Cochrane Library

Cochrane Database of Systematic Reviews

Computed tomography for diagnosis of acute appendicitis in adults (Review)

Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jørgensen P

Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jørgensen P.
Computed tomography for diagnosis of acute appendicitis in adults.
Cochrane Database of Systematic Reviews 2019, Issue 11. Art. No.: CD009977.
DOI: 10.1002/14651858.CD009977.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

HEADER 1
ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 4
BACKGROUND 6
OBJECTIVES 8
METHODS 8
RESULTS 10
Figure 1. 11
Figure 2. 13
Figure 3. 14
Figure 4. 19
Figure 5. 21
Figure 6. 23
Figure 7. 24
Figure 8. 26
Figure 9. 27
Figure 10. 29
Figure 11. 32
Figure 12. 34
Figure 13. 36
DISCUSSION 37
AUTHORS' CONCLUSIONS 39
Figure 14. 40
ACKNOWLEDGEMENTS 40
REFERENCES 42
CHARACTERISTICS OF STUDIES 57
DATA 191
Test 1. CT (unenhanced). 192
Test 2. CT (IV contrast). 192
Test 3. CT (oral contrast). 192
Test 4. CT (rectal contrast). 192
Test 5. CT (IV+oral contrast). 192
Test 6. CT (oral+rectal contrast). 192
Test 7. CT (IV+oral+rectal contrast). 192
Test 8. Low-dose CT. 192
Test 9. CT (overall). 192
Test 10. Standard-dose CT. 192
ADDITIONAL TABLES 192
APPENDICES 199
CONTRIBUTIONS OF AUTHORS 203
DECLARATIONS OF INTEREST 203
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 204

[Diagnostic Test Accuracy Review]

Computed tomography for diagnosis of acute appendicitis in adults

Bo Rud[1], Thomas S Vejborg[2], Eli D Rappeport[2], Johannes B Reitsma[3], Peer Wille-Jørgensen[4]

[1]Gastrounit, Surgical Division, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark. [2]Department of Radiology R, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark. [3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands. [4]Department of Surgical Gastroenterology K, Bispebjerg Hospital, Copenhagen NV, Denmark

Contact address: Bo Rud, Gastrounit, Surgical Division, Copenhagen University Hospital Hvidovre, Kettegaards Alle 30, Hvidovre, 2650, Denmark. mail@borud.dk, borud99@gmail.com.

Editorial group: Cochrane Colorectal Group
Publication status and date: New, published in Issue 11, 2019.

Citation: Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jørgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database of Systematic Reviews* 2019, Issue 11. Art. No.: CD009977. DOI: 10.1002/14651858.CD009977.pub2.

ABSTRACT

Background

Diagnosing acute appendicitis (appendicitis) based on clinical evaluation, blood testing, and urinalysis can be difficult. Therefore, in persons with suspected appendicitis, abdominopelvic computed tomography (CT) is often used as an add-on test following the initial evaluation to reduce remaining diagnostic uncertainty. The aim of using CT is to assist the clinician in discriminating between persons who need surgery with appendicectomy and persons who do not.

Objectives

Primary objective

Our primary objective was to evaluate the accuracy of CT for diagnosing appendicitis in adults with suspected appendicitis.

Secondary objectives

Our secondary objectives were to compare the accuracy of contrast-enhanced versus non-contrast-enhanced CT, to compare the accuracy of low-dose versus standard-dose CT, and to explore the influence of CT-scanner generation, radiologist experience, degree of clinical suspicion of appendicitis, and aspects of methodological quality on diagnostic accuracy.

Search methods

We searched MEDLINE, Embase, and Science Citation Index until 16 June 2017. We also searched references lists. We did not exclude studies on the basis of language or publication status.

Selection criteria

We included prospective studies that compared results of CT versus outcomes of a reference standard in adults (> 14 years of age) with suspected appendicitis. We excluded studies recruiting only pregnant women; studies in persons with abdominal pain at any location and with no particular suspicion of appendicitis; studies in which all participants had undergone ultrasonography (US) before CT and the decision to perform CT depended on the US outcome; studies using a case-control design; studies with fewer than 10 participants; and studies that did not report the numbers of true-positives, false-positives, false-negatives, and true-negatives. Two review authors independently screened and selected studies for inclusion.

Data collection and analysis

Two review authors independently collected the data from each study and evaluated methodological quality according to the Quality Assessment of Studies of Diagnostic Accuracy - Revised (QUADAS-2) tool. We used the bivariate random-effects model to obtain summary estimates of sensitivity and specificity.

Main results

We identified 64 studies including 71 separate study populations with a total of 10,280 participants (4583 with and 5697 without acute appendicitis). Estimates of sensitivity ranged from 0.72 to 1.0 and estimates of specificity ranged from 0.5 to 1.0 across the 71 study populations. Summary sensitivity was 0.95 (95% confidence interval (CI) 0.93 to 0.96), and summary specificity was 0.94 (95% CI 0.92 to 0.95). At the median prevalence of appendicitis (0.43), the probability of having appendicitis following a positive CT result was 0.92 (95% CI 0.90 to 0.94), and the probability of having appendicitis following a negative CT result was 0.04 (95% CI 0.03 to 0.05). In subgroup analyses according to contrast enhancement, summary sensitivity was higher for CT with intravenous contrast (0.96, 95% CI 0.92 to 0.98), CT with rectal contrast (0.97, 95% CI 0.93 to 0.99), and CT with intravenous and oral contrast enhancement (0.96, 95% CI 0.93 to 0.98) than for unenhanced CT (0.91, 95% CI 0.87 to 0.93). Summary sensitivity of CT with oral contrast enhancement (0.89, 95% CI 0.81 to 0.94) and unenhanced CT was similar. Results show practically no differences in summary specificity, which varied from 0.93 (95% CI 0.90 to 0.95) to 0.95 (95% CI 0.90 to 0.98) between subgroups. Summary sensitivity for low-dose CT (0.94, 95% 0.90 to 0.97) was similar to summary sensitivity for standard-dose or unspecified-dose CT (0.95, 95% 0.93 to 0.96); summary specificity did not differ between low-dose and standard-dose or unspecified-dose CT. No studies had high methodological quality as evaluated by the QUADAS-2 tool. Major methodological problems were poor reference standards and partial verification primarily due to inadequate and incomplete follow-up in persons who did not have surgery.

Authors' conclusions

The sensitivity and specificity of CT for diagnosing appendicitis in adults are high. Unenhanced standard-dose CT appears to have lower sensitivity than standard-dose CT with intravenous, rectal, or oral and intravenous contrast enhancement. Use of different types of contrast enhancement or no enhancement does not appear to affect specificity. Differences in sensitivity and specificity between low-dose and standard-dose CT appear to be negligible. The results of this review should be interpreted with caution for two reasons. First, these results are based on studies of low methodological quality. Second, the comparisons between types of contrast enhancement and radiation dose may be unreliable because they are based on indirect comparisons that may be confounded by other factors.

PLAIN LANGUAGE SUMMARY

How accurate is computed tomography for the diagnosis of acute appendicitis in adults?

Why is improving the diagnosis of appendicitis important?
The purpose of using computed tomography (CT) in persons with suspected appendicitis is to assist the clinician in differentiating between persons who need surgery with resection of the appendix (appendicectomy) and persons who do not need this procedure.

What is the aim of this review?
The aim of this Cochrane Review was to find out how accurate CT of the abdomen and pelvis is for diagnosing appendicitis in adults. Researchers at Cochrane included 64 studies in the review to answer this question.

What was studied in the review?
A CT-scan can be performed in several ways. Image quality can be improved by using intravenous contrast material, and visualization of the appendix can be better when oral or rectal contrast material is used. CT can also be performed with low-dose radiation. The radiation exposure related to CT may increase lifetime risk of cancer. This Cochrane Review studied the accuracy of the following types of CT: any type of CT, CT according to type of contrast material, and low-dose CT.

What are the main results of this review?
This review included 64 relevant studies that reported results for 71 separate study populations with a total of 10,280 participants. Overall results of these studies indicate that in theory, if CT of any type were to be used in an emergency department in a group of 1000 people, of whom 43% have appendicitis, then:
• an estimated 443 people would have a CT result indicating appendicitis, and of these, 8% would not have acute appendicitis; and
• of the 557 people with a CT result indicating that appendicitis is not present, 4% would actually have acute appendicitis.

Low-dose CT appeared to be as accurate as standard-dose CT for diagnosing appendicitis. CT with intravenous, rectal, or oral and intravenous contrast material appeared to be equally accurate, and more accurate than CT without use of contrast material.

How reliable are the results of the studies in this review?
Among the included studies, the final diagnosis of appendicitis was based on operative findings or microscopic examination of the resected appendix. Among participants who did not have surgery, appendicitis was ruled out by following up to see whether their symptoms resolved without appendicectomy. This is likely to have been a reliable method for deciding whether patients really had appendicitis when follow-up was careful and complete. Unfortunately, this was not so in a substantial proportion of the included studies. In general, some

problems with how the studies were conducted were evident. This may have resulted in CT appearing more accurate than it really is, thereby increasing the number of correct CT results (green rectangles) in the diagram.

To whom do the results of this review apply?
Studies included in the review were carried out mainly in emergency departments. Appendicitis was suspected in all participants following clinical examination and blood testing. Included studies evaluated a wide range of types of CT. Participants' average age ranged from 25 to 46 years across studies, and the percentage of women varied between 26% and 100%. The percentage of study participants with a final diagnosis of appendicitis varied between 13% and 92% across studies (average, 43%).

What are the implications of this review?
CT is an accurate test that is likely to assist clinicians in treating persons with possible appendicitis. Results of this review indicate that the chance of a clinician wrongly diagnosing acute appendicitis appears to be low (8% among those whose CT results suggest they have appendicitis). The chance of missing a diagnosis of appendicitis is also low (4% among those whose CT results suggest they do not have appendicitis).

How up-to-date is this review?

The review authors searched for and included studies published up to 16 June 2017.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table

Population	Adults (> 14 years of age) with suspected acute appendicitis based on history, physical examination, and/or blood tests					
Settings	Emergency and Radiology Departments in secondary and tertiary care settings					
Index test	Computed tomography of the abdomen					
Reference standard	Histological examination of the resected appendix or intraoperative findings in persons who had surgery. Clinical follow-up for persons who did not have surgery					
Target condition	Acute appendicitis					
Number of studies	64 studies including 71 separate study populations with a total of 10,280 participants - 4583 with and 5697 without acute appendicitis					
Methodological concerns	The methodological quality was generally poor, particularly with respect to the reference test and the flow and timing domains. For these domains, few studies were at low risk of bias. Differential verification was used in most studies because some of the participants with suspected acute appendicitis did not have surgery. Clinical follow-up for these participants was inadequate, incomplete, or poorly described in most studies					
Results	Number of studies (study populations)[a]	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Prevalence of appendicitis (25% percentile 50% percentile 75% percentile)[b]	Post-test probability following a positive CT outcome (95% CI)	Post-test probability following a negative CT outcome (95% CI)
CT overall	64 (71)	0.95 (0.93-0.96)	0.94 (0.92-0.95)	0.32 0.43 0.58	0.88 (0.85-0.90) 0.92 (0.90-0.94) 0.96 (0.94-0.96)	0.02 (0.02-0.03) 0.04 (0.03-0.05) 0.07 (0.05-0.09)
Unenhanced CT	19 (19)	0.91 (0.87-0.93)	0.94 (0.90-0.96)	0.32 0.43 0.58	0.87 (0.82-0.92) 0.92 (0.88-0.95) 0.95 (0.93-0.97)	0.04 (0.03-0.06) 0.07 (0.05-0.09) 0.12 (0.09-0.16)
CT with intravenous	17 (18)	0.96 (0.92-0.98)	0.93 (0.90-0.95)	0.32 0.43	0.87 (0.82-0.90) 0.91 (0.88-0.94)	0.02 (0.01-0.04) 0.03 (0.02-0.06)

contrast enhancement				0.58	0.95 (0.93-0.96)	0.06 (0.03-0.11)
CT with rectal contrast enhancement	9 (9)	0.97 (0.93-0.99)	0.95 (0.90-0.98)	0.32	0.91 (0.81-0.96)	0.02 (0.01-0.04)
				0.43	0.94 (0.87-0.97)	0.03 (0.01-0.06)
				0.58	0.97 (0.93-0.99)	0.05 (0.02-0.10)
CT with oral contrast enhancement	7 (7)	0.89 (0.81-0.94)	0.94 (0.90-0.97)	0.32	0.88 (0.81-0.93)	0.05 (0.03-0.09)
				0.43	0.92 (0.87-0.96)	0.08 (0.04-0.14)
				0.58	0.96 (0.92-0.98)	0.14 (0.08-0.22)
CT with oral and intravenous contrast enhancement	15 (15)	0.96 (0.93-0.98)	0.94 (0.92-0.96)	0.32	0.89 (0.85-0.92)	0.02 (0.01-0.03)
				0.43	0.93 (0.90-0.95)	0.03 (0.02-0.05)
				0.58	0.96 (0.94-0.97)	0.05 (0.03-0.09)
Low-dose CT	7 (8)	0.94 (0.90-0.97)	0.94 (0.91-0.96)	0.32	0.88 (0.82-0.92)	0.03 (0.02-0.05)
				0.43	0.92 (0.88-0.95)	0.04 (0.02-0.08)
				0.58	0.96 (0.93-0.97)	0.08 (0.04-0.13)
Conclusion	Sensitivity and specificity of CT for diagnosing acute appendicitis in adults are high. Unenhanced standard-dose CT appears to have lower sensitivity than standard-dose CT with intravenous, rectal, or oral+intravenous contrast enhancement. Use of different types of contrast enhancement or no enhancement does not appear to affect specificity. Differences in sensitivity and specificity between low-dose and standard-dose CT appear to be negligible. The results of this review should be interpreted with caution for 2 reasons. First, the results are based on studies of low methodological quality. Second, the comparisons between types of contrast enhancement and radiation dose may be unreliable because they are based on indirect comparisons that may be confounded by other factors					

CI: confidence interval.
CT: computed tomography.
[a]In five studies, participants were randomly allocated to two CT-protocols, and in another study to three CT-protocols. These protocols differed with respect to contrast enhancement and radiation dose. This generated seven additional study populations, which were included as separate studies in the meta-analyses.
[b]The distribution of the prevalence of appendicitis was roughly similar in the included studies across subgroups. Therefore, to facilitate comparison of post-test probabilities between subgroups, these probabilities were calculated for the 25%, 50%, and 75% percentiles of prevalence for all 71 study populations.

