


Discriminating complicated from uncomplicated appendicitis by ultrasound imaging, computed tomography or magnetic resonance imaging: systematic review and meta-analysis of diagnostic accuracy

W. J. Bom ^{1,2,*}, M. D. Bolmers¹, S. L. Gans¹, C. C. van Rossem², A. A. W. van Geloven³, P. M. M. Bossuyt⁴, J. Stoker⁵ and M. A. Boermeester¹

¹Department of Surgery, Amsterdam University Medical Centre, Location Academic Medical Centre, Amsterdam Gastroenterology and Metabolism, University of Amsterdam, Amsterdam, the Netherlands

²Department of Surgery, Tergooi Hospital Hilversum, Hilversum, the Netherlands

³Department of Surgery, Maastad Ziekenhuis, Rotterdam, the Netherlands

⁴Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, Amsterdam Public Health, University of Amsterdam, Amsterdam, the Netherlands

⁵Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centre, Location Academic Medical Centre, Amsterdam Gastroenterology and Metabolism, University of Amsterdam, Amsterdam, the Netherlands

*Correspondence to: Department of Surgery (G4-124), Amsterdam UMC, Academic Medical Centre, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands (e-mail: w.j.bom@amsterdamumc.nl)

Abstract

Background: Discriminating complicated from uncomplicated appendicitis is crucial. Patients with suspected complicated appendicitis are best treated by emergency surgery, whereas those with uncomplicated appendicitis may be treated with antibiotics alone. This study aimed to obtain summary estimates of the accuracy of ultrasound imaging, CT and MRI in discriminating complicated from uncomplicated appendicitis

Methods: A systematic literature review was conducted by an electronic search in PubMed, Embase and the Cochrane Library for studies describing the diagnostic accuracy of complicated versus uncomplicated appendicitis. Studies were included if the population comprised adults, and surgery or pathology was used as a reference standard. Risk of bias and applicability were assessed with QUADAS-2. Bivariable logitnormal random-effect models were used to estimate mean sensitivity and specificity.

Results: Two studies reporting on ultrasound imaging, 11 studies on CT, one on MRI, and one on ultrasonography with conditional CT were included. Summary estimates for sensitivity and specificity in detecting complicated appendicitis could be calculated only for CT, because of lack of data for the other imaging modalities. For CT, mean sensitivity was 78 (95 per cent c.i. 64 to 88) per cent, and mean specificity was 91 (85 to 99) per cent. At a median prevalence of 25 per cent, the positive predictive value of CT for complicated appendicitis would be 74 per cent and its negative predictive value 93 per cent.

Conclusion: Ultrasound imaging, CT and MRI have limitations in discriminating between complicated and uncomplicated appendicitis. Although CT has far from perfect sensitivity, its negative predictive value for complicated appendicitis is high.

Introduction

Imaging is part of the standard workup for diagnosing appendicitis. Ultrasound imaging, CT and MRI are used most frequently. A diagnosis of acute appendicitis can be made adequately based on radiological findings¹⁻⁶. The first choice of diagnostic modality differs. In Europe, ultrasonography is often used as the first approach to diagnose acute appendicitis, combined with CT if necessary in patients with inconclusive or negative results⁶⁻⁸; in Northern America, CT first is preferred over US⁹.

As well as confirming the diagnosis of appendicitis, imaging may also help in distinguishing between complicated and uncomplicated appendicitis. Nowadays, it is believed that uncomplicated

and complicated appendicitis are different entities, and may require different treatment strategies. In patients with complicated appendicitis, early surgical treatment is necessary to avert a complicated postoperative course¹⁰. In uncomplicated appendicitis, semi-urgent surgery and antibiotic treatment may be an option, and even a wait-and-see policy is currently being investigated^{11,12}.

Several studies¹³⁻¹⁵ have evaluated non-operative treatment of uncomplicated appendicitis. Treatment with antibiotics may be just as safe and effective as surgical treatment, without the risk of surgical complications, similar to management strategies for other inflammatory bowel diseases, such as diverticulitis or colitis¹⁴⁻¹⁶. Although studies on antibiotic treatment have shown

Received: June 20, 2020. Revised: September 20, 2020. Accepted: September 28, 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of BJS Society Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

a low initial failure rate (below 10 per cent), 22 per cent of patients need an appendectomy within 1 year of follow-up and up to 40 per cent within 5 years^{15,17}. An essential factor in the success of non-operative treatment is the selection of patients with truly uncomplicated appendicitis.

Identifying uncomplicated appendicitis is improved by ruling out complicated appendicitis, which indirectly improves the selection of patients in need of urgent surgery. Thus, to employ different treatment strategies, a discriminatory test with high sensitivity and high negative predictive value (NPV) for ruling out complicated appendicitis is of the utmost importance. The two largest published RCTs^{18,19} assessing the effectiveness of antibiotic treatment for uncomplicated appendicitis used CT alone to discriminate between uncomplicated and complicated appendicitis. They reported 18 per cent¹⁸ and 1.5 per cent¹⁹ complicated appendicitis in the surgery arm. Therefore, diagnosing complicated appendicitis by CT alone may not be good enough. Moreover, if ultrasound imaging or MRI is adequate in discriminating these entities, perforated appendicitis may be detected without the use of radiation or intravenous contrast agent. This systematic review was designed to obtain summary estimates of the accuracy of ultrasonography, CT and MRI in discriminating complicated from uncomplicated appendicitis.

Methods

The review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) with registration number CRD42020150771. PRISMA-DTA guidelines²⁰ for reporting systematic reviews of diagnostic test accuracy studies were used to prepare this manuscript.

