:** NR | **Sens:** 88%<br>**Spec:** 71%<br>**Acc:** NR | NR (stated as "still high") | CT outperforms US. Misdiagnosis rates: US 35%, CT 17%. Both still have considerable equivocal results (12% for US, 33% for CT). |
| **Noori *et al.* (2023)** [19] | Histopathology | **Sens:** NR*<br>**Spec:** NR*<br>**Acc:** NR | **Sens:** NR*<br>**Spec:** NR*<br>**Acc:** NR | 10.9% | *CT sensitivity/specificity were significantly higher than US only in patients with **equivocal** clinical scores (4-6). CT adds little for high scores (≥7). AIR score was superior to Alvarado. |
| **Coccolini *et al.* (2024)** [20] | Histopathology | **Sens:** NR<br>**Spec:** NR<br>**Acc:** NR | **Sens:** NR<br>**Spec:** NR<br>**Acc:** NR | Overall: 3.8%<br>US group: 2.9%<br>CT group: 4.1% | Clinical score + US showed the best balance (p=0.004). CT association was weaker (p=0.08). US is particularly useful in low-resource settings. |

**Table 3:** Risk of Bias Assessment Using QUADAS-2

| **Study (Author, Year [Ref.])** | **Risk of Bias** | | | | **Applicability Concerns** | | | **Overall Risk of Bias Judgment** |
|---|---|---|---|---|---|---|---|---|
| | **Patient Selection** | **Index Test** | **Reference Standard** | **Flow and Timing** | **Patient Selection** | **Index Test** | **Reference Standard** | |
| **Anan *et al.* (2025) [13]** | Unclear | Unclear | Low | Unclear | Low | Low | Low | Unclear |
| **Ahmed *et al.* (2022) [14]** | Unclear | Unclear | Unclear | Unclear | Unclear | Low | Unclear | Unclear |
| **Haentjens *et al.* (2023) [15]** | Unclear | Unclear | Low | Low | Low | Low | Low | Unclear |
| **Raffa *et al.* (2022) [16]** | Unclear | Unclear | Low | Low | Low | Low | Low | Unclear |
| **Lu *et al.* (2025) [17]** | Unclear | Unclear | Low | Unclear | Low | Low | Low | Unclear |
| **Zhao *et al.* (2025) [18]** | High | Unclear | Unclear | Unclear | Unclear | Low | Unclear | High |
| **Noori *et al.* (2023) [19]** | Low | Unclear | Low | Low | Low | Low | Low | Unclear |
| **Coccolini *et al.* (2024) [20]** | Low | Unclear | Low | Low | Low | Low | Low | Unclear |

## DISCUSSION

Computed tomography (CT) consistently outperforms ultrasound (US) in the diagnostic evaluation of acute appendicitis, although the optimal diagnostic strategy may vary according to clinical context and patient subpopulations. The pooled sensitivity of CT across the included studies ranged from 88% to 99.4%, while US sensitivity ranged from 40% to 85.7%, representing a substantial and clinically meaningful difference [13-20]. These findings are remarkably consistent with the broader body of evidence in the literature.

A comprehensive systematic review and meta-analysis by **Arruzza *et al.*** [3] evaluated 31 studies on CT and 10 studies on US for adult acute appendicitis, reporting a pooled CT sensitivity of 97.2% (95% CI: 95.8–98.1%) and specificity of 95.6% (95% CI: 94.1–96.7%), compared to US sensitivity of 82.1% (95% CI: 73.8–88.2%) and specificity of 85.9% (95% CI: 72.7–93.3%) [3].

The meta-analysis by **Weston *et al.*** (2005) similarly concluded that CT has better sensitivity and specificity than ultrasound in both unselected patients and those with equivocal diagnoses [9]. The consistently superior diagnostic performance of CT observed in our included studies—particularly the 99.4% sensitivity reported by **Haentjens *et al.*** and the 100% accuracy reported by **Anan *et al.*** [13]—thus aligns with decades of accumulated evidence establishing CT as the most accurate imaging modality for acute appendicitis.

The diagnostic superiority of CT over US can be attributed to several well-established technical and anatomical factors. CT provides superior spatial resolution, lack of operator dependence, and the ability to visualize the appendix regardless of patient body habitus, bowel gas, or retrocecal location—factors that frequently limit ultrasound examination.