BACKGROUND

Target condition being diagnosed

Acute appendicitis (appendicitis) is a common cause of abdominal pain, with an incidence of around 1 per 1000 per year (Hall 2010), and with a lifetime risk of 7% to 9% in developed countries (Anderson 2012). Appendicitis is an inflammation of the vermiform appendix, but the etiology of the inflammation and its progression remains poorly understood. Obstruction of the appendix lumen by a fecalith, stool, or caecum tumour may elicit appendicitis, but it appears that genetic and environmental factors are also important for the development of appendicitis (Sadr 2009). The characteristic medical history is one of central abdominal pain followed by nausea, vomiting, anorexia, and migration of pain to the right iliac fossa. Clinical and laboratory findings include mild pyrexia, exacerbation of pain on coughing, maximum tenderness in the right lower fossa, and elevated white blood cell count and C-reactive protein concentration (Bhangu 2015; Humes 2006; Paulson 2003; Wagner 2009). Migration of pain and signs of peritoneal irritation (guarding, percussion, and rebound tenderness) appear to be the most reliable clinical features (Andersson 2004), but these features may be absent in up to 70% of patients with suspected appendicitis (Lameris 2009). Hence, the diagnosis based on history, clinical findings, and laboratory results is often difficult, particularly in women of childbearing age, because persons with a wide range of intra-abdominal and pelvic pathology may have a similar clinical presentation. The treatment of choice for most persons is appropriate supportive therapy followed by expedient surgical excision of the appendix (appendicectomy). Based on intraoperative findings, appendicitis is classified as simple or complex (gangrenous or perforated appendix with or without abscess formation). Accordingly, the clinical spectrum of appendicitis is wide-ranging - from uncomplicated disease that may be self-limiting to severe complicated disease with generalised peritonitis, sepsis, abscess formation, bowel obstruction, and rarely death (Blomqvist 2001). Over the past decade, several randomised controlled trials (RCTs) have shown that antibiotic therapy can be successful in 70% to 75% of persons with uncomplicated appendicitis on computed tomography (CT); remaining persons will need subsequent appendicectomy within the following year (Salminen 2015; Vons 2011). Laparoscopic appendicectomy is generally recommended over open appendicectomy due to less postoperative pain, lower incidence of surgical site infection, and reduced length of hospital stay (Di Saverio 2016). Conservative therapy with antibiotics and percutaneous drainage is recommended for persons presenting with an appendiceal abscess (Andersson 2007).

Index test(s)

Computed tomography (CT) is an imaging method that uses a series of X-ray measurements from different angles and computer software to generate cross-sectional images of the body. CT of the abdomen and pelvis has been used since the late 1980s to assess persons with suspected appendicitis (Balthazar 1986). With modern multi-slice CT or multi-detector row CT (MDCT), an abdominopelvic CT-scan is acquired in a few seconds once the patient is positioned. The most common approach is to visualise the entire abdomen and pelvis via thin-section images (≤ 5 mm), but protocols focusing on the lower abdomen and pelvis are also used to reduce radiation exposure at the expense of missing disease processes in the upper abdomen (Brown 2008). Enhancement by intravenous (IV), oral, or rectal contrast material is often used to optimise image quality and aid visualisation of the appendix; however, use of oral contrast is time-consuming, rectal contrast is uncomfortable for the patient, and IV contrast may cause allergic reactions. Moreover, it is controversial whether contrast enhancement is needed for the radiological diagnosis of appendicitis (Neville 2009); hence, no consensus has been reached about the most appropriate CT-protocol for persons with suspected appendicitis (Drake 2014; Tan 2017). The introduction of 16-MDCT in 2002 enabled high-quality multi-planar re-formations with coronal and sagittal cross-sectional images that facilitate identification of the appendix (Paulson 2005). CT criteria used in most studies to detect an inflamed appendix have included an appendiceal diameter exceeding 6 mm and the finding of periappendiceal inflammation, an appendicolith, or thickening of the caecal wall (Terasawa 2004). Radiation exposure (effective dose) related to contrast-enhanced abdominopelvic CT varies between 8 and 16 mSv (Smith-Bindman 2009; Yun 2017), which roughly corresponds to three to six years of background radiation in most parts of the world. The estimated increased lifetime risk of cancer following an abdominopelvic CT-scan is 0.02% to 0.14%; the lower the age at the time of CT-scan, the higher the estimated risk (Brenner 2007). Many studies have evaluated the diagnostic accuracy of different types of CT (CT-protocols) for appendicitis; accuracy has been high in previous meta-analyses with summary estimates of sensitivity and specificity above 0.9 (Al-Khayal 2007; Anderson 2005; Dahabreh 2015; Hlibczuk 2010; Terasawa 2004; Weston 2005; Xiong 2015). Several recent studies have demonstrated that low-dose CT (effective dose around 2 mSv) is as accurate as standard-dose CT for diagnosing appendicitis (Yun 2017). By contrast, the accuracy of CT in separating simple from complex appendicitis is more heterogeneous, with estimates of sensitivity and specificity ranging from 0.28 to 0.95, and from 0.88 to 1.0, respectively (Foley 2005; Horrow 2003; Oliak 1999; Suh 2011).

Clinical pathway

Adult persons admitted with acute pain in the right lower abdomen or possible appendicitis are routinely assessed by a general surgeon or an emergency physician via history-taking, physical examination, urinalysis, and blood testing, including a differential white blood cell count and C-reactive protein (CRP) concentration. In women of childbearing age, a gynaecological examination is performed and blood tests or urinalysis includes a pregnancy test (human chorionic gonadotropin analysis) (Humes 2006). Based on weighting and integration of collected information, the clinician must decide the appropriate course of action. If the risk of appendicitis is considered low, the clinician may decide on discharge; conversely, if the risk is high, the clinician will plan to perform surgery. If the risk is intermediate due to an equivocal clinical presentation, the clinician is likely to perform imaging tests or diagnostic laparoscopy, or to admit for observation. The proportion of persons with suspected appendicitis who have imaging tests varies considerably between settings. Assessment of risk of appendicitis may be subjective, or it may be based on one of several clinical decision rules developed to assist the clinician in decision-making. Such decision rules include the Alvarado Score (Alvarado 1986), the Appendicitis Inflammatory Response Score (Andersson 2008), the Adult Appendicitis Score (Sammalkorpi 2014), and the Raja Isteri Pengiran Anak Saleha Appendicitis (RIPASA) Score (Chong 2010). Imaging tests often used include ultrasonography (US), CT, or sequential US and CT (i.e. CT following

inconclusive findings on US). Magnetic resonance imaging (MRI) is typically reserved for children and pregnant women (Di Saverio 2016). The use of CT is common in the USA, where more than 90% of persons have CT before appendicectomy in some regions (Coursey 2010; Drake 2014). In England, the corresponding proportion was 13% in 2012 (National Surgical Research Collaborative 2013). In the Netherlands, almost all persons who undergo appendicectomy have preoperative sequential US and CT (van Rossem 2016). If the diagnosis of appendicitis is confirmed by imaging tests, most persons proceed to surgery. If the diagnosis is not confirmed, persons may be discharged or admitted for observation. Among the elderly with suspected appendicitis, CT is often performed to rule out conditions such as right-sided colon cancer and diverticulitis.

Role of index test(s)

CT serves as an add-on test to reduce diagnostic uncertainty following clinical evaluation, blood testing, and urinalysis in persons with suspected appendicitis. If accurate, CT can play an important role in reducing both unnecessary surgery and delay of surgery. When appendicitis is not confirmed by CT, CT images are often helpful for diagnosing other causes of abdominal pain, such as cholecystitis, diverticulitis, renal calculi, epiploic appendagitis, bowel obstruction, and gynaecological conditions. Historically, the negative appendicectomy rate (NAR) for persons operated on for acute appendicitis has exceeded 20% due to the low accuracy of clinical assessment and a low threshold to perform surgery to avoid potential disease progression through perforation and abscess formation (Lewis 1975; Velanovich 1992). The NAR is the proportion of resected appendices without histological evidence of inflammation out of all resected appendices. Along with the perforation rate, NAR is an often used indicator of the accuracy of the preoperative evaluation of persons with suspected appendicitis. A systematic review with meta-analysis of results from 20 studies found a significantly lower NAR in persons who had clinical evaluation and preoperative CT compared to those who had clinical evaluation only (9% vs 17%, respectively; P = 0.001; Krajewski 2011). The time from emergency department to operating room was examined in 10 studies, and the mean waiting time was longer for those who had preoperative CT than for those who did not (800 vs 468 minutes; no statistical analysis due to lack of standard deviations), but no statistical difference in summary estimates of perforation rates was evident. Additionally, two studies from the USA have demonstrated a drop in NAR from 23% to 24% to 2% to 3% from the 1990s to 2007, coinciding with an increase in the use of preoperative CT from 10% to 20% to more than 85% (Raja 2010; Raman 2008). Results from other studies indicate that the effects of preoperative CT on NAR are limited to women younger than 45 years, whereas there is little or no effect on men (Coursey 2010; Wagner 2008). The accuracy of clinical assessment alone versus clinical assessment and CT has been compared in three RCTs with a total of 400 participants. The sensitivity of the former was 1.0 for all studies compared to 0.90 to 0.94 for the latter. Conversely, specificity was generally lower for clinical assessment alone (0.73 to 0.88) compared to clinical assessment and CT (0.93 to 1.0) (Hong 2003; Lopez 2007; Walker 2000). Two of the studies concluded that the accuracy of clinical assessment and CT was not superior to the accuracy of clinical assessment alone; the third study reached the opposite conclusion (Walker 2000).

Alternative test(s)

Alternative add-on tests used to reduce diagnostic uncertainty following clinical evaluation are ultrasonography (US), magnetic resonance imaging (MRI), and diagnostic laparoscopy (DL). US has been used since the 1980s in persons with suspected appendicitis (Rybkin 2007); the main advantages are that US is free from radiation exposure, widely available, quick to perform, and cheap. Refinements in US technology and use of Doppler sonography and the graded compression technique have improved both visualisation of the appendix and accuracy (Birnbaum 2000). However, the utility of US is hampered because the appendix can be difficult to visualise even for experienced radiologists due to obesity and overlying bowel gas, resulting in inconclusive examinations in up to 30% to 50% of cases (D'Souza 2015; Leeuwenburgh 2013; Poletti 2011; Poortman 2009). Several meta-analyses have compared the accuracy of US and CT (Doria 2006; Terasawa 2004; van Randen 2008), revealing lower sensitivity and specificity for US compared to CT. In the most recent meta-analysis, summary sensitivity and specificity were 0.85 and 0.90, respectively, for US, and 0.96 and 0.96, respectively, for CT (Dahabreh 2015). Nevertheless, in some settings, US is used as the primary imaging test in most persons with suspected appendicitis, and CT is primarily reserved for persons with inconclusive US findings (van Rossem 2016).

Over the past 10 years, MRI has been increasingly used for assessment of persons with possible appendicitis. Advances in MRI hardware and software as well as in radiologists' expertise have led to increasing accuracy and quicker scan times (Leeuwenburgh 2012). Although MRI offers disadvantages such as high costs, long acquisition times, and limited availability, the features of high accuracy and non-ionising radiation make MRI particularly attractive for pregnant women and children with an inconclusive US examination (Basaran 2009). Summary estimates of sensitivity and specificity in the currently most comprehensive meta-analysis of results from 30 studies were 0.96 (95% confidence interval (CI) 0.95 to 0.97) and 0.96 (95% CI 0.95 to 0.97), respectively (Duke 2016). Summary estimates were similar in subgroups of children and pregnant women. A recent study used a paired design to compare MRI and CT in participants older than 11 years (Repplinger 2018). Sensitivity and specificity were 0.97 and 0.81 for unenhanced MRI, and 0.98 and 0.90 for IV contrast-enhanced CT, respectively. The difference in specificity was statistically significant. Another paired study compared the accuracy of IV contrast-enhanced CT and unenhanced MRI in persons with suspected appendicitis following a negative or inconclusive US examination (Leeuwenburgh 2012). Sensitivity and specificity were 0.98 and 0.88 for MRI, and 0.97 and 0.91 for CT, respectively. The difference in specificity was not statistically significant.