Review question

This review aimed to obtain summary estimates of the (comparative) accuracy of ultrasonography, CT and MRI in discriminating complicated from uncomplicated appendicitis. The secondary aim was to explore sources of heterogeneity of the accuracy of these modalities.

Eligibility criteria

Diagnostic accuracy studies were eligible in which ultrasound imaging, CT, MRI, or a combination of these imaging modalities were used to discriminate between complicated and uncomplicated appendicitis. The definition of complicated appendicitis could differ among studies, but was used as defined in the original publications. In studies comparing 'perforated' and 'non-perforated appendicitis', perforated appendicitis was considered as complicated appendicitis and non-perforated appendicitis as uncomplicated appendicitis. Both histopathological and perioperative findings, or a combination, were valid as a reference standard.

An estimate of sensitivity and specificity for complicated appendicitis was mandatory for inclusion. Both retrospective and prospective studies were eligible. Studies had to mention the radiological diagnosis of complicated appendicitis, and either report or allow the construction of 2×2 tables on accuracy. If the reported data were unclear, the authors were contacted by e-mail. Only studies reporting only or predominantly on adults (at least 75 per cent; at least 15 years of age) were included, as diagnostic accuracy and workup are different in children²¹. If age was not mentioned, or the study did not report the incidence in adults, this was reported and marked as high risk of bias. The electronic search did not use any limitations, but in the full-text

selection only studies reported in English, German or Dutch were selected.

Information sources

An electronic search was performed in PubMed, Embase and the Cochrane Library. Reference lists of included full-text study reports were searched manually for missing relevant articles. Keywords assigned to the retrieved articles were used for the additional search. The final search date was 12 November 2019.

Literature search

The search strategy is described in [Appendix S1](#).

Study selection

Two reviewers independently evaluated potentially eligible studies, assessed these for risk of bias, and extracted data. Disagreements were discussed. If no consensus was reached, a consensus meeting with a third reviewer was decisive.

Data extraction and critical appraisal

Data from the included studies were systematically, independently, and blindly extracted by two reviewers using a structured study record form. Disagreements were resolved in consensus meetings.

The following items were extracted: title, year of publication, journal of publication, name of first author, number of patients, study design, country, inclusion criteria, exclusion criteria, true positives, true negatives, false positives, false negatives, proportion of male patients, median (range) age, radiological features for complicated appendicitis, reference standard, definition of complicated appendicitis, imaging characteristics and protocols if reported, number of observers and observers' experience.

Risk of bias and concerns about applicability to the review questions were assessed with the QUality Assessment of Diagnostic Accuracy Studies tool, version 2 (QUADAS-2)²². Two reviewers independently assessed the included articles, and disagreements were resolved in consensus meetings including a third reviewer, if necessary. Studies that included patients retrospectively based on an appendectomy registration code were marked as at high risk of bias in the patient selection domain, as this might have influenced radiologist judgement.

Statistical analysis

For all included studies, true positives, true negatives, false positives and false negatives for patients with complicated appendicitis were extracted. If data were available, the total number of patients was the number of patients with appendicitis according to the reference standard. Patients considered to have appendicitis by the radiologist, but without appendicitis in the reference group, were included for analyses, if the data were available.

When only sensitivity and specificity were mentioned, counts were combined with the prevalence of complicated appendicitis in the study and 2×2 tables for the diagnosis of complicated appendicitis were reconstructed. If possible, 3×3 tables were extracted, adding the diagnosis 'not appendicitis' in both reference and imaging groups. If two imaging observers were used, mean counts from contingency tables were used and rounded.

Study-specific estimates of sensitivity and specificity with 95 per cent c.i. are presented in forest plots. Bivariable logitnormal random-effect models were used to estimate mean sensitivity and specificity with 95 per cent confidence intervals. A hierarchical summary receiver operating characteristic (HSROC) curve

was plotted. Projected post-test probabilities were calculated based on the median prevalence of complicated appendicitis in the eligible studies, and the summary estimates of sensitivity and specificity. As complicated appendicitis needs to be ruled out, a sensitivity of at least 90 per cent and a specificity of at least 50 per cent were deemed necessary. Data were analysed with Review Manager version 5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) and R (The R Foundation for Statistical Computing, Vienna, Austria) using the *mada* package.

Potential sources of heterogeneity were explored, related to imaging characteristics, reader experience, use of intravenous contrast medium, slice thickness, and CT with a standard versus low radiation dose. Bivariable meta-regression was planned if sufficient studies were available.

Results

The search identified 5285 studies, of which 147 potentially eligible studies were selected for full-text evaluation. After evaluation of eligibility, 13 studies^{23–35} were included in the review (Fig. 1). Eleven studies^{23,25–33,35} reported on CT, one study²⁴ on ultrasound imaging alone, and one study³⁴ described ultrasound imaging, ultrasonography with conditional CT (CT after negative or inconclusive ultrasound imaging) and MRI. Thus data on ultrasound imaging could be obtained from two studies, CT from 11 studies, and MRI from one study. The combination of ultrasonography with conditional CT was reported in one study³⁴.