The National Survey in China by **Ma *et al.*** (2024) reported that in adults, US sensitivity and specificity were 92% and 84%, respectively, while CT achieved 94% sensitivity and 79% specificity, with US performance being significantly higher in specialized children's hospitals compared to general hospitals [21]. This operator-dependent variability is a critical limitation of US that our findings corroborate; **Ahmed *et al.*** [14] reported a strikingly low US sensitivity of only 40%, which fell far below the Royal College of Radiologists' recommended standard of 70%, highlighting the substantial impact of operator experience and institutional expertise on US diagnostic yield.

The meta-analysis by **Arruzza *et al.*** [3] further demonstrated that US studies which excluded equivocal findings produced significantly inflated sensitivity and specificity values ($p < 0.0001$), suggesting that the true diagnostic performance of US in real-world clinical practice may be even lower than reported figures. The original Alvarado score, first described in 1986, combined with imaging has been shown to enhance diagnostic accuracy when applied in a structured algorithm, as demonstrated by multiple contemporary studies [5].

An important nuance emerging from our review is that while CT demonstrates superior overall diagnostic accuracy, ultrasound retains a valuable role as a first-line triage tool, particularly in specific patient populations and clinical contexts. The international multicenter study by **Coccolini *et al.*** [20], which included 3,123 patients, demonstrated that the combination of clinical scores with ultrasound achieved the best diagnostic balance ($p = 0.004$), with a negative appendectomy rate of only 2.9% in the US group—lower than the 4.1% observed in the CT group. This finding suggests that a diagnostic algorithm incorporating clinical assessment and ultrasound may be sufficient for a substantial proportion of patients, with CT reserved for equivocal or high-risk cases. The Appendicitis Inflammatory Response (AIR) score, introduced by Andersson in 2008, has been validated as superior to the Alvarado score in stratifying patients and reducing unnecessary imaging [22].

The randomized controlled trial by **Noori *et al.*** [19] provided further granularity, showing that CT is only beneficial for patients with equivocal clinical scores (Alvarado or AIR scores 4–6), while it adds little diagnostic value for patients with high scores (≥7) or low scores (≤4). This stratified approach is consistent with the findings of Anan ***et al.*** [13], who demonstrated that patients with concordant positive findings on both Alvarado score and focused ultrasound achieved 100% diagnostic accuracy, equivalent to CT, suggesting that the "double-check test" could serve as a reliable positive diagnostic tool, reserving CT for cases where the two tests are discordant or both negative. The Pediatric Appendicitis Score (PAS) combined with standardized ultrasound scoring, as proposed by **Nemeh *et al.*** [26], similarly demonstrated that scores of 7 or more achieved 90% sensitivity and 90.2% specificity, potentially eliminating the need for CT in many pediatric cases.

Subgroup analyses from our included studies revealed important patient-specific factors that influence the comparative performance of US and CT.

**Lu *et al.*** [17] demonstrated that while US and CT had similar diagnostic accuracy for patients with normal body mass index (BMI), CT provided significantly better accuracy for overweight, obese, and elderly patients ($p < 0.05$). This finding is clinically significant given the rising global prevalence of obesity and the known limitations of ultrasound in patients with increased abdominal adipose tissue, which impairs acoustic window and reduces visualization of the appendix.

The retrospective pediatric study by **Baykara and Kılıç** [23] reported a CT diagnostic accuracy of 81.8% compared to US accuracy of 70.5%, with US sensitivity of 77.2% and specificity of 52.6% versus CT sensitivity of 88.1% and specificity of 57.1% [23]. The contemporary systematic review and meta-analysis by **Castro-Luna *et al.*** [8] evaluated 37 articles including 22 conventional US studies (20,897 patients), 4 point-of-care US studies (280 patients), and 4 CT studies (1,389 patients) in pediatric populations, reporting pooled sensitivity and specificity of 93% and 89% for conventional US, 80% and 93% for POCUS, and 96% and 98% for CT [8]. While all four modalities demonstrated high diagnostic accuracy with no statistically significant differences ($p = 0.07$), the substantially higher CT sensitivity (96% vs. 93% for conventional US) and specificity (98% vs. 89%) underscore the continued role of CT as the most accurate imaging modality, particularly in complex or equivocal cases. A recent prospective study emphasizes that ultrasound (US) achieves over 90% sensitivity for pediatric appendicitis when performed by experienced pediatric radiologists, but this high accuracy is not replicable in general emergency settings [28].