Diagnostic laparoscopy (DL) is a surgical procedure performed under general anaesthesia by which two or three cannulas are inserted through the abdominal wall after pneumoperitoneum with carbon dioxide has been established. A laparoscope and a grasper are inserted though the cannulas, loops of small bowel are swept away from the right lower quadrant, and the appendix is visualised. If the appendix appears inflamed, it is resected; if it appears normal, other causes of abdominal pain are sought. It remains controversial whether a macroscopically normal looking appendix should be resected or left in situ (Bijnen 2003; Grunewald 1993; Strong 2015; Teh 2000; van den

Broek 2001). DL is used more often in European countries than in the USA, where CT is the most commonly used add-on test following clinical evaluation (Di Saverio 2016; Jaunoo 2012; National Surgical Research Collaborative 2013). A recent review included 54 studies evaluating the accuracy of diagnostic laparoscopy; median sensitivity and specificity were 1.00 and 0.89, respectively (Dahabreh 2015). However, estimates showed wide variability, with sensitivity ranging from 0.37 to 1.0 (interquartile range 0.95 to 1.0), and specificity ranging from 0 to 1.0 (interquartile range 0.73 to 1.0). Complications of DL appear to be infrequent (< 2% in most studies); however in many studies, it was difficult to distinguish complications related to the diagnostic phase of laparoscopy from complications related to the therapeutic phase (appendicectomy). The most common complications were wound infection, postoperative ileus, deep venous thrombosis, haematoma, and intra-abdominal infection (Dahabreh 2015).

Rationale

Assessment of persons with suspected appendicitis is a common and often difficult task for emergency physicians and general surgeons. Imaging tests are frequently used when the diagnosis is uncertain following clinical examination, blood testing, and urinalysis. The magnitude and importance of this assessment task are reflected by the fact that appendicectomy is the most frequently performed abdominal emergency procedure, with approximately 50,000 and 300,000 appendicectomies performed annually in the UK and the USA, respectively (Hospital Episode Statistics 2015; Weiss 2014). As part of the ongoing effort to develop an evidence-based algorithm for the treatment of persons with suspected appendicitis, it is important to systematically review the accuracy of these imaging tests. Ideally, such a review should summarise and compare the accuracy of US, CT, and MRI, and the sequential use of these tests; however, the resources needed to perform such a review are extensive. Because CT appears to be the imaging test used most often (Jaunoo 2012), we limited our task to reviewing the accuracy of CT as a first-line imaging test in adults and exploring differences in accuracy between CT-protocols defined by the use of contrast enhancement and radiation dose. We excluded studies in children because US is usually the first-line imaging test used in children, and CT is reserved for those with negative or inconclusive US findings to reduce radiation exposure (Frush 2009; Hernanz-Schulman 2010; Strouse 2010). In our view, the methodological issues related to sequential use of imaging tests in children with suspected appendicitis require special attention in a separate review. Other Cochrane Review author teams are currently engaged in reviews of the accuracy of MRI and US for appendicitis.

OBJECTIVES

Primary objective

Our primary objective was to evaluate the accuracy of CT for diagnosing appendicitis in adults with suspected appendicitis.

Secondary objectives

Our secondary objectives were to compare the accuracy of contrast-enhanced versus non-contrast-enhanced CT, to compare the accuracy of low-dose versus standard-dose CT, and to explore the influence of CT-scanner generation, radiologist experience, degree of clinical suspicion of appendicitis, and aspects of methodological quality on diagnostic accuracy.

METHODS

Criteria for considering studies for this review

Types of studies

We included prospective studies comparing the results of CT to the results of a reference standard test for appendicitis. We excluded studies with a case-control design and studies with fewer than 10 participants. We considered studies in which all participants had histologically verified appendicitis as irrelevant because such studies cannot estimate specificity. In cases of duplicate publications, we considered the study report with the largest number of participants or the most information as the primary study report. We applied no language restrictions. We excluded studies using retrospectively collected data to reduce potential bias from partial verification.

Participants

We included studies in adults (> 14 years of age) with suspected appendicitis based on history, physical examination, and/or blood testing. We accepted authors' definitions of suspected appendicitis and applied no restrictions regarding the degree of suspicion of appendicitis. We excluded studies recruiting only pregnant women, as well as studies in persons with abdominal pain at any location and no particular suspicion of appendicitis. We also excluded studies in which all participants had US before CT, and the decision to perform CT depended on the outcome of US. In the protocol, we accepted studies with a mixed adult-paediatric population if the paediatric fraction accounted for 10% or less of the group. We planned to contact study authors with a request for results for the adult subgroup when more than 10% of participants were younger than 15 years, but this turned out to be not feasible. Therefore, we decided to include studies with mixed adult-paediatric populations, and we planned sensitivity analyses to explore whether summary sensitivity and specificity differed in such studies compared to studies including only adults (see Differences between protocol and review).

Index tests

Index tests included a sequential or helical abdominopelvic CT-scan whereby the interpreter was assessing the appendix and its surroundings for signs of appendicitis. We applied no restrictions related to image acquisition, CT-scanner generation, the part of the abdomen included in the scan (lower vs entire abdomen), radiation dose, or the use of enhancement by IV, oral, or rectal contrast material. We included no comparator tests.

Target conditions

The target condition was acute appendicitis. We did not distinguish between simple and complex appendicitis. We excluded studies evaluating the accuracy of CT for differentiating between simple and complex appendicitis.

Reference standards

We included studies that used one of the following two reference standards.

- Histological examination of the removed appendix as well as clinical follow-up of participants who did not have surgery.

- Laparoscopic assessment of the appendix by the surgeon as inflamed or normal, as well as clinical follow-up of participants who did not have surgery.

We included studies in which all participants had surgery if intraoperative assessment or histological examination was used as the reference standard. We also included studies that combined the two reference standards because only macroscopically inflamed appendices were resected and examined histologically. We considered intraoperative assessment by laparotomy and laparoscopy as equal. As stated above, we found wide variation in estimates of sensitivity and specificity for the laparoscopic appendix assessment when histological assessment was used as the reference standard, and whether a normal looking appendix should be resected or left in situ in persons undergoing laparoscopy for suspected appendicitis remains controversial. For this reason, we decided to consider laparoscopic assessment as a legitimate reference standard for appendicitis. We performed a sensitivity analysis to explore the potential consequences thereof. This analysis was not planned in the protocol (see Differences between protocol and review).

Search methods for identification of studies

Electronic searches

We searched MEDLINE and Embase via OVID by using an electronic search strategy that combines indexing terms and text words to capture the index test and the target disease. We developed our search strategy in collaboration with the medical information specialist of the Colorectal Cancer Group. We applied no filters in our electronic searches to target diagnostic test accuracy studies. We have presented our search strategies for MEDLINE in Appendix 1, and for Embase in Appendix 2. We performed the latest update of these searches on 16 June 2017. We also searched the Science Citation Index for study reports that had cited the included studies. We did not restrict studies on the basis of language or publication status.

Searching other resources

We screened the reference lists of included studies and existing systematic reviews for relevant studies.

Data collection and analysis

Selection of studies

Two review authors independently applied the selection criteria to the titles and abstracts of study reports identified by the search strategy. If the decision to exclude a study could not be made on the basis of the title and the abstract, we retrieved the entire study report for assessment. We based the final decision on inclusion on the entire study report. We resolved disagreements between review authors by discussion, or if necessary, by consultation with a third review author. We contacted study authors when information was insufficient to indicate whether a study could be included.

Data extraction and management

Two review authors independently extracted information from included studies using a data collection form. We collected the following information: country, publication language, selection criteria, recruitment procedure, study design, clinical setting, and age and gender distribution. For each study, we noted if participants were recruited regardless of the suspicion of appendicitis, or if recruitment was limited to those with intermediate suspicion due to an equivocal presentation. If all participants had surgery, we classified the degree of suspicion as high. For the index test, we collected information on CT manufacturer, model name, CT-scanner generation (sequential/ helical, single slice/ multi-slice), slice thickness, slice interval, voltage, mAs level, use of multi-planar reformations, use of contrast enhancement, use of a low-dose protocol, radiologist experience, criteria for CT diagnosis of appendicitis, and whether CT was compared to other tests. We also extracted counts of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) CT assessments. Finally, we collected information to support the assessment of methodological quality, particularly features related to the reference standard and patient flow. We piloted the data collection form on five studies assessing the accuracy of CT for appendicitis in children. We contacted study authors if information needed for quantitative analyses was unclear or was not reported.

Assessment of methodological quality

We used the Quality Assessment of Studies of Diagnostic Accuracy - Revised (QUADAS-2) tool to assess methodological quality. To promote consistent assessments, we developed a rating guideline with operational criteria for answering signalling questions and assessing risk of bias and concern regarding applicability (Appendix 3). Two review authors independently applied the QUADAS-2 tool and resolved disagreements by discussion. We piloted our adaptation of the QUADAS-2 tool on five studies assessing the accuracy of CT for appendicitis in children. We have presented the outcome of the methodological quality assessment graphically in standard figures. We explored the influence of bias risk on summary estimates of sensitivity and specificity in sensitivity analyses when feasible.

Statistical analysis and data synthesis

We used the bivariate random-effects model to summarise sensitivity and specificity because we anticipated little variation between studies in the CT features that were used to diagnose appendicitis (Reitsma 2005). We performed an overall meta-analysis with results from all studies regardless of contrast enhancement and radiation dose. If studies reported results for two or more independent study populations (i.e. randomised studies), we included the results for each study population in the analyses. In case accuracy analyses were reported for several CT criteria (i.e. thresholds), we focused on the criterion that conferred the highest degree of homogeneity with other studies. If results were reported for several observers without overall estimates of sensitivity and specificity, we calculated average values across observers for TP, FP, FN, and TN and rounded them to integers. To present and visually explore the variation between studies in sensitivity and specificity, we plotted study results in forest plots and in receiver operating characteristic (ROC) plots. For each analysis, we calculated a 95% prediction region around the summary estimate from the parameters of the bivariate model and added it to the plot. This region covers the range of sensitivity and specificity that would be expected in 95% of future large studies if it is assumed that the statistical model is adequate. We calculated summary likelihood ratios from summary estimates of sensitivity and specificity. We also calculated post-test probabilities for appendicitis following positive and negative CT results for the 25%, 50%, and 75% percentiles of prevalence in the included studies.

In subgroup analyses, we explored and compared the accuracy of CT according to types of contrast enhancement (IV, oral, rectal, IV and oral) using unenhanced CT as the reference. We also compared the accuracy of low-dose and standard-dose CT (this subgroup analysis was not planned in the protocol). In the subgroup analyses, we applied the following rules if several CT-protocols were used in the same study.

- If the CT-protocol differed in 20% of participants or less, we analysed the study according to the CT-protocol used in the majority of persons.
- If the CT-protocol differed in more than 20% of participants, we contacted study authors to request subgroup data. If we received no reply from study authors, we excluded the study from the subgroup analysis.

We performed meta-regression analyses to explore potential sources of heterogeneity (see below). We performed these analyses by adding one covariate at a time to the bivariate model. We used a likelihood ratio test to compare nested models with and without covariates and to test whether summary sensitivity and specificity differed between groups. If the number of studies made it meaningful to add parameters to the models, we tested whether the assumption of equal variances for the random-effects model across groups was reasonable. Fitting models with separate variances for the random-effects model for each group did not improve the fit of any of the models ($P > 0.12$; likelihood ratio test), hence we used equal variances for the random-effects model in all analyses. Using parameter estimates from the bivariate model, we calculated absolute differences in summary sensitivity and specificity between different types of contrast enhancement and unenhanced CT. We also calculated these differences between low-dose and standard-dose CT. We calculated a 95% confidence interval for these differences by using the delta method. We used the *metandi, xtmelogit,* and *nlcom* commands in Stata version 13 (Stata-Corp, College Station, Texas, USA) to perform the analyses.

Investigations of heterogeneity

We explored the following study characteristics as sources of heterogeneity.

- CT-scanner generation: number of detector rows fewer than 16 versus equal to or greater than 16.
- Assessment by senior radiologist versus another individual.
- Participants with intermediate suspicion of appendicitis due to an equivocal presentation versus participants with any suspicion of appendicitis (In the protocol, this analysis was planned as a sensitivity analysis of studies in participants with intermediate suspicion).

Sensitivity analyses

We performed sensitivity analyses to explore the effects of methodological quality on summary estimates of sensitivity and specificity. We implemented these analyses as a subgroup analysis in studies with low risk of bias across the four domains in QUADAS-2 (in the protocol, it was planned to investigate the impact of each of the four domains in meta-regression analyses).

We also performed a sensitivity analysis to explore whether inclusion of studies with a mix of paediatric and adult participants affected the summary estimates. Moreover, we explored whether summary estimates were affected by the inclusion of studies that used laparoscopic assessment of the appendix as a reference standard. Finally, we explored the impact of selecting different analyses from paired studies that reported two or more analyses in the same study population. These analyses were not planned in the protocol.