Study and patient characteristics

Study characteristics are shown in Table 1. Nine studies were retrospective, and two studies were prospective (1 study³³ reported on CT, and the other³⁴ on ultrasonography alone or with conditional CT, and MRI). In two studies^{23,25}, it was unclear whether the study was performed prospectively or retrospectively. Three studies described the role of low- versus high-dose radiation CT, of which two^{29,31} were performed retrospectively and one³³ prospectively. Ten studies^{23,24,26,28–33,35} used perforated appendicitis (rather than complicated appendicitis) as the outcome. Only 3^{25,27,34} of the 13 studies evaluated complicated appendicitis (rather than perforated appendicitis). One study²³ only used histopathology as the reference standard. One study³⁴ used an expert panel for final diagnosis.

In total, data on 1892 patients were reported, of which 620 had complicated appendicitis (Table 2). The median prevalence of complicated appendicitis was 25 per cent. For ultrasound imaging data were available on 218 patients, for CT 1667, for MRI 120, and the combination of US and CT 125. In four studies, the proportion of adults was unclear^{24–26,28}. Two^{27,32} of the studies included approximately 75 per cent adults (based on mean(s.d.) age) and were therefore included. The remaining seven studies^{23,29–31,33–35} included only adult patients (above 15 years of age). For further details, see Tables S1 and S2.

Risk of bias

The risk of bias was high in 9 of 13 studies in the following domains: patient selection, index test and reference standard. Applicability concerns were considered high in 9 studies for the domains patient selection, index test and reference standard, and low in three studies. The QUADAS-2 characteristics and summary are depicted in Fig. 2 and Table 3.

Diagnostic accuracy

Ultrasound imaging

One retrospective cohort²⁴, based on patients with surgically and histopathology proven appendicitis, reported on the diagnostic accuracy of ultrasound imaging for complicated appendicitis, with a sensitivity of 86 per cent (19 of 22) and a specificity of 60 per cent (47 of 78). Based on a prevalence of 25 per cent, the positive predictive value (PPV) was calculated to be 42 per cent and the NPV as 93 per cent.

Another study³⁴, of a prospective cohort, included all patients with clinically suspected appendicitis. If ultrasonography was inconclusive, conditional CT was performed. Outcomes for ultrasound imaging alone were not reported separately, but could be calculated from the data. For ultrasound imaging, the sensitivity was 32 per cent (10 of 31) and specificity was 93 per cent (81 of 87). Based on a prevalence of 25 per cent, PPV and NPV were 60 and 80 per cent respectively. Patients with a diagnosis other than appendicitis were excluded in these calculations.

Because of high heterogeneity, caused by differences in study design (retrospective versus prospective studies) and patient selection (proven versus suspected appendicitis), no meta-analysis was performed. For both studies^{24,34}, it was not possible to construct a 3×3 table that included the diagnosis 'no appendicitis'.

CT

Eleven studies evaluated the diagnostic accuracy of CT for detecting complicated appendicitis: eight retrospective studies^{26–32,35}, one prospective study³³, and two studies^{23,25} with unclear design. Reported estimates ranged from 28 to 95 per cent for sensitivity and 71 to 100 per cent for specificity. The forest plot is depicted in Fig. 3. The summary estimates were 78 (95 per cent c.i. 64 to 88) per cent for sensitivity, with a specificity of 91 (85 to 99) per cent. See Fig. 4 for the HSROC curve. At a median prevalence of complicated appendicitis of 25 per cent, the PPV of CT would be 74 per cent and the NPV 93 per cent.

The only prospective study³³ on the diagnostic accuracy of CT for complicated appendicitis reported a sensitivity of 55 per cent and a specificity of 88 per cent. For the eight retrospective studies^{26–32,35}, the summary estimate of sensitivity was 81 (95 per cent c.i. 62 to 91) per cent with a specificity of 93 (84 to 97) per cent. The two studies^{23,25} with unclear design were not analysed separately.

Three studies^{29,31,33} compared accuracy for low versus standard radiation dose CT. The estimates for sensitivity and specificity were not significantly different ($P=0.286$). For regular-dose CT, the sensitivity was 68 (95 per cent c.i. 45 to 85) per cent at a specificity of 88 (80 to 93) per cent; for low-dose CT, these estimates were 58 (32 to 80) and 75 (40 to 94) per cent respectively. There were insufficient data to test for differences related to radiologist experience, intravenous contrast or slice thickness (Table S3). It was not possible to construct a 3×3 table that included the diagnosis of no appendicitis for any of these studies.

MRI

One study³⁴ reported on the diagnostic accuracy of MRI in detecting complicated appendicitis: sensitivity was 57 (95 per cent c.i. 37 to 75) per cent and specificity was 86 (77 to 92) per cent. At a prevalence of 25 per cent, this would indicate a PPV of 58 per cent and NPV of 86 per cent. The 3×3 table, including the diagnosis of no appendicitis, is shown in Table S4.