The role of CT in discriminating complicated from uncomplicated appendicitis represents another important dimension of diagnostic imaging that has significant therapeutic implications. The systematic review and meta-analysis by **Bom *et al.*** [24] evaluated the accuracy of ultrasound, CT, and MRI in distinguishing complicated from uncomplicated appendicitis, including 11 CT studies and 2 US studies [24]. The summary estimates for CT demonstrated a mean sensitivity of 78% (95% CI: 64–88%) and mean specificity of 91% (95% CI: 85–99%) for detecting complicated appendicitis, with a negative predictive value of 93% at a median prevalence of 25%. This high negative predictive value is clinically valuable, as it can help identify patients who may be candidates for non-operative management with antibiotics alone, avoiding unnecessary emergency surgery.

However, **Bom *et al.*** [24] concluded that all three imaging modalities have limitations in discriminating between complicated and uncomplicated appendicitis, with CT having "far from perfect sensitivity". Our included studies did not specifically report on the ability to distinguish complicated from uncomplicated appendicitis, but the consistently high sensitivity and specificity of CT observed across all studies suggest that it remains the preferred modality when this distinction is clinically relevant, particularly given the 99.4% sensitivity reported by **Haentjens *et al.*** and the 92% sensitivity reported by **Raffa *et al.*** [16].

A recent study confirmed that CT findings such as the presence of extraluminal air, abscess, or phlegmon were highly specific for complicated appendicitis, reinforcing the role of CT in guiding non-operative management decisions [29].

The clinical implications of these findings extend beyond simple diagnostic accuracy to encompass important considerations of radiation exposure, cost, accessibility, and patient safety. The superior diagnostic performance of CT must be weighed against the potential risks of ionizing radiation, particularly in pediatric and young adult populations where radiation-induced malignancies are a concern. The American College of Radiology Appropriateness Criteria recommend ultrasound as the initial imaging modality for suspected appendicitis in children and pregnant women, with CT reserved for cases where US is equivocal or non-diagnostic. The implementation of ultrasound-first pathways has been shown to significantly reduce CT utilization; Steinl ***et al.*** demonstrated a reduction in CT use from 32% to 2% following the implementation of a US and rapid MRI pathway in children [30]. However, the relatively low sensitivity of US—particularly the 40% reported by Ahmed ***et al.*** and the 73% reported by **Zhao *et al.*** [18]—raises concerns about missed diagnoses and delayed treatment if US is relied upon as the sole imaging modality. The negative appendectomy rate across our included studies ranged from 2% to 22.9%, with the highest rate (22.9%) observed in the study by **Raffa *et al.*** [16], which also reported the lowest US specificity (50%). This suggests that over-reliance on US without confirmatory CT in equivocal cases may contribute to unnecessary surgeries, particularly in institutions with limited operator expertise. The concept of "concurrent imaging"—where both US and CT are performed during the same encounter—was examined by **Scaife *et al.*** [25], who found that increased US utilization was negatively correlated with concurrent CT use ($r = -0.27$, $p < 0.001$), suggesting that ultrasound-first approaches can effectively reduce overall CT utilization when implemented appropriately. A cost-effectiveness analysis demonstrated that an ultrasound-first strategy, followed by CT only for equivocal cases, was the most cost-effective approach in both adult and pediatric populations, provided that ultrasound had a sensitivity of at least 75% [27].

## LIMITATIONS

Several limitations of this systematic review impact its conclusions. Heterogeneity in study design, patient populations, and methods reduces generalizability. The presence of retrospective studies introduces potential selection bias. Variability in sample sizes affects statistical power, with some studies being too small to identify meaningful differences. Inadequate reporting of methodological details hampers risk assessment. Inconsistent definitions of "equivocal" ultrasound findings and varied criteria for advancing to CT could introduce verification bias. Additionally, using histopathology as a reference may cause spectrum bias, excluding patients treated conservatively. The findings are not generalizable to specific subpopulations, such as pregnant women and young children, and publication bias may affect the prevalence of published results.