Assessment of reporting bias

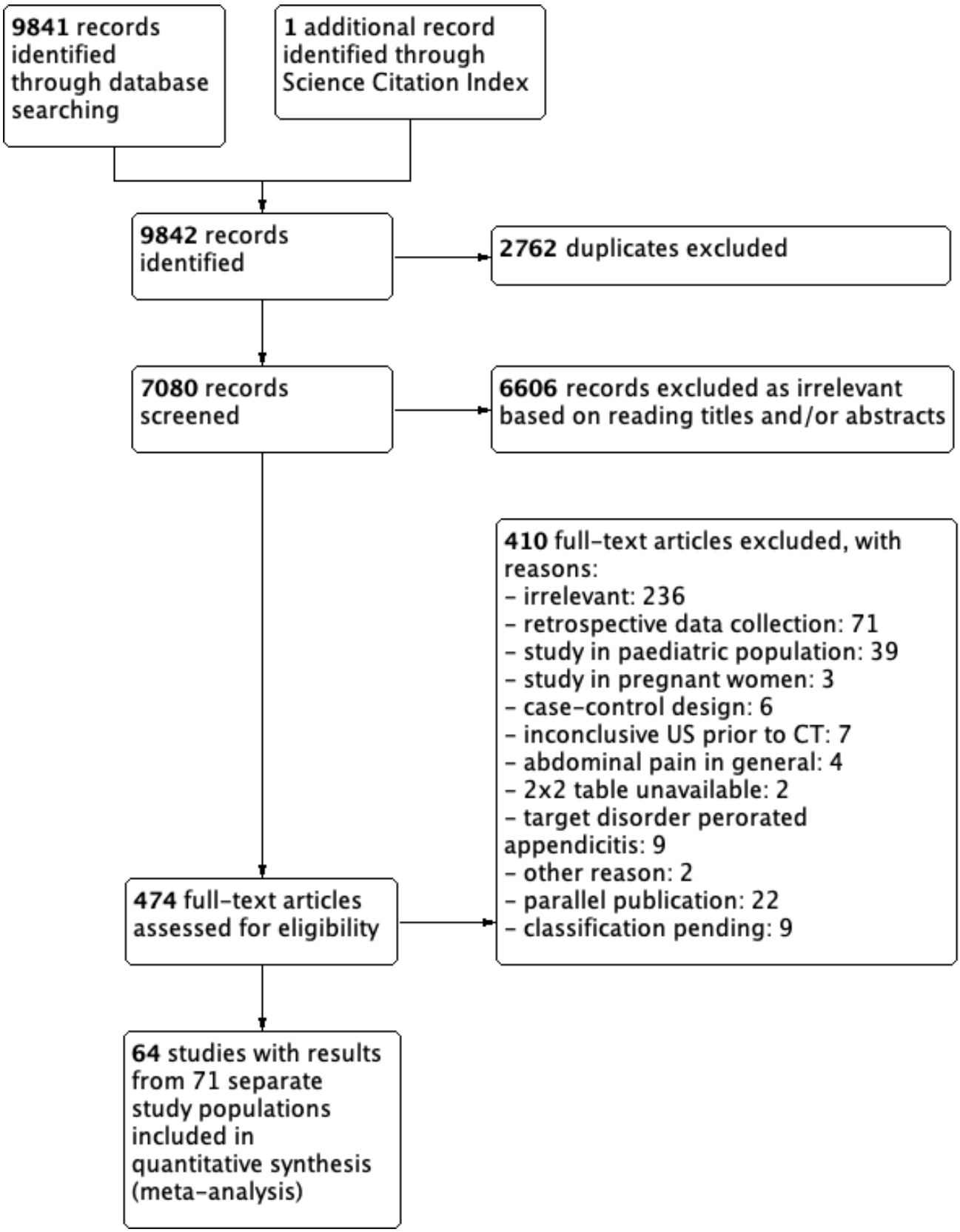
We performed no assessment of reporting bias.

RESULTS

Results of the search

Through our electronic search of MEDLINE and Embase, we identified 9841 references; 2762 of these were duplicates. Science Citation Index provided one additional reference. We excluded 6606 irrelevant references after reading titles and abstracts, and we collected the full text of 474 articles for further assessment. Of these, 236 did not report a diagnostic accuracy study of CT in persons with suspected appendicitis, and we excluded 174 for the reasons stated in Figure 1. Sixty-four studies complied with the selection criteria, and these studies provided data for the review. We contacted the corresponding authors of 26 studies; ten replied, and nine provided supplementary information (Holloway 2003; Jo 2010; Keyzer 2004; Ozturk 2014; Repplinger 2015; Scott 2015; Sim 2013; Tan 2015; Uzunosmanoglu 2017).

Figure 1. Study flow diagram. CT: computed tomography. US: ultrasonography.



Characteristics of included studies

In 55 of the 64 included studies, the outcome of a single CT-protocol was compared to the result of the reference standard. Six studies randomly allocated participants to have one of two CT-protocols (Hekimoglu 2011; Kepner 2012; Keyzer 2009; Kim 2012; Mittal 2004), or to have one of three CT-protocols (Hershko 2007). Four studies compared different CT-protocols in the same participants, three studies compared two protocols (Jacobs 2001; Keyzer 2004; Platon 2009), and one study compared four CT-protocols in each of two randomised groups (Keyzer 2009). Hence, the review includes 80 analyses of accuracy from 71 separate study populations with a total of 10,280 participants (4583 with and 5697 without acute appendicitis). The median number of participants in the 71 separate study populations was 100, with interquartile range 65 to 157, and range 26 to 738.

All studies were reported in full-text publications except two. One was published as a letter to the editor (Cougard 2002), and the other was published as a conference abstract (Repplinger 2015). The authors of the latter provided an unpublished full-text manuscript (Repplinger 2018). The publication language was English in 58 studies, French in two studies, and Spanish, Turkish, Russian, or German in four studies. The studies were performed in 22 countries; 30 studies were performed in the USA. Three studies were multi-centre studies conducted at two (in't Hof 2004), two (Kim 2008), and six participating centres (Atema 2015).

The accuracy of CT was compared to the accuracy of US in 13 studies, to clinical decision rules or clinical assessments in nine studies, to MRI in one study, and to CT conditional on US results in one study. These were randomised trials or paired diagnostic accuracy studies.

Settings and features of the study populations

The clinical settings were emergency departments, general surgery departments, and radiology departments in 34, one, and 15 studies, respectively. In 14 studies, the setting was unclear. All studies were performed in secondary or tertiary care hospitals. Among the 71 separate study populations, the median prevalence of appendicitis was 0.43, with interquartile range 0.32 to 0.58, and range 0.13 to 0.92. The gender distribution was reported for 67 study populations, and the median percentage of women was 55%, with interquartile range 49% to 61%, and range 26% to 100%. The median or mean age of study participants was available for 59 study populations, and the median of these was 33 years, with interquartile range 30 to 38 years, and range 25 to 46 years. Participants younger than 15 years of age were included in 30 study populations. The percentage of paediatric participants was available for five of these populations; it ranged from 3% to 15%. The authors of one study provided subgroup results for participants aged 15 years or older (Sim 2013). All participants were 15 years of age or older in 39 study populations, and two studies provided no information about the age distribution (Holloway 2003; Megibow 2002). Based on available information, we considered it most likely that the latter two studies included adults or a mix of adults and children.

No study reports mentioned that a course of antibiotic therapy was used as an alternative to surgery, or that antibiotic therapy in participants with a negative CT result was a reason for exclusion.

CT-scanners and CT-protocols

A single CT-scanner was used in 50 studies, two were used in 12 studies, three were used in one study, and six were used in a multi-centre study at six centres (Atema 2015). Hence, overall 83 CT-scanners were used in the included studies. Of these, 68 were helical, seven were non-helical, and eight were not described as helical or non-helical. Of the 68 helical CT-scanners, 22 were single detector row devices, 35 were multi-detector row devices, and it was unclear for 11 CT-scanners if they were single or multi-detector row devices. For the multi-detector row CT-scanners, the number of detector rows was 2, 4, 16, 64, 128, 256, and unclear for 1, 7, 10, 6, 3, 2, and 6 scanners, respectively. The entire abdomen and pelvis was included in the CT-scan in 34 study populations, whereas the scan included only the lower abdomen and pelvis in 29 study populations. The field of view was not reported for eight study populations. Additional details about CT-protocols are presented in Table 1. We have described the use of contrast enhancement and low-dose protocols below under subgroup analyses.

Methodological quality of included studies

The outcome of our assessment of methodological quality is described below and is summarised in Figure 2 and Figure 3. None of the included studies were high-quality studies defined as studies with low risk of bias for all four domains. Three studies had low risk of bias for three domains (in't Hof 2004; Keyzer 2009; Pakaneh 2008). Fifteen studies had high or unclear risk of bias for all four domains. Insufficient reporting defined as one or more domains with unclear risk of bias was noted in 52 studies. Our assessments of the signalling questions for each study are presented under Characteristics of included studies

Figure 2. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.

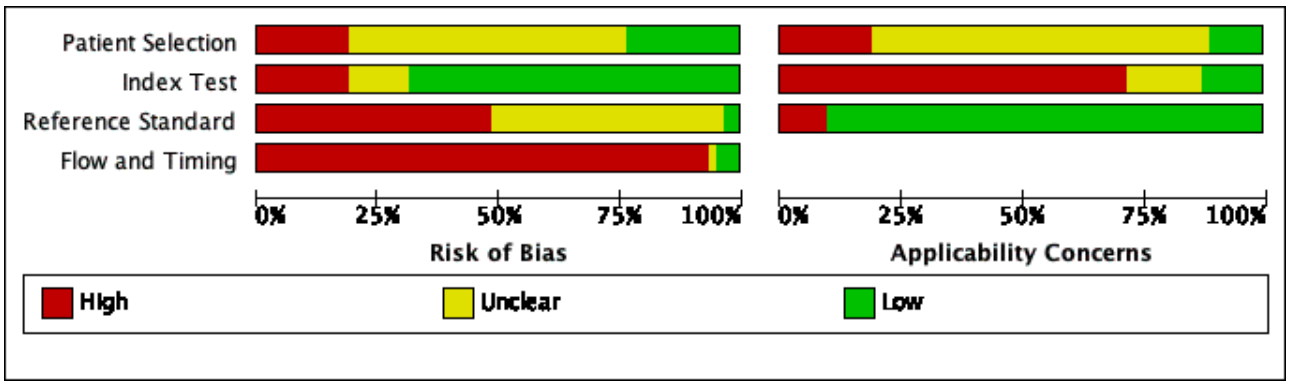


Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

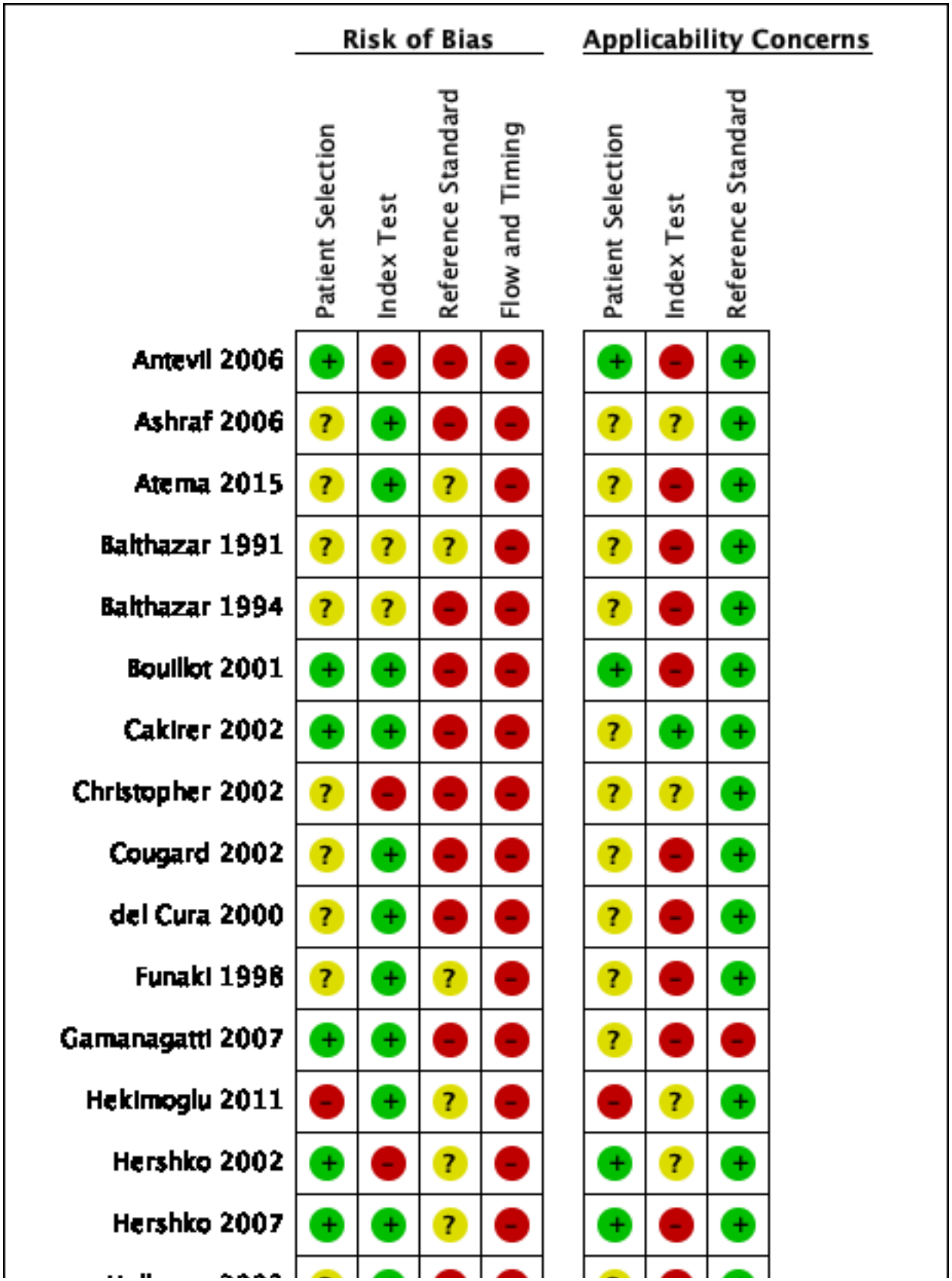


Figure 3. (Continued)