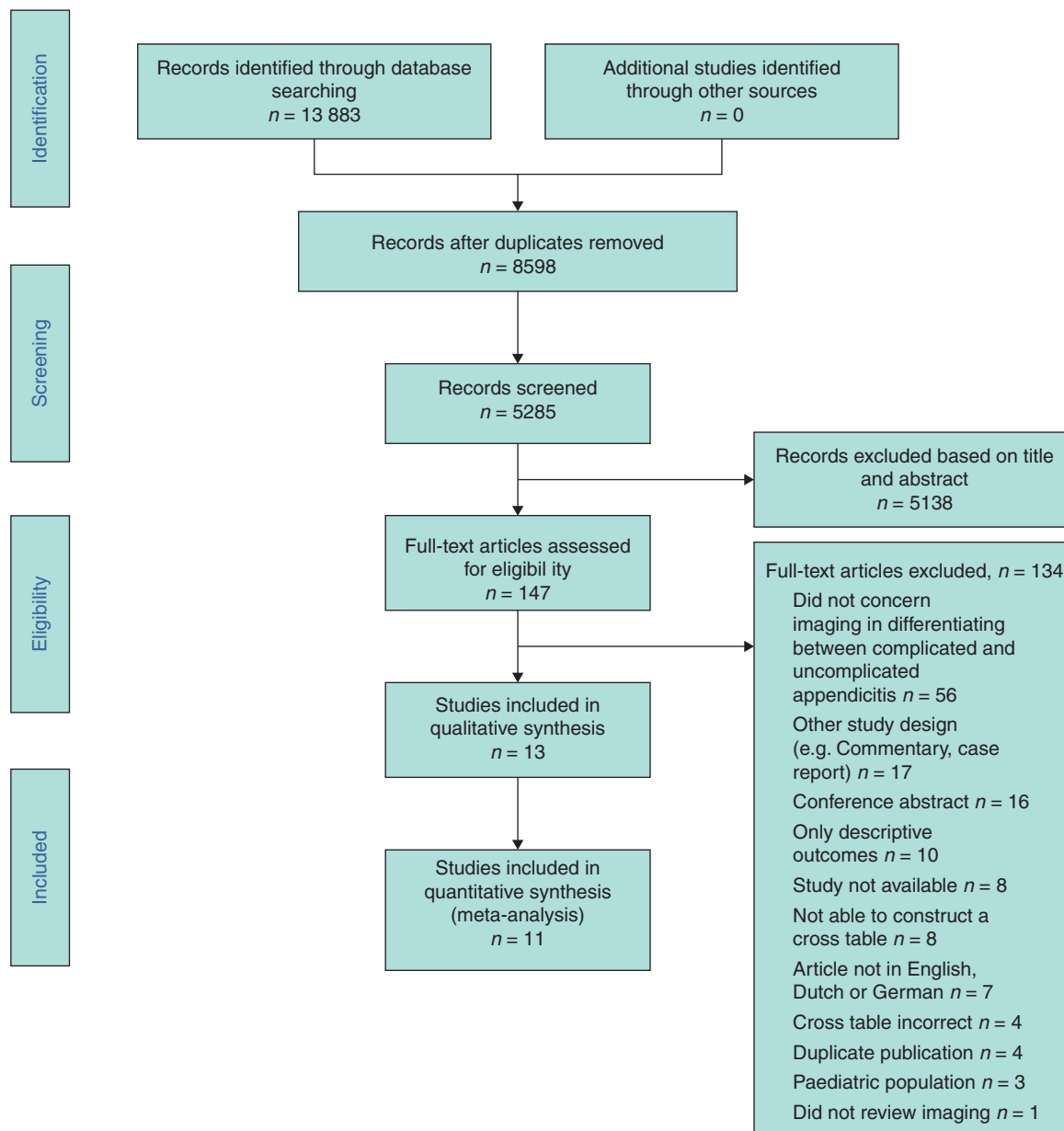


Fig. 1 PRISMA diagram for the review

Ultrasonography with conditional CT

One study³⁴ reported accuracy estimates for ultrasound imaging followed by CT in patients with an inconclusive or negative ultrasound scan. The sensitivity was 48 (95 per cent c.i. 30 to 67) per cent at a specificity of 93 (85 to 97) per cent. At a prevalence of 25 per cent, a PPV of 70 per cent and NPV of 84 per cent was calculated. The 3×3 table, including the diagnosis of no appendicitis, is seen in Table S4.

Discussion

Based on currently available evidence, ultrasonography, CT, MRI alone, or a combination ultrasound imaging and CT (conditional CT) have limitations in discriminating between complicated and uncomplicated appendicitis. Imaging cannot reliably rule out a complicated presentation of acute appendicitis in need of urgent surgery, as a sensitivity of 90 per cent does not appear to be

reached and there is high heterogeneity between studies. With respect to ruling in complicated appendicitis, CT seems to reach a specificity above 90 per cent but still is not perfect, and ruling in is considered less important. In the absence of comparative studies, no head-to-head comparisons could be made between imaging techniques or strategies. In a meta-regression comparing low-dose with normal-dose radiation CT, no significant difference that caused the heterogeneity in CT studies was found. Diagnostic accuracy was worse in prospective studies, which is important as these are the closest to daily clinical practice.

A recent meta-analysis³⁷ described 10 CT features to discriminate between complicated and uncomplicated appendicitis, nine of which showed high individual specificity but low sensitivity. Periappendicular fat infiltration had a sensitivity of 94 per cent, but a specificity of 40 per cent³⁷. Another study³⁸ built a regression model based on radiological features. Although the authors of that study did not report results for diagnostic accuracy, only

Table 1 Study characteristics

Reference	Year	Country	Study design	Modality	Reference standard	Definition target condition
Borushok et al. ²⁴	1990	USA	R	US	S+PA	Perforated
Choi et al. ²⁵	1998	USA	U	CT	S+PA	Necrotizing and perforated appendicitis
Foley et al. ²⁶	2005	USA	R	CT	S+PA	Perforated
Miki et al. ²⁷	2005	Japan	R	CT	S+PA	Gangrenous and perforated appendicitis
Tsuboi et al. ²⁸	2008	Japan	R	CT	S+PA	Perforated
Seo et al. ²⁹	2009	South Korea	R	CT H/L	S+PA	Perforated
Suthikeeree et al. ³⁰	2010	Thailand	R	CT	S+PA	Perforated
Kim et al. ³¹	2011	South Korea	R	CT H/L	S+PA	Perforated
Suh et al. ³²	2011	South Korea	R	CT	S+PA	Perforated
Kim et al. ³³	2012	South Korea	P	CT H/L	S+PA	Perforated
Leeuwenburgh et al. ³⁴	2014	Netherlands	P	US, US+CT, MRI	S+PA	Perforated appendicitis or pus in abdomen
Liu et al. ³⁵	2015	China	R	CT	S+PA	Perforated
Ali et al. ²³	2018	Pakistan	U	CT	PA	Perforated

R, retrospective; US, ultrasonography; S, surgery; PA, histopathology; U, uncertain; H/L, high and low radiation dose; P, prospective.