## CONCLUSION

Computed tomography (CT) excels in diagnostic accuracy for acute appendicitis compared to ultrasound (US), with CT sensitivity between 88% and 99.4%, while US sensitivity ranges from 40% to 85.7%. CT's advantages include higher sensitivity, specificity, and overall accuracy, alongside its capability to visualize the appendix independent of body habitus. However, an optimal diagnostic strategy requires a risk-stratified, patient-centered approach that incorporates clinical presentation, patient demographics, and available resources. Employing clinical scoring systems with focused ultrasound can provide a cost-effective, radiation-free initial diagnosis; CT should be used for inconclusive or complex cases. Future research should aim at standardizing ultrasound protocols, enhancing operator training, and conducting studies on US versus CT effectiveness in subpopulations like children and pregnant women to refine diagnostic guidelines and improve patient care.

### Ethics Approval and Consent to Participate

As this study is a systematic review of previously published literature, it did not involve the collection of primary data from human or animal subjects. Therefore, formal ethical approval and individual participant consent were not required. All included studies were conducted in accordance with the ethical standards of their respective institutional and/or national research committees.

### Consent for Publication

Not applicable, as this manuscript does not contain any individual person’s data in any form (e.g., individual details, images, or videos).

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Competing Interests

The authors declare that they have no competing interests.

### Authors' Contributions

All authors cooperated in drafting, wrtiting, synthesis of the study and they reviewed and approved the final manuscript.