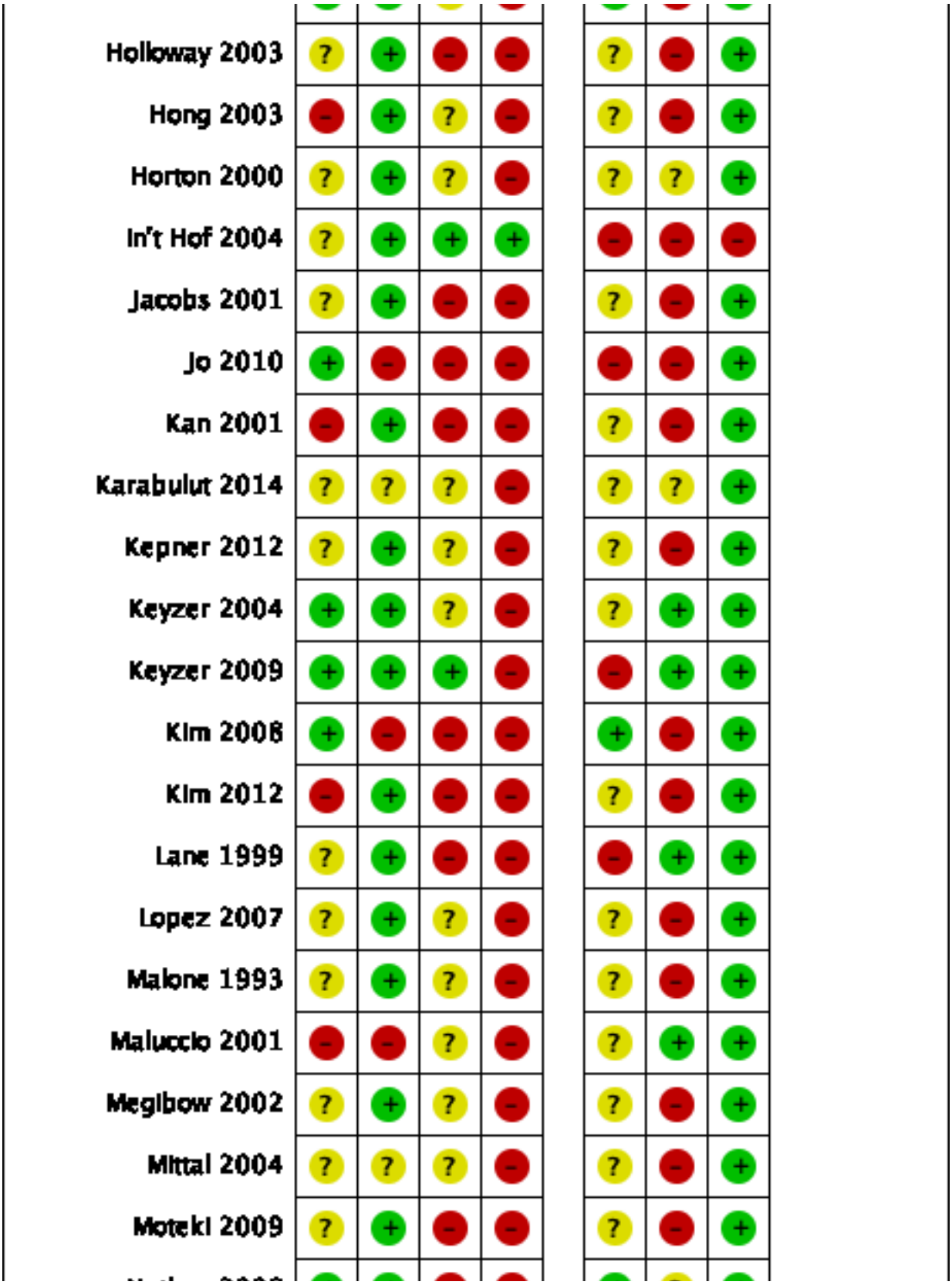


Figure 3. (Continued)

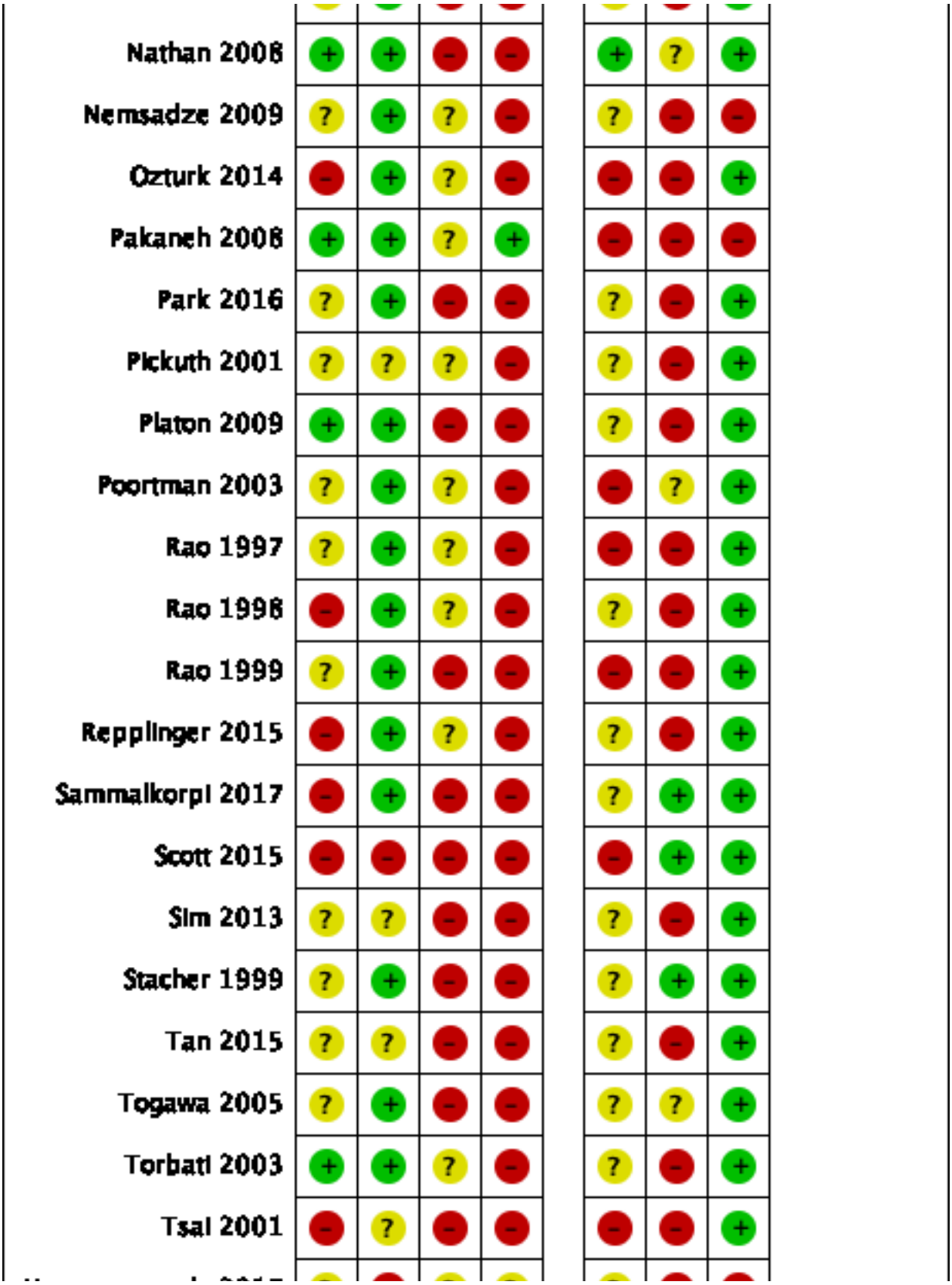
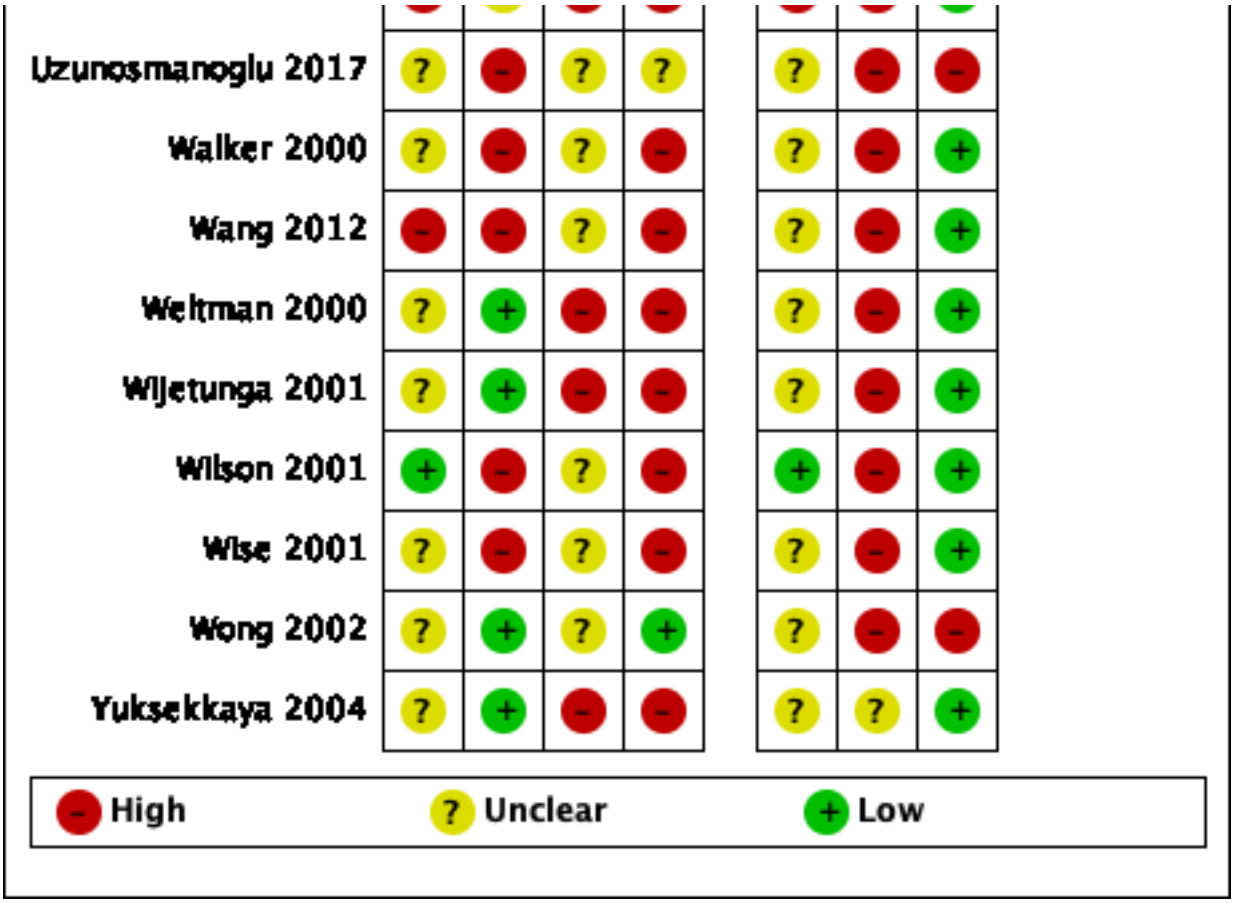


Figure 3. (Continued)



Domain 1: patient selection

A consecutive or a random sample of persons was enrolled in 24 studies, and inappropriate exclusions were avoided in 32 studies. Fifteen studies complied with both of these signalling questions and were considered to have low risk of bias for the patient selection domain. Both signalling questions were scored as unclear for 17 studies. As regards applicability, we considered the study population to represent an unselected sample of persons with suspected appendicitis in seven studies, whereas this was not so for 10 studies. In 47 studies, it was unclear if the study population was representative.

Domain 2: index test

In 58 studies, the CT-scan was evaluated without knowledge of the reference standard. This information was unclear in six studies. The criteria for the CT diagnosis of appendicitis were prespecified in 48 studies. This was not done in 13 studies and was unclear in three studies. We assessed the risk of bias introduced by execution and interpretation of CT-scans as low, high, and unclear in 44, 12, and 8 studies, respectively.

The description of the CT-scanner (manufacturer, model name, helical vs non-helical, number of detector rows) and the CT-protocol (use of contrast enhancement, low dose vs standard dose, slice thickness, slice interval, voltage and mAs product, use of multi-planar reconstruction) was adequate in 19 studies and was inadequate in 44 studies, whereas it was unclear for one study. In 11 studies, it was explicitly stated that coronal and/or sagittal reformations were used in the assessments.

The features included in the CT analyses were reported in 52 studies. The six most common features were appendix diameter (41 studies, diameter > 6 mm in 34 studies), periappendicular inflammation (40 studies), appendicolith (29 studies), abscess or phlegmon (19 studies), thickened or layered appendix wall (13 studies), and periappendiceal free fluid (13 studies).

The incorporation of equivocal CT assessments in the analyses was reported by 19 studies. Equivocal CT assessments were counted as positive for appendicitis in six studies, negative in eight, and excluded from analyses in two. Other incorporations were used in three studies. Results were based on initial assessment of the CT-scan in 31 studies; this was not so in 18 studies and was unclear in 15 studies. Overall, our concern regarding applicability of the execution and interpretation of CT-scans was high, low, and unclear for 46, 10, and 8 studies, respectively.

Domain 3: reference standard

A single reference standard was used in six studies in which all participants had surgery. Among these studies, histological examination of the resected appendix was performed in three (Pakaneh 2008; Uzunosmanoglu 2017; Wong 2002), intraoperative

findings were used in two (Gamanagatti 2007; in't Hof 2004), and it is unclear if the reference standard was based on intraoperative findings or on histological assessments of resected appendices in one (Nemsadze 2009). In another study, all participants had surgery, but macroscopically normal looking appendices were left in situ if participants had a laparoscopy, hence the reference standard was macroscopic findings during laparoscopy combined with histological examination of removed appendices (Poortman 2003). In the remaining 57 studies, only a subset of participants had surgery with or without appendectomy. Various follow-up regimens were used as a reference standard in those who did not have surgery. These regimens were highly heterogenous and ranged from checking hospital records for readmission to using systematic and standardised regimens including one or more telephone interviews, mailed questionnaires, or outpatient consultations within a predefined time frame. Telephone interviews, mailed questionnaires, outpatient visits, and review of medical records were conducted in 27, 3, 7, and 14 studies, respectively. Some studies used more than one of these methods for follow-up. The follow-up interval after CT or discharge was reported in 40 studies: it was up to one month, one to three months, and four or more months in 11, 15, and 8 studies, respectively. In six studies, the upper limit of the follow-up interval was not reported.