Table 2 Patient characteristics

Reference	No. with appendicitis	No. with complicated appendicitis	Age (years)*	% aged ≥ 15 years
Ali et al. ²³	236	42 (17.8)	40 (15–n.r.)	100
Borushok et al. ²⁴	100	22 (22.0)	29 (1–71)	n.r.
Choi et al. ²⁵	105	69 (65.7)	n.r.	n.r.
Foley et al. ²⁶	86	21 (24)	34 (8–87)	n.r.
Kim et al. ³¹	52	7 (13)	28 (15–40)	100
Kim et al. ³³	180	42 (23.3)	30 (22–37)	100
Leeuwenburgh et al. ³⁴ (US)	125	31 (26.3)	35 (24–49) [†]	100
Leeuwenburgh et al. ³⁴ (US+CT)	125	31 (24.8)	35 (24–49) [†]	100
Leeuwenburgh et al. ³⁴ (MRI)	120	30 (25.0)	35 (24–49) [†]	100
Liu et al. ³⁵	187	41 (21.9)	48 (19–87)	100
Miki et al. ²⁷	64	28 (44)	32 (4–78)	~75
Seo et al. ²⁹	79	24 (30)	39 (15–80)	100
Suh et al. ³²	528	226 (42.8)	29 28 (15 20)	~75
Suthikeeree et al. ³⁰	48	27 (56)	56 (15–96)	100
Tsuboi et al. ²⁸	102	40 (39.2)	37 (4–82)	n.r.

Values in parentheses are percentages unless indicated otherwise; *values are mean (range), except [†] median (i.q.r.). n.r., not reported.

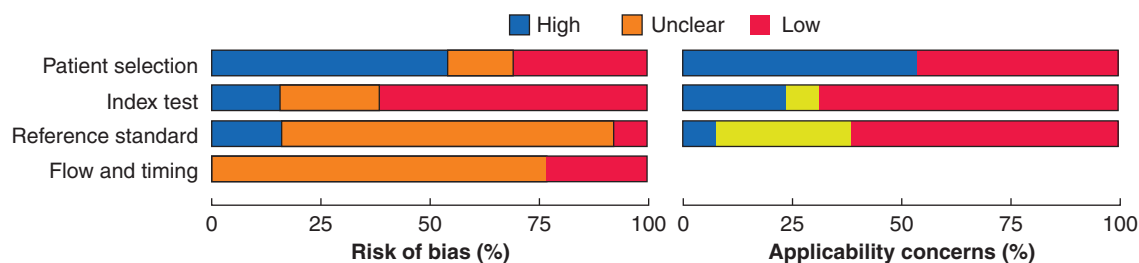


Fig. 2 Risk of bias and applicability concerns

Review authors' judgements about each domain presented as percentages across the included studies.

Table 3 Summary of risk of bias and applicability concerns: review authors' judgements about each domain for each study

Reference	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ali <i>et al.</i> ²³	?	?	-	?	+	?	-
Borushok <i>et al.</i> ²⁴	-	-	+	+	-	-	+
Choi <i>et al.</i> ²⁵	-	?	?	?	-	+	+
Foley <i>et al.</i> ²⁶	-	+	?	?	-	+	+
Kim <i>et al.</i> ³¹	+	-	?	?	+	-	+
Kim <i>et al.</i> ³³	+	+	?	?	+	+	+
Leeuwenburgh <i>et al.</i> ³⁴	+	+	?	?	+	+	+
Liu <i>et al.</i> ³⁵	?	+	?	?	+	+	?
Miki <i>et al.</i> ²⁷	-	+	?	?	-	+	?
Seo <i>et al.</i> ²⁹	+	+	?	+	+	+	+
Suh <i>et al.</i> ³²	-	?	?	?	-	-	?
Suthikeeree <i>et al.</i> ³⁰	-	+	-	?	-	+	?
Tsuboi <i>et al.</i> ²⁸	-	+	?	+	-	+	+

?, Unclear; +, high; -, low.

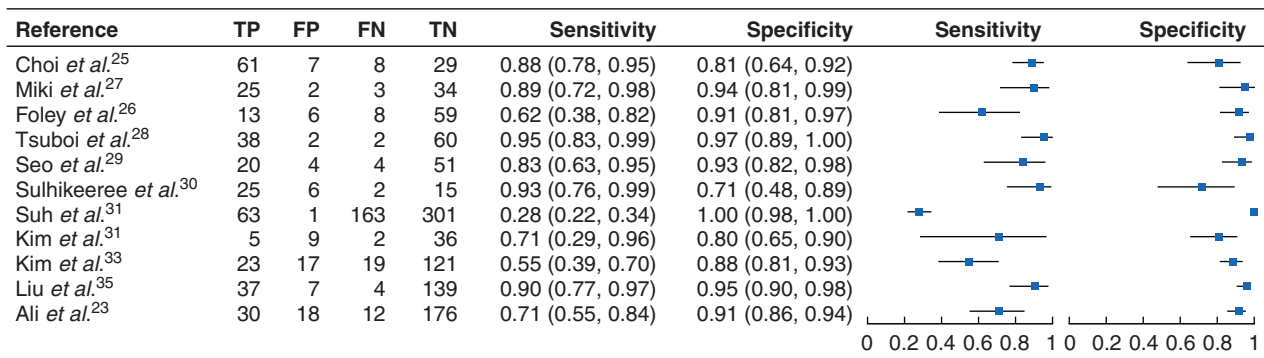


Fig. 3 Forest plots of sensitivity and specificity for the diagnostic accuracy of CT in detecting complicated appendicitis