### Acknowledgements

None

## REFERENCES


1. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World journal of emergency surgery. 2020 Apr 15;15(1):27.
2. American College of Emergency Physicians. 2023 ACEP Guidelines for Emergency Department Patients with Suspected Appendicitis. Ann Emerg Med. 2023. Available from: https://www.guidelinecentral.com/guideline/2556960/#section-2556964
3. Arruzza E, Milanese S, Li LS, Dizon J. Diagnostic accuracy of computed tomography and ultrasound for the diagnosis of acute appendicitis: A systematic review and meta-analysis. Radiography. 2022;28(4):1127-41.
4. Bahrami M, Mirgaloyebayat H, Mohajeri Z, Mohammadi H, Afshari SA, Fazeli P, et al. The diagnostic value of the computed tomography scan and ultrasonography in acute appendicitis. American Journal of Nuclear Medicine and Molecular Imaging. 2023 Feb 15;13(1):11.
5. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Annals of emergency medicine. 1986;15(5):557-64.
6. Han H, Letourneau ID, Abate YH, Abdelmasseh M, Abu-Gharbieh E, Adane TD, et al. Trends and levels of the global, regional, and national burden of appendicitis between 1990 and 2021: findings from the Global Burden of Disease Study 2021. The Lancet Gastroenterology & Hepatology. 2024;9(9):825-58.
7. Dietrich CF, Sirlin CB, O'Boyle M, Dong Y, Jenssen C. Editorial on the current role of ultrasound. Applied Sciences. 2019 Aug 26;9(17):3512. **https://doi.org/10.3390/app9173512**
8. Castro-Luna DI, Porras-Hernandez JD, Flores-Garcia JA, Dies-Suarez P, Servin-Martinez MF, Pierdant-Perez M. Contemporary ultrasound, computed tomography, or magnetic resonance imaging for acute appendicitis diagnosis in children and adolescents: systematic review and meta-analysis. Pediatric Radiology. 2025;55(7):1448-64.
9. Weston AR, Jackson TJ, Blamey S. Diagnosis of appendicitis in adults by ultrasonography or computed tomography: a systematic review and meta-analysis. International journal of technology assessment in health care. 2005;21(3):368-79.
10. Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
11. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. doi:10.1186/s13643-016-0384-4
12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009
13. Anan M, Khafagy W, Abdelnaby M, Hussien AF, Farid M. Tailoring of computed tomography scanning in patients with acute appendicitis according to combined assessment by Alvarado score and focused ultrasound. The Egyptian Journal of Surgery. 2024;43(4):1525-31.
14. Ahmed S, Hassan K, Abubaker M, Wilson J, Magee C. EP-142 assessing the diagnostic sensitivity of CT and ultrasound in suspected appendicitis. British Journal of Surgery. 2022;109(Supplement_5): znac245-037. https://doi.org/10.1093/bjs/znac245.037
15. Haentjens L, Coussement L, Vuylsteke M. Value of ultrasound and computed tomography in the diagnosis of acute appendicitis with histopathology as gold standard. Acta Chirurgica Belgica. 2023;123(6):654-8.
16. Raffa A, Abduljabbar A, Alharthy A. Comparison of ultrasound and computed tomography scanning accuracy in diagnosing acute appendicitis at King Abdulaziz university hospital. Cureus. 2022 Nov 25;14(11).
17. Lu K, Zhong F, Miao J, Sun C, Zhou K, Wang W, et al. Assessment of diagnostic value of ultrasound and multi-slice spiral computed tomography in acute appendicitis: a retrospective study. Abdominal Radiology. 2025;50(3):1117-22.
18. Zhao Y, Liu H, Guo D. Diagnostic efficacy of ultrasound and computed tomography for acute appendicitis: A single center retrospective study. Medicine. 2025;104(13):e41968.
19. Noori IF, Jabbar AS, Noori AF. Clinical scores (Alvarado and AIR scores) versus imaging (ultrasound and CT scan) in the diagnosis of equivocal cases of acute appendicitis: a randomized controlled study. Annals of Medicine and Surgery. 2023;85(4):676-83.
20. Coccolini F, Martinez-Perez A, Licitra G, De Angelis N, Cremonini C, Strambi S, et al. Diagnostic pathways in acute appendicitis: clinical score plus ultrasound outperform CT scan or upfront surgery, especially in doubtful cases. World Journal of Surgery. 2024;48(6):1350-9.
21. Ma L, Niu Z, Xu W, Wang R, Gao L, Tao X, et al. Current usage and diagnostic performance of cross-sectional imaging in patients with suspected acute appendicitis: a national survey in China. Abdominal Radiology. 2025;50(5):1979-85.
22. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World journal of surgery. 2008;32(8):1843-9.
23. Baykara AS, Baykara S. Acute appendicitis in children: evaluation of the diagnostic efficacy of ultrasonography and computed tomography. Cureus. 2023;15(8).
24. Bom WJ, Bolmers MD, Gans SL, Van Rossem CC, Van Geloven AA, Bossuyt PM, et al. Discriminating complicated from uncomplicated appendicitis by ultrasound imaging, computed tomography or magnetic resonance imaging: systematic review and meta-analysis of diagnostic accuracy. BJS open. 2021 Mar;5(2):zraa030.
25. Scaife JH, Iantorno SE, Bucher BT. Rates of concurrent computed tomography imaging following ultrasound for pediatric patients with appendicitis. Journal of Surgical Research. 2024;302:134-43.
26. Nemeh C, Hassan IN, Walsh A, Iyer S, Gitzelmann CA. Pediatric ultrasound appendicitis score for the diagnosis of

acute appendicitis to reduce computed tomography scan utilization. The American Surgeon™. 2024;90(2):245-51.

27. Jennings R, Guo H, Goldin A, Wright DR. Cost-effectiveness of Imaging Protocols for Suspected Appendicitis. *Pediatrics*. 2020;145(2):e20191352. doi:10.1542/peds.2019-1352
28. Mittal MK, Dayan PS, Macias CG, Bachur RG, Bennett J, Dudley NC, et al. Pediatric Emergency Medicine Collaborative Research Committee of the American Academy of Pediatrics. Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. Acad Emerg Med. 2013;20(7):697-702. doi: 10.1111/acem.12161.
29. Iamwat J, Teerasamit W, Apisarnthanarak P, Noppakunsomboon N, Kaewlai R. Predictive ability of CT findings in the differentiation of complicated and uncomplicated appendicitis: a retrospective investigation of 201 patients undergone appendectomy at initial admission. Insights into imaging. 2021;12(1):143.
30. Steinl G, Grabski D, Fleming M 2nd, Levin D, McGahren E, McCullough W, et al. Implementation of ultrasound and fast magnetic resonance imaging pathway reduces computed tomography utilization in children with suspected appendicitis. Pediatr Surg Int. 2023;39(1):238. doi:10.1007/s00383-023-05521-y