In our assessment, the reference standard was likely to correctly classify participants as having or not having acute appendicitis in 22 studies; this was not the case in 29 studies, and it was unclear in 13 studies. Inadequate or insufficiently described follow-up was the reason that 42 studies did not comply with our criteria for correct classification.

In 24 of these 42 studies, follow-up methods as well as follow-up intervals were inadequate or were not reported. In three studies, the follow-up interval was within 31 days, which was the longest duration we accepted, but the follow-up method was inadequate (checking for readmissions, reviewing hospital records, or method not stated). In the remaining 15 studies, the method of follow-up was adequate, but the follow-up interval was not; length of follow-up after CT was within three months in five studies, was longer than three months in six, and was not stated in four.

Histological evaluations, intraoperative findings, and results of follow-up were assessed without knowledge of the CT outcome in two studies (in't Hof 2004; Keyzer 2009). In 59 studies, this information was unclear, and in three studies, the reference standard included intraoperative assessment of the appendix by an unblinded surgeon (Gamanagatti 2007; Jacobs 2001; Platon 2009).

Overall, there was low risk of bias in the reference standard domain for two studies (in't Hof 2004; Keyzer 2009), high risk for 30 studies, and unclear risk for 32 studies. Our concern regarding applicability of the reference standard was low for the 58 studies with differential verification because the reference standard in these studies reflects clinical practice wherein only some persons with suspected appendicitis have surgery.

Domain 4: flow and timing

More than 95% of participants received a reference standard in 44 studies. This assessment was liberal, as it was often difficult to determine if participants scheduled for follow-up had received follow-up as intended. The choice of reference standard was considered independent of the CT result in eight studies; in seven of these, all participants had surgery. In five studies, it was unclear if the reference standard was independent of CT outcome. As stated above, all participants received the same reference test in six studies.

All participants with a CT diagnosis of appendicitis had surgery in 21 studies. In 18 studies, a few participants with a CT diagnosis of appendicitis were followed up. Likewise, all participants without CT signs of appendicitis were followed up in three studies, whereas a few participants without CT signs of appendicitis had surgery in 46 studies.

All participants were included in the analyses in 50 studies, in 13 studies they were not, and in one study this was unclear. Reasons why participants were not included in analyses included because they did not have surgery in three studies, because they were lost to follow-up in four studies, because CT findings were inconclusive in three studies, and for other reasons in three studies.

In our assessment, there was low risk of bias in the flow and timing domain for three studies (in't Hof 2004; Pakaneh 2008; Wong 2002), risk was high for 60 studies, and risk was unclear for one study.

Findings

Overall, the diagnostic accuracy of CT was reported for 71 separate study populations in the 64 included studies. Estimates of sensitivity ranged from 0.72 to 1.0, and estimates of specificity from 0.5 to 1.0. Sensitivity and specificity were higher than 0.90 in 40 study populations. The forest plot is presented in Figure 4, and the summary ROC plot in Figure 5. In the overall meta-analysis of results from the 71 study populations, summary sensitivity was 0.95 (95% confidence interval (CI) 0.93 to 0.96), and summary specificity was 0.94 (95% CI 0.92 to 0.95). The summary positive likelihood ratio was 15 (95% CI 12 to 19), and the summary negative likelihood ratio was 0.05 (95% CI 0.04 to 0.07). At the median appendicitis prevalence of 0.43, the probability of appendicitis following a positive and a negative CT result was 0.92 (95% CI 0.90 to 0.94) and 0.04 (95% CI 0.03 to 0.05), respectively. At the 25% percentile prevalence of 0.32, the probability following a positive and a negative CT result was 0.70 (95% CI 0.65 to 0.74) and 0.01 (95% CI 0.01 to 0.01), respectively. At the 75% percentile prevalence of 0.58, the probability following a positive and a negative CT result was 0.96 (95% CI 0.94 to 0.96) and 0.07 (95% CI 0.05 to 0.09), respectively.

Figure 4. Forest plot: CT regardless of contrast enhancement and radiation dose.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Hershko 2007	43	5	0	36	1.00 [0.92, 1.00]	0.88 [0.74, 0.96]
Kan 2001	4	2	0	25	1.00 [0.40, 1.00]	0.93 [0.76, 0.99]
Kepner 2012	34	4	0	75	1.00 [0.90, 1.00]	0.95 [0.88, 0.99]
Kepner 2012	41	1	0	72	1.00 [0.91, 1.00]	0.99 [0.93, 1.00]
Keyzer 2009	13	1	0	51	1.00 [0.75, 1.00]	0.98 [0.90, 1.00]
Mittal 2004	36	0	0	3	1.00 [0.90, 1.00]	1.00 [0.29, 1.00]
Platon 2009	37	2	0	47	1.00 [0.91, 1.00]	0.96 [0.86, 1.00]
Rao 1997	56	2	0	41	1.00 [0.94, 1.00]	0.95 [0.84, 0.99]
Rao 1999	32	2	0	66	1.00 [0.89, 1.00]	0.97 [0.90, 1.00]
Sim 2013	320	37	0	381	1.00 [0.99, 1.00]	0.91 [0.88, 0.94]
Tsai 2001	4	0	0	22	1.00 [0.40, 1.00]	1.00 [0.85, 1.00]
Wang 2012	26	1	0	32	1.00 [0.87, 1.00]	0.97 [0.84, 1.00]
Jo 2010	108	6	1	72	0.99 [0.95, 1.00]	0.92 [0.84, 0.97]
Tan 2015	153	22	2	173	0.99 [0.95, 1.00]	0.89 [0.83, 0.93]
Sammalkorpi 2017	253	7	4	225	0.98 [0.96, 1.00]	0.97 [0.94, 0.99]
Balthazar 1991	63	6	1	30	0.98 [0.92, 1.00]	0.83 [0.67, 0.94]
Repplinger 2015	63	15	1	125	0.98 [0.92, 1.00]	0.89 [0.83, 0.94]
Holloway 2003	188	6	3	226	0.98 [0.95, 1.00]	0.97 [0.94, 0.99]
Rao 1998	52	1	1	46	0.98 [0.90, 1.00]	0.98 [0.89, 1.00]
Weltman 2000	47	1	1	51	0.98 [0.89, 1.00]	0.98 [0.90, 1.00]
Kim 2008	88	3	2	64	0.98 [0.92, 1.00]	0.96 [0.87, 0.99]
Mittal 2004	43	4	1	4	0.98 [0.88, 1.00]	0.50 [0.16, 0.84]
Horton 2000	37	0	1	11	0.97 [0.86, 1.00]	1.00 [0.72, 1.00]
Hekimoglu 2011	31	1	1	67	0.97 [0.84, 1.00]	0.99 [0.92, 1.00]
Antevil 2006	92	5	3	429	0.97 [0.91, 0.99]	0.99 [0.97, 1.00]
Funaki 1998	29	4	1	66	0.97 [0.83, 1.00]	0.94 [0.86, 0.98]
Megibow 2002	29	2	1	21	0.97 [0.83, 1.00]	0.91 [0.72, 0.99]
Balthazar 1994	52	5	2	41	0.96 [0.87, 1.00]	0.89 [0.76, 0.96]
Wilson 2001	48	16	2	33	0.96 [0.86, 1.00]	0.67 [0.52, 0.80]
Gamanagatti 2007	46	1	2	3	0.96 [0.86, 0.99]	0.75 [0.19, 0.99]
Lane 1999	110	4	5	181	0.96 [0.90, 0.99]	0.98 [0.95, 0.99]
Stacher 1999	21	0	1	34	0.95 [0.77, 1.00]	1.00 [0.90, 1.00]
In't Hof 2004	83	0	4	16	0.95 [0.89, 0.99]	1.00 [0.79, 1.00]
Ozturk 2014	79	10	4	32	0.95 [0.88, 0.99]	0.76 [0.61, 0.88]
Kim 2012	171	16	9	244	0.95 [0.91, 0.98]	0.94 [0.90, 0.96]
Hershko 2007	37	3	2	36	0.95 [0.83, 0.99]	0.92 [0.79, 0.98]
Atema 2015	238	22	13	149	0.95 [0.91, 0.97]	0.87 [0.81, 0.92]
Cakirer 2002	89	3	5	33	0.95 [0.88, 0.98]	0.92 [0.78, 0.98]
Pickuth 2001	88	3	5	24	0.95 [0.88, 0.98]	0.89 [0.71, 0.98]
Wong 2002	35	1	2	12	0.95 [0.82, 0.99]	0.92 [0.64, 1.00]
Kim 2012	156	18	9	250	0.95 [0.90, 0.97]	0.93 [0.90, 0.96]
Nathan 2008	17	0	1	82	0.94 [0.73, 1.00]	1.00 [0.96, 1.00]
Scott 2015	32	5	2	47	0.94 [0.80, 0.99]	0.90 [0.79, 0.97]
Walker 2000	30	0	2	25	0.94 [0.79, 0.99]	1.00 [0.86, 1.00]
Wijetunga 2001	28	2	2	68	0.93 [0.78, 0.99]	0.97 [0.90, 1.00]
Park 2016	39	5	3	60	0.93 [0.81, 0.99]	0.92 [0.83, 0.97]
Nemsadze 2009	38	1	3	13	0.93 [0.80, 0.98]	0.93 [0.66, 1.00]
Yuksekkaya 2004	25	3	2	35	0.93 [0.76, 0.99]	0.92 [0.79, 0.98]
Karabulut 2014	36	10	3	55	0.92 [0.79, 0.98]	0.85 [0.74, 0.92]
Jacobs 2001	47	8	4	151	0.92 [0.81, 0.98]	0.95 [0.90, 0.98]
Torbati 2003	43	5	4	166	0.91 [0.80, 0.98]	0.97 [0.93, 0.99]
Ashraf 2006	21	0	2	38	0.91 [0.72, 0.99]	1.00 [0.91, 1.00]
Hong 2003	30	3	3	42	0.91 [0.76, 0.98]	0.93 [0.82, 0.99]
Hershko 2007	19	5	2	30	0.90 [0.70, 0.99]	0.86 [0.70, 0.95]
Lopez 2007	17	1	2	22	0.89 [0.67, 0.99]	0.96 [0.78, 1.00]
Hershko 2002	67	7	8	124	0.89 [0.80, 0.95]	0.95 [0.89, 0.98]
Boulliot 2001	63	4	9	24	0.88 [0.78, 0.94]	0.86 [0.67, 0.96]
Christopher 2002	27	4	4	66	0.87 [0.70, 0.96]	0.94 [0.86, 0.98]
Keyzer 2004	26	5	4	59	0.87 [0.69, 0.96]	0.92 [0.83, 0.97]
Malone 1993	65	4	10	132	0.87 [0.77, 0.93]	0.97 [0.93, 0.99]
Keyzer 2009	17	1	3	45	0.85 [0.62, 0.97]	0.98 [0.88, 1.00]
Uzunosmanoglu 2017	39	5	7	9	0.85 [0.71, 0.94]	0.64 [0.35, 0.87]

Figure 4. (Continued)

Keyzer 2009	17	1	3	45	0.85 [0.62, 0.97]	0.98 [0.88, 1.00]
Uzunosmanoglu 2017	39	5	7	9	0.85 [0.71, 0.94]	0.64 [0.35, 0.87]
Cougard 2002	37	3	7	42	0.84 [0.70, 0.93]	0.93 [0.82, 0.99]
Wise 2001	20	11	4	65	0.83 [0.63, 0.95]	0.86 [0.76, 0.93]
del Cura 2000	54	4	12	82	0.82 [0.70, 0.90]	0.95 [0.89, 0.99]
Togawa 2005	47	4	11	38	0.81 [0.69, 0.90]	0.90 [0.77, 0.97]
Maluccio 2001	28	6	7	63	0.80 [0.63, 0.92]	0.91 [0.82, 0.97]
Hekimoglu 2011	20	5	6	69	0.77 [0.56, 0.91]	0.93 [0.85, 0.98]
Pakaneh 2008	32	1	10	7	0.76 [0.61, 0.88]	0.88 [0.47, 1.00]
Poortman 2003	100	11	32	56	0.76 [0.68, 0.83]	0.84 [0.73, 0.92]
Moteki 2009	62	30	24	138	0.72 [0.61, 0.81]	0.82 [0.75, 0.88]