Bivariable logitnormal random-effect models were used for meta-analysis. Mean sensitivity and specificity values are shown with 95 per cent confidence intervals. TP, true positives; FP, false positives; FN, false negatives; TN, true negatives.

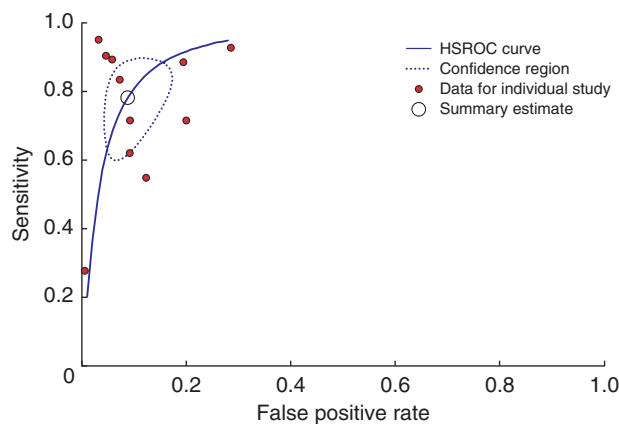


Fig. 4 Hierarchical summary receiver operating characteristic curve for CT

HSROC, hierarchical summary receiver operating characteristic.

12 of 21 patients with a gangrenous appendicitis were identified as such, leading to a sensitivity of 57 per cent. However, in clinical practice, radiological features are not interpreted separately. The radiologist's decision will be based on a combination of specific features, together with the severity of the feature.

Ultrasound imaging has limitations as a single technique in detecting acute appendicitis (sensitivity 69–83 per cent and specificity 81–93 per cent)^{16,19}. Thus it is intrinsically limited in distinguishing between complicated and uncomplicated appendicitis. This was also shown in the one prospective study³⁴, which reported a sensitivity of 32 per cent and specificity of 93 per cent

for ultrasonography. CT and MRI were better at detecting acute appendicitis. CT had a sensitivity for acute appendicitis of 91–94 per cent and specificity of 90–94 per cent^{3,21}, whereas MRI had a sensitivity of 96 per cent and specificity of 96 per cent³⁹. The 3×3 tables, including the diagnosis of no appendicitis, were performed in only one study³⁴, which evaluated conditional CT (CT after ultrasound imaging if necessary) and MRI. The 3×3 tables, including the diagnosis of no appendicitis, were performed in only one study³⁴, describing both ultrasonography with CT if necessary and MRI. Merely reporting the discriminatory capacity of imaging, in which the radiologist is confident of the final diagnosis of appendicitis, might lead to bias. Incorrect classification of appendicitis type may be associated with the wrong treatment choice. In addition, patients without acute appendicitis, but falsely diagnosed as having simple appendicitis, are overtreated.

Studies not included in this review have reported data on the performance of the Alvarado and Appendicitis Inflammatory Response (AIR) scores, with respect to discrimination between complicated and uncomplicated appendicitis^{35,40–42}. The authors concluded there might be an association between score level and complexity of appendicitis, but no data on diagnostic accuracy measures were presented. Other authors^{43,44} created a prediction model for complicated appendicitis using clinical features, but also failed to report diagnostic accuracy statistics. Two further studies^{45,46} reported on a combination of clinical and radiological features. Atama and colleagues⁴⁵ developed two scoring systems, combining clinical and radiological features: one with ultrasonographic and one with CT features. The ultrasonographic system reached a sensitivity of 96.6 per cent and a specificity of 45.7 per cent; for CT, sensitivity and specificity were 90.2 and 70.3 per cent respectively⁴⁵. Avanesov and co-workers⁴⁶ developed a scoring system with CT features; the sensitivity was 82 per cent and the specificity 93 per cent. However, these scoring models have not yet been validated externally. Current studies on surgery versus antibiotics in uncomplicated appendicitis have shown that, on average, 18 per cent of patients considered before surgery to have uncomplicated appendicitis were found to have complicated appendicitis at operation¹⁴. Most studies used imaging to rule out complicated appendicitis, but misclassified about one in six patients. The only study¹⁹ with a low proportion of false negatives for patients with complicated appendicitis in the surgery arm used CT alone, and was probably biased by the preselection of patients eligible for non-operative treatment.

Limitations of the studies included in this review involve the predominantly retrospective designs, which may have resulted in an overestimation of accuracy. Another limitation is the variation in the diagnostic accuracy of the radiologists reading the examinations. For two studies^{26,29} in the meta-analysis, the outcomes of two readers were reported; both reported differences in diagnostic accuracy. Kim et al.⁴⁷ reported on the performance of CT in differentiating between complicated and uncomplicated appendicitis. Their study was not included in the present review, as they used CT scans from an RCT that already had been included³. However, in that more recent report⁴⁷, the sensitivity of the radiologists' assessment for complicated appendicitis ranged from 31 to 81 per cent, and the specificity ranged from 60 to 93 per cent, indicating that the performance of radiologists in detecting complicated appendicitis is likely to vary.