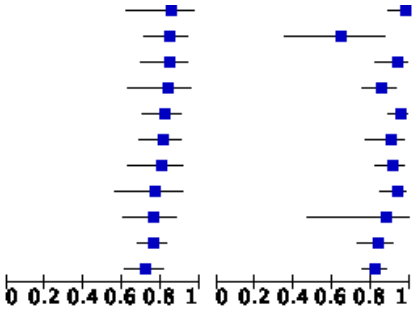
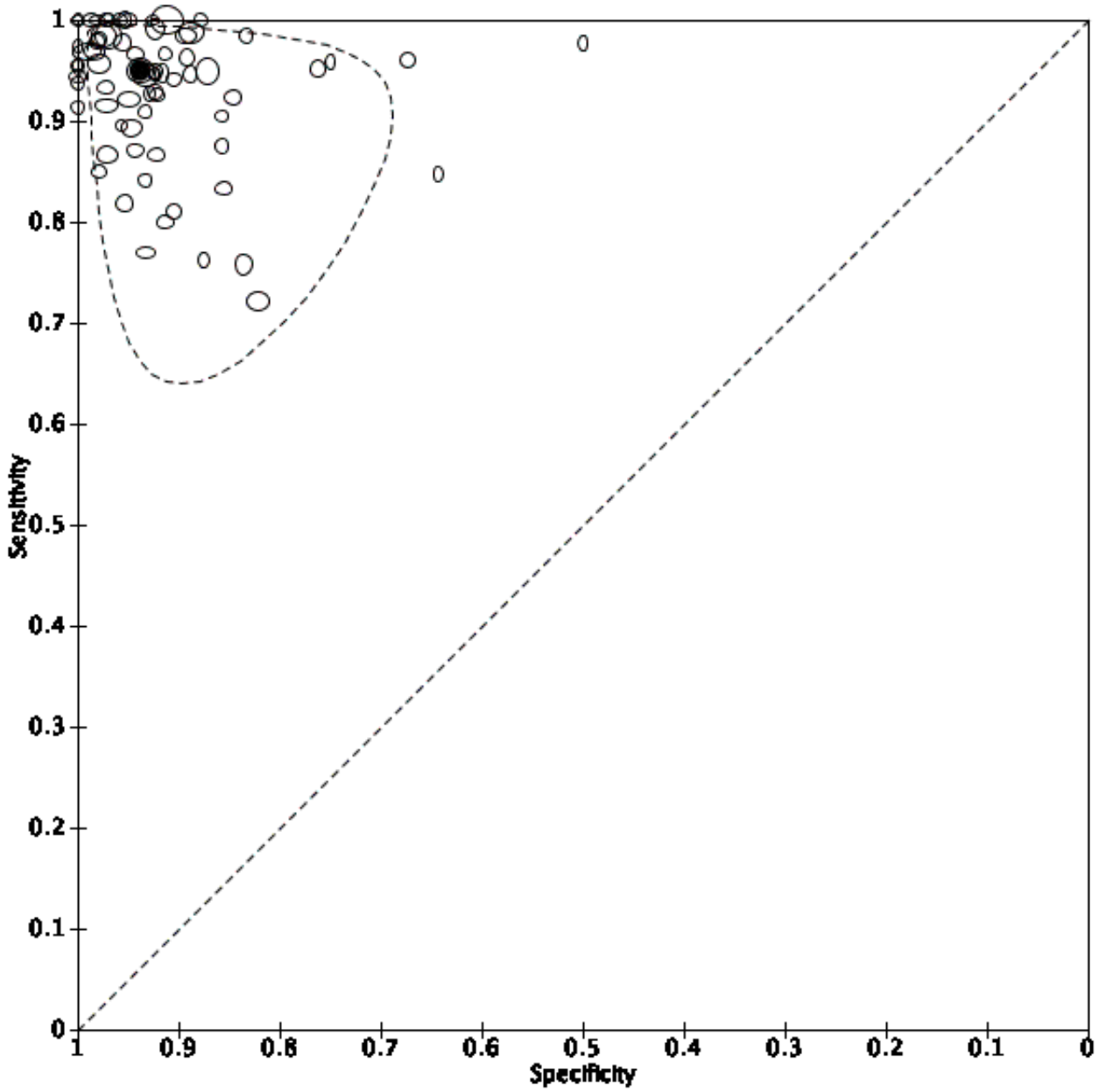


Figure 5. Summary ROC plot of CT for diagnosis of acute appendicitis (any contrast enhancement and radiation dose). The hollow symbols represent the pairs of sensitivity and specificity from the included studies; the symbols are scaled according to sample sizes of the studies. The solid circle represents the summary sensitivity and specificity. This summary point is surrounded by a 95% prediction region (interrupted line).



Comparative subgroup analyses according to contrast enhancement and radiation dose

Unenhanced CT was evaluated in 19 study populations, and CT with IV, rectal, oral, and IV+oral contrast enhancement was evaluated in 18, 9, 7, and 15 study populations, respectively. Summary sensitivity varied between 0.89 (95% CI 0.81 to 0.94) and 0.97 (95% CI 0.93 to 0.99) across subgroups defined by the use of contrast enhancement, and summary specificity varied from 0.93 (95% CI 0.90 to 0.95) to 0.95 (95% CI 0.90 to 0.98). Summary sensitivity was lowest for CT with oral contrast 0.89 (95% CI 0.81 to 0.94) and unenhanced CT 0.91 (95% CI 0.87 to 0.93), whereas the variation was marginal between CT with IV contrast, rectal contrast, and IV+oral contrast. These results correspond with the finding of lower sensitivity but similar specificity in three studies comparing CT with oral contrast enhancement to CT with IV+oral contrast enhancement using a paired or a randomised design (Jacobs 2001; Kepner 2012; Keyzer 2009) (Table 2). Likewise, sensitivity of unenhanced CT was lower than sensitivity of CT with any type of contrast enhancement in two studies with a paired or a randomised design (Hershko 2007; Keyzer 2009).

Low-dose protocols were evaluated in eight study populations. Summary sensitivity and specificity for low-dose CT was 0.94 (95% CI 0.90 to 0.97) and 0.94 (95% CI 0.91 to 0.96), respectively. These estimates were similar to summary estimates in the overall meta-analysis. This finding corresponds closely with the findings in four studies with direct comparisons of low-dose and standard-dose CT (Keyzer 2004; Keyzer 2009; Kim 2012; Platon 2009) (Table 3).

Results of the subgroup analyses are summarised in Table 4 presented graphically in Figure 6, Figure 7, Figure 8, Figure 9, and Figure 10, and described below. In addition to the types of contrast enhancement covered by the subgroup analyses, CT with oral+rectal contrast was evaluated in three study populations (Funaki 1998; Kan 2001; Rao 1997), and CT with IV+oral+rectal contrast was evaluated in one study (Mittal 2004). Several types of contrast enhancement were used in three study populations, and results from these populations were excluded from the subgroup analyses (Nemsadze 2009; Pickuth 2001; Weltman 2000). In the protocol, some of the subgroup analyses were planned as sensitivity analyses (see Differences between protocol and review).

Figure 6. Summary ROC plot of CT with intravenous contrast enhancement versus unenhanced CT. See the caption for Figure 5 for a description of symbols and lines.

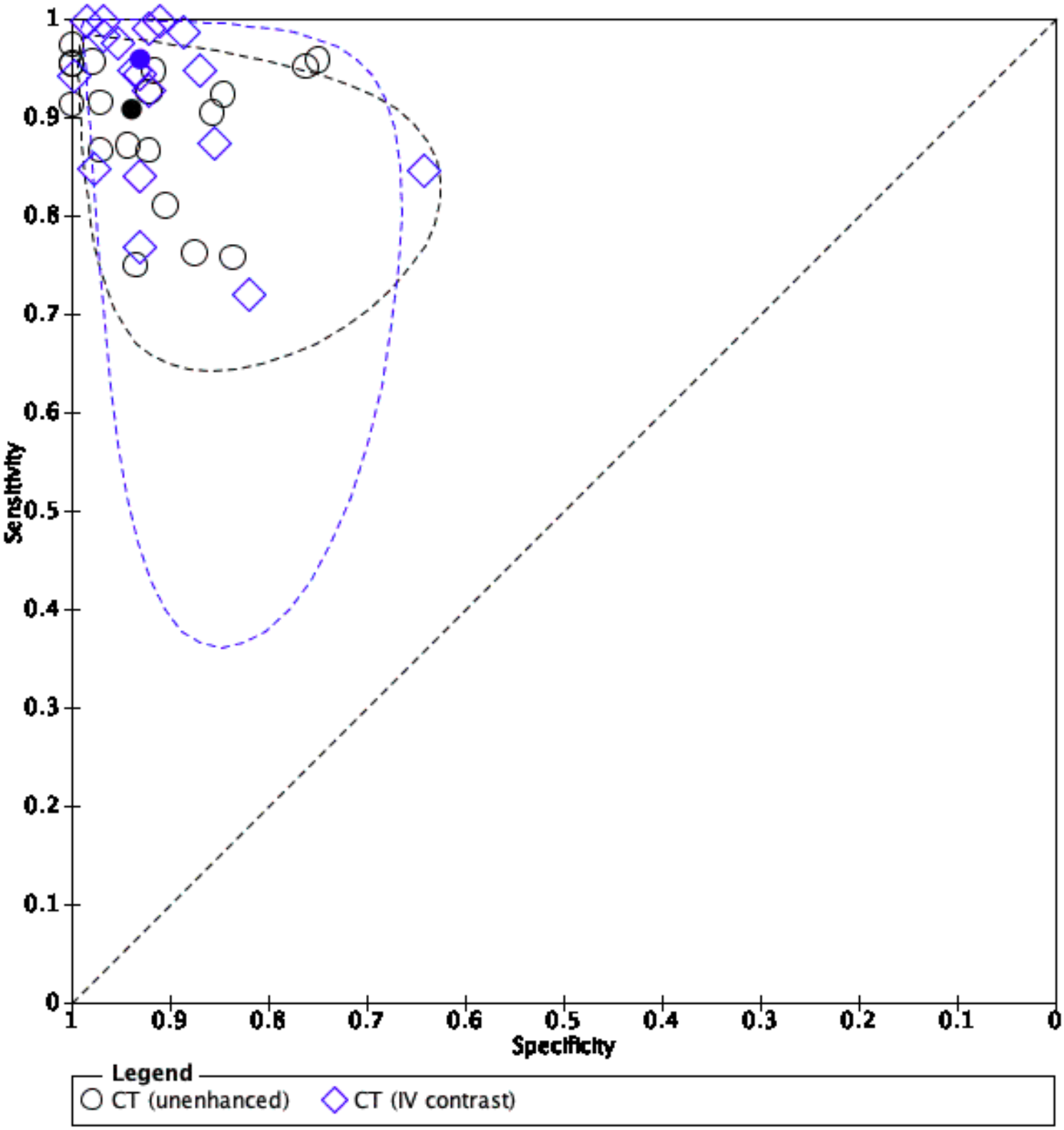


Figure 7. Summary ROC plot of CT with rectal contrast enhancement versus unenhanced CT. See the caption for Figure 5 for a description of symbols and lines.

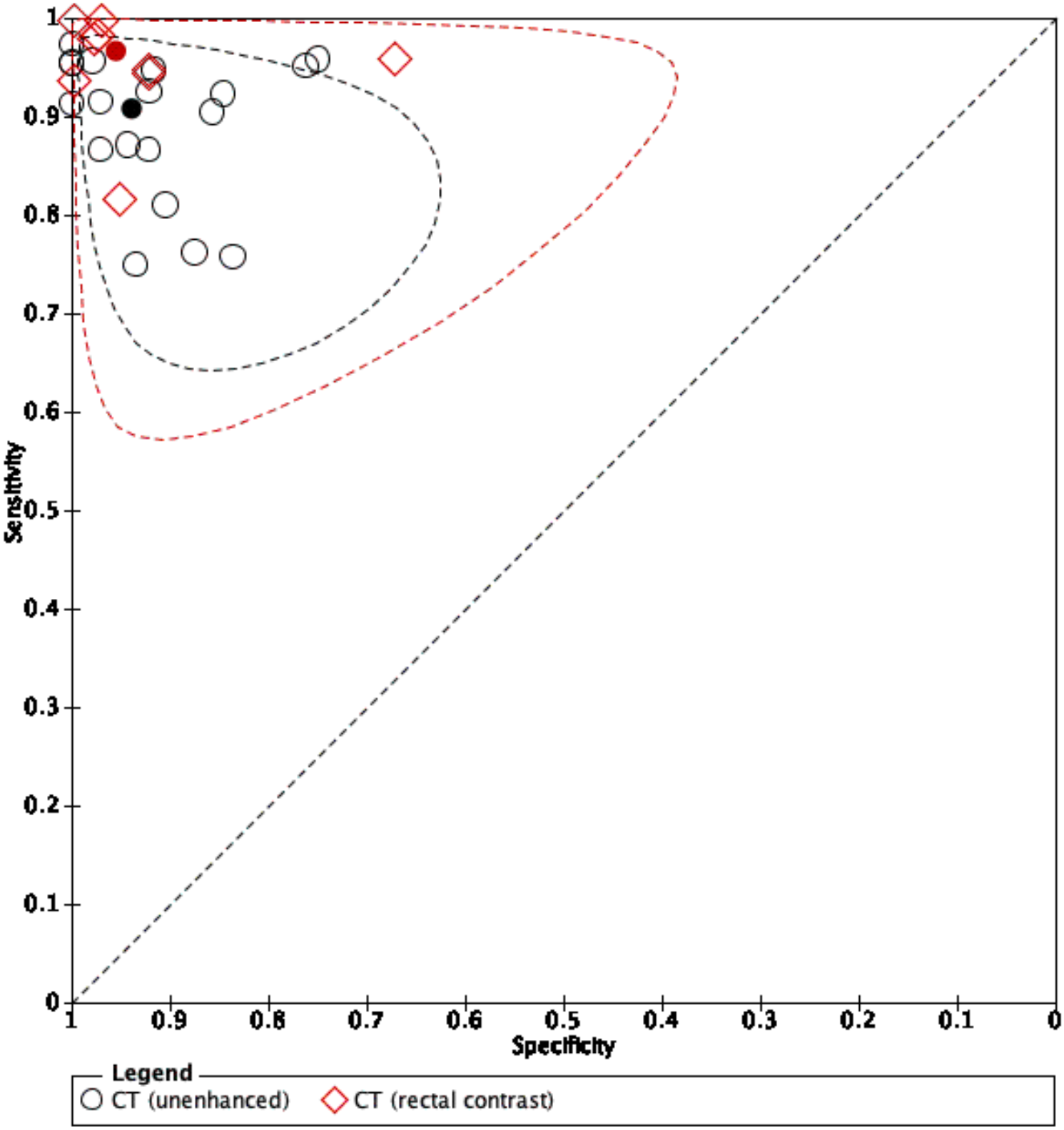


Figure 8. Summary ROC plot of CT with oral contrast enhancement versus unenhanced CT. See the caption for Figure 5 for a description of symbols and lines.