The debate regarding the best strategy to discriminate complicated from uncomplicated appendicitis will probably continue. Scoring systems seem to perform better in ruling out complicated appendicitis, but there are no adequately validated scoring systems yet. Most studies reporting on the effect of non-operative

treatment for uncomplicated appendicitis have included patients based on CT alone. Although CT has a far from perfect, and highly varying, sensitivity, its NPV for complicated appendicitis is high. Ultrasound imaging, CT and MRI all have limitations in discriminating between complicated and uncomplicated appendicitis.

Acknowledgements

W.J.B. and M.D.B. contributed equally to this publication

Disclosure. The authors declare no conflict of interest.

Supplementary material

Supplementary material is available at *BJS Open* online.

References

- Giljaca V, Nadarevic T, Poropat G, Nadarevic VS, Stimac D. Diagnostic accuracy of abdominal ultrasound for diagnosis of acute appendicitis: systematic review and meta-analysis. *World J Surg* 2017;41:693–700
- Replinger MD, Levy JF, Peethumnongsin E, Gussick ME, Svenson JE, Golden SK et al. Systematic review and meta-analysis of the accuracy of MRI to diagnose appendicitis in the general population. *J Magn Reson Imaging* 2016;43:1346–1354
- van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester MA. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. *Radiology* 2008;249:97–106
- Smith MP, Katz DS, Lalani T, Carucci LR, Cash BD, Kim DH et al. ACR Appropriateness criteria[®] right lower quadrant pain—suspected appendicitis. *Ultrasound Quarterly* 2015;31:85–91. doi: 10.1097/RUQ.0000000000000118
- Di Saverio S, Birindelli A, Kelly MD, Catena F, Weber DG, Sartelli M et al. WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. *World J Emerg Surg* 2016;11:34
- Gorter RR, Eker HH, Gorter-Stam MAW, Abis GSA, Acharya A, Ankersmit M et al. Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. *Surg Endosc* 2016;30:4668–4690
- National Surgical Research Collaborative. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. *Br J Surg* 2013;100:1240–1252
- van Rossem CC, Bolmers MD, Schreinemacher MH, van Geloven AA, Bemelman WA. Prospective nationwide outcome audit of surgery for suspected acute appendicitis. *Br J Surg* 2016;103:144–151
- Cuschieri J, Florence M, Flum DR, Jurkovich GJ, Lin P, Steele SR et al. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. *Ann Surg* 2008;248:557–563
- van Dijk ST, van Dijk AH, Dijkgraaf MG, Boermeester MA. Meta-analysis of in-hospital delay before surgery as a risk factor for complications in patients with acute appendicitis. *Br J Surg* 2018; 105:933–945
- Sippola S, Grönroos J, Sallinen V, Rautio T, Nordström P, Rantanen T et al. A randomised placebo-controlled double-blind multicentre trial comparing antibiotic therapy with placebo in the treatment of uncomplicated acute appendicitis: APPAC III trial study protocol. *BMJ Open* 2018;8:e023623

12. Park HC, Kim MJ, Lee BH. Randomized clinical trial of antibiotic therapy for uncomplicated appendicitis. *Br J Surg* 2017;104:1785–1790
13. Livingston EH. Antibiotic treatment for uncomplicated appendicitis really works: results from 5 years of observation in the APPAC trial. *JAMA* 2018;320:1245–1246
14. Rollins KE, Varadhan KK, Neal KR, Lobo DN. Antibiotics versus appendectomy for the treatment of uncomplicated acute appendicitis: an updated meta-analysis of randomised controlled trials. *World J Surg* 2016;40:2305–2318
15. Sallinen V, Akl EA, You JJ, Agarwal A, Shoucair S, Vandvik PO et al. Meta-analysis of antibiotics versus appendectomy for non-perforated acute appendicitis. *Br J Surg* 2016;103:656–667
16. Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. *BMJ* 2012;344:e2156
17. Salminen P, Tuominen R, Paaajanen H, Rautio T, Nordström P, Aarnio M et al. Five-year follow-up of antibiotic therapy for uncomplicated acute appendicitis in the APPAC randomized clinical trial. *JAMA* 2018;320:1259–1265
18. Vons C, Barry C, Maitre S, Pautrat K, Leconte M, Costaglioli B et al. Amoxicillin plus clavulanic acid versus appendectomy for treatment of acute uncomplicated appendicitis: an open-label, non-inferiority, randomised controlled trial. *Lancet* 2011;377:1573–1579
19. Salminen P, Paaajanen H, Rautio T, Nordström P, Aarnio M, Rantanen T et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: the APPAC randomized clinical trial. *JAMA* 2015;313:2340–2348
20. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–396
21. Doria AS, Moineddin R, Kellenberger CJ, Epelman M, Beyene J, Schuh S et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. *Radiology* 2006;241:83–94
22. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536
23. Ali M, Iqbal J, Sayani R. Accuracy of computed tomography in differentiating perforated from nonperforated appendicitis, taking histopathology as the gold standard. *Cureus* 2018;10:e3735
24. Borushok KF, Jeffrey RB Jr, Laing FC, Townsend RR. Sonographic diagnosis of perforation in patients with acute appendicitis. *AJR Am J Roentgenol* 1990;154:275–278
25. Choi YH, Fischer E, Hoda SA, Rubenstein WA, Morrissey KP, Hertford D et al. Appendiceal CT in 140 cases diagnostic criteria for acute and necrotizing appendicitis. *Clin Imaging* 1998;22:252–271
26. Foley TA, Earnest F, Nathan MA, Hough DM, Schiller HJ, Hoskin TL. Differentiation of nonperforated from perforated appendicitis: accuracy of CT diagnosis and relationship of CT findings to length of hospital stay. *Radiology* 2005;235:89–96
27. Miki T, Ogata S, Uto M, Nakazono T, Urata M, Ishibe R et al. Enhanced multidetector-row computed tomography (MDCT) in the diagnosis of acute appendicitis and its severity. *Radiat Med* 2005;23:242–255
28. Tsuboi M, Takase K, Kaneda I, Ishibashi T, Yamada T, Kitami M et al. Perforated and nonperforated appendicitis: defect in enhancing appendiceal wall—depiction with multi-detector row CT. *Radiology* 2008;246:142–147
29. Seo H, Lee KH, Kim HJ, Kim K, Kang S-B, Kim SY et al. Diagnosis of acute appendicitis with sliding slab ray-sum interpretation of low-dose unenhanced CT and standard-dose i.v. contrast-enhanced CT scans. *AJR Am J Roentgenol* 2009;193:96–105
30. Suthikeeree W, Lertdomrongdej L, Charoensak A. Diagnostic performance of CT findings in differentiation of perforated from nonperforated appendicitis. *J Med Assoc Thai* 2010;93:1422–1429
31. Kim SY, Lee KH, Kim K, Kim TY, Lee HS, Hwang S-S et al. Acute appendicitis in young adults: low- versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. *Radiology* 2011;260:437–445
32. Suh SW, Choi YS, Park JM, Kim BG, Cha SJ, Park SJ et al. Clinical factors for distinguishing perforated from nonperforated appendicitis: a comparison using multidetector computed tomography in 528 laparoscopic appendectomies. *Surg Laparosc Percutan Tech* 2011;21:72–75
33. Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med* 2012;366:1596–1605
34. Leeuwenburgh MMN, Wiezer MJ, Wiarda BM, Bouma WH, Phoa SSKS, Stockmann HBAC et al. Accuracy of MRI compared with ultrasound imaging and selective use of CT to discriminate simple from perforated appendicitis. *Br J Surg* 2014;101:e147–e155
35. Liu W, Wei Qiang J, Xun Sun R. Comparison of multislice computed tomography and clinical scores for diagnosing acute appendicitis. *J Int Med Res* 2015;43:341–349
36. Xu Y, Jeffrey RB, Chang ST, DiMaio MA, Olcott EW. Sonographic differentiation of complicated from uncomplicated appendicitis: implications for antibiotics-first therapy. *J Ultrasound Med* 2017;36:269–277
37. Kim HY, Park JH, Lee YJ, Lee SS, Jeon JJ, Lee KH. Systematic review and meta-analysis of CT features for differentiating complicated and uncomplicated appendicitis. *Radiology* 2018;287:104–115
38. Hansen AJ, Young SW, De Petris G, Tessier DJ, Hernandez JL, Johnson DJ. Histologic severity of appendicitis can be predicted by computed tomography. *Arch Surg* 2004;139:1304–1308
39. Duke E, Kalb B, Arif-Tiwari H, Daye ZJ, Gilbertson-Dahdal D, Keim SM et al. A systematic review and meta-analysis of diagnostic performance of MRI for evaluation of acute appendicitis. *AJR Am J Roentgenol* 2016;206:508–517
40. Yilmaz EM, Kapci M, Celik S, Manoglu B, Avcil M, Karacan E. Should Alvarado and Ohmann scores be real indicators for diagnosis of appendicitis and severity of inflammation? *Ulus Trauma Acil Cerrahi Derg* 2017;23:29–33
41. Yesiltas M, Karakas DO, Gokcek B, Hot S, Egin S. Can Alvarado and Appendicitis Inflammatory Response scores evaluate the severity of acute appendicitis? *Ulus Trauma Acil Cerrahi Derg* 2018;24:557–562
42. Deiters A, Drozd A, Parikh P, Markert R, Shim JK. Use of the Alvarado score in elderly patients with complicated and uncomplicated appendicitis. *Am Surg* 2019;85:397–402
43. Kang CB, Li WQ, Zheng JW, Li XW, Lin DP, Chen XF et al. Preoperative assessment of complicated appendicitis through stress reaction and clinical manifestations. *Medicine* 2019;98:e15768

44. Broker ME, van Lieshout EM, van der Elst M, Stassen LP, Schepers T. Discriminating between simple and perforated appendicitis. *J Surg Res* 2012;176:79–83
45. Atema JJ, van Rossem CC, Leeuwenburgh MM, Stoker J, Boermeester MA. Scoring system to distinguish uncomplicated from complicated acute appendicitis. *Br J Surg* 2015;102:979–990
46. Avanesov M, Wiese NJ, Karul M, Guerreiro H, Keller S, Busch P et al. Diagnostic prediction of complicated appendicitis by combined clinical and radiological appendicitis severity index (APSI). *Eur Radiol* 2018;28:3601–3610
47. Kim HY, Park JH, Lee SS, Lee WJ, Ko Y, Andersson RE et al. CT in differentiating complicated from uncomplicated appendicitis: presence of any of 10 CT features versus radiologists' gestalt assessment. *AJR Am J Roentgenol* 2019;213:W218–W227