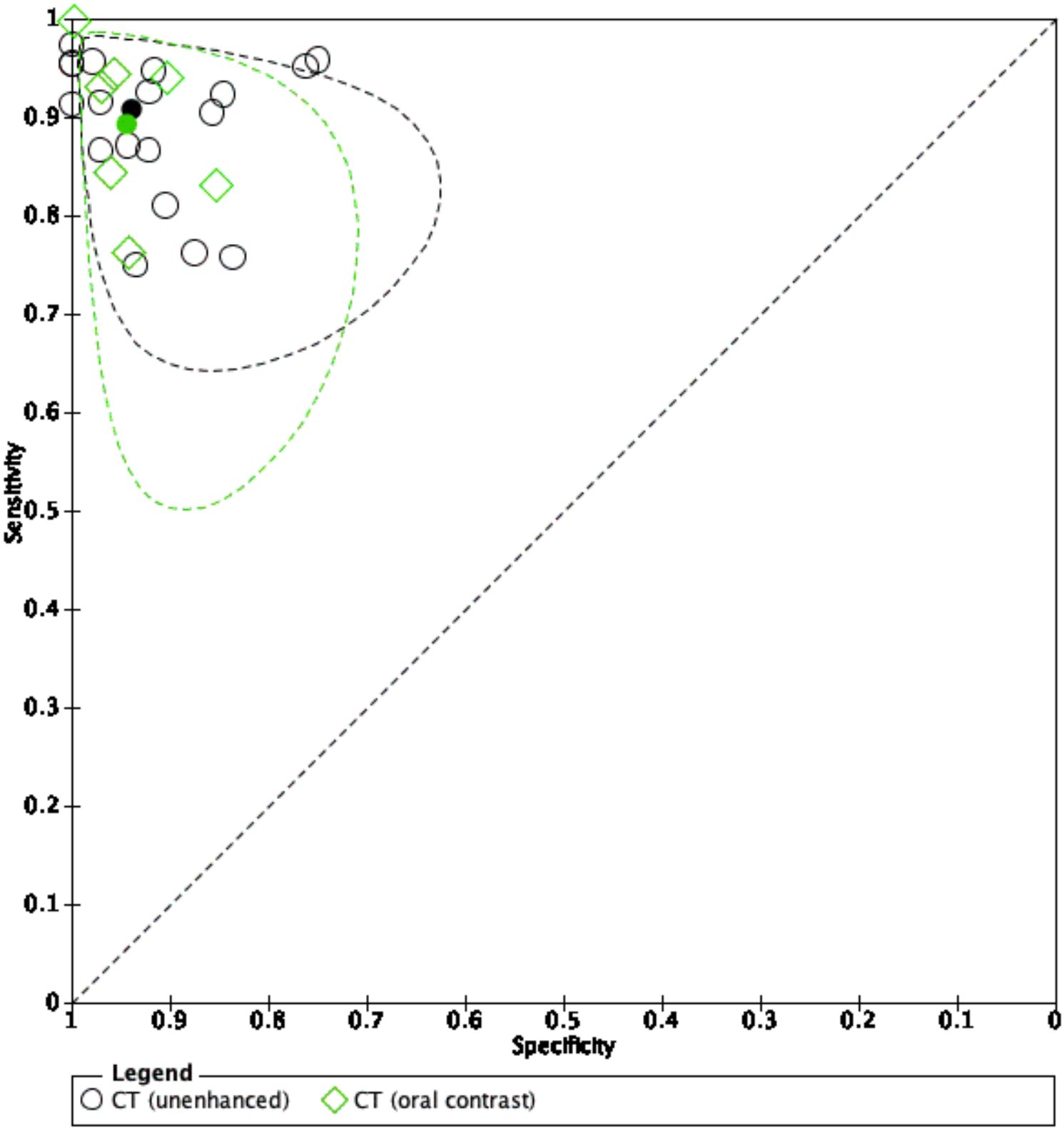


Figure 9. Summary ROC plot of CT with intravenous and oral contrast enhancement versus unenhanced CT. See the caption for Figure 5 for a description of symbols and lines.

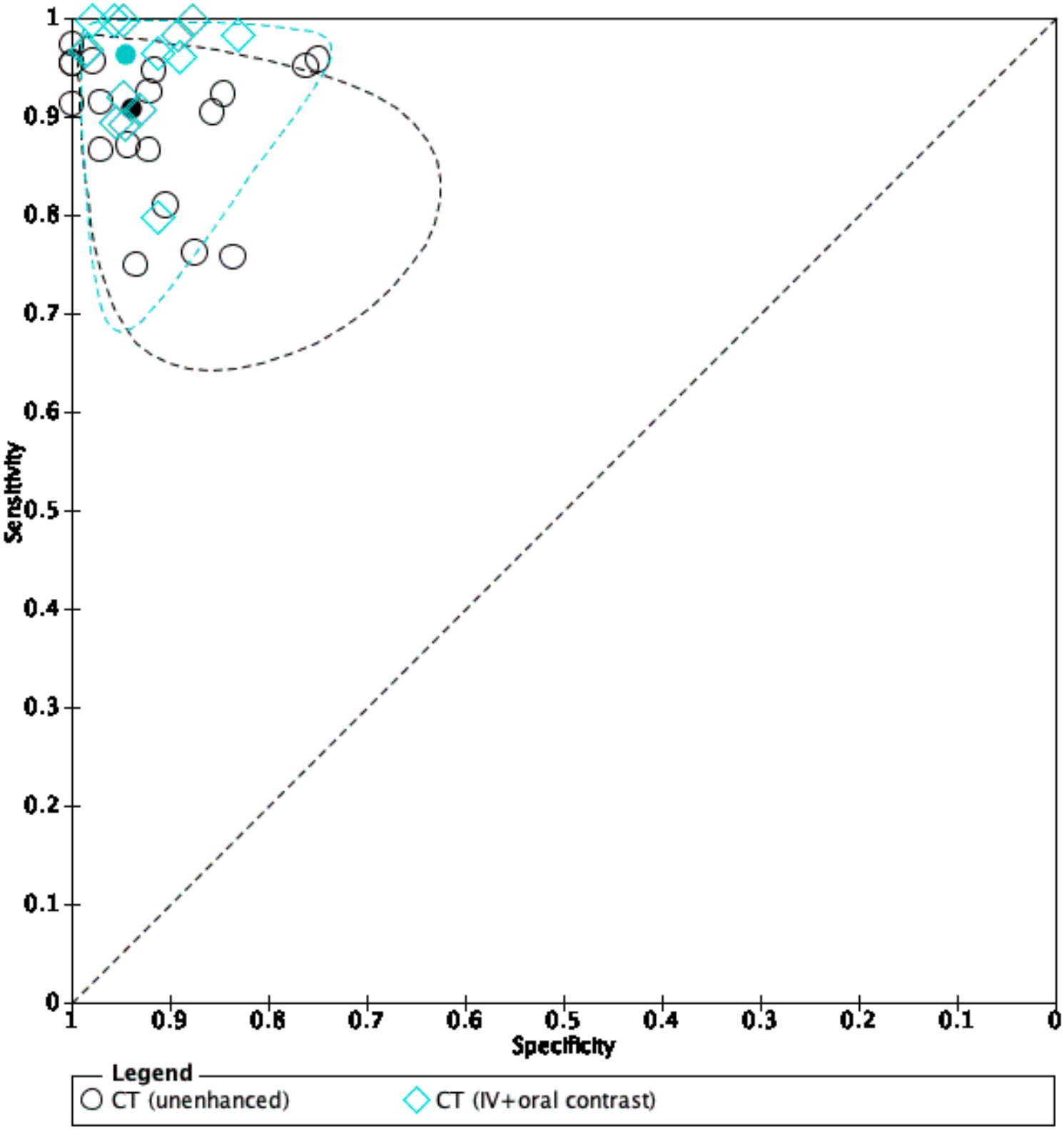
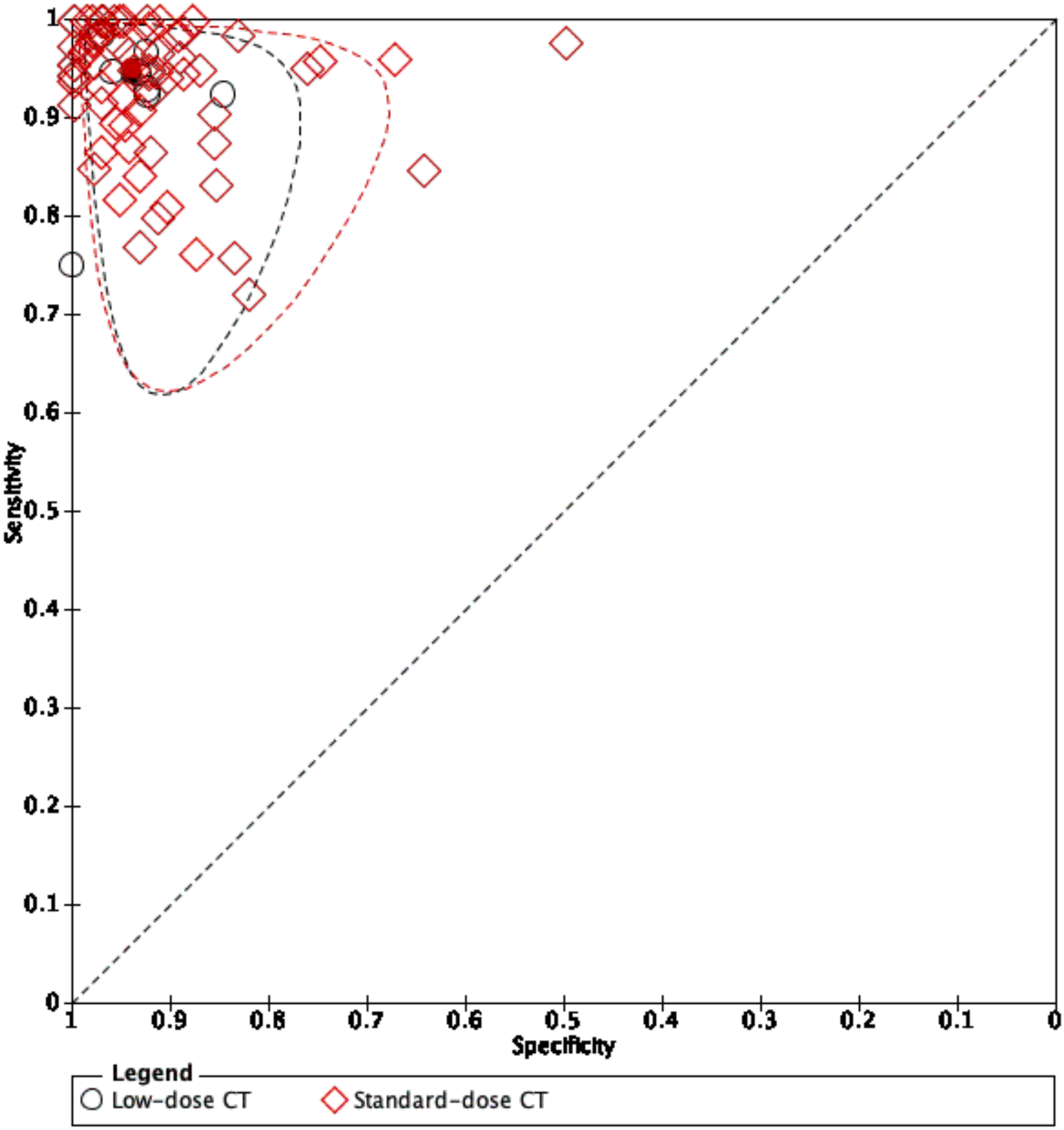


Figure 10. Summary ROC plot of low-dose versus standard-dose CT. See the caption for Figure 5 for a description of symbols and lines.



Unenhanced CT

Estimates of sensitivity and specificity for unenhanced CT were available for 19 study populations reported in 19 studies. Two studies reported results for unenhanced standard-dose CT and unenhanced low-dose CT in the same participants (Keyzer 2004; Keyzer 2009). Results for standard-dose CT were selected for this analysis. The median prevalence of appendicitis in these populations was 0.39, with interquartile range 0.36 to 0.72, and range 0.22 to 0.92. Estimates of sensitivity ranged from 0.75 to 0.97, and estimates of specificity ranged from 0.75 to 1.0. The summary sensitivity was 0.91 (95% CI 0.87 to 0.93), and the summary specificity was 0.94 (95% CI 0.90 to 0.96).

CT with intravenous contrast enhancement

Estimates of sensitivity and specificity for CT with IV contrast enhancement were available for 18 study populations reported in 17 studies. One study provided results for standard-dose CT and low-dose CT in the same study population (Keyzer 2009). Results for standard-dose CT were selected for this analysis. The median prevalence of appendicitis in these populations was 0.44, with interquartile range 0.36 to 0.57, and range 0.18 to 0.77. Estimates of sensitivity ranged from 0.72 to 1.0, and estimates of specificity from 0.64 to 1.0. The summary sensitivity was 0.96 (95% CI 0.92 to 0.98), and the summary specificity was 0.93 (95% CI 0.90 to 0.95).

Meta-regression analyses showed a trend for higher summary sensitivity for CT with IV contrast enhancement compared to unenhanced CT (0.96, 95% CI 0.92 to 0.98 vs 0.90, 95% CI 0.87 to 0.93) (likelihood ratio test, Chi^2 = 3.35, 1 df, P = 0.07). There was no statistically significant difference for summary specificity (0.93, 95% CI 0.90 to 0.95 vs 0.94, 95% CI 0.90 to 0.96) (likelihood ratio test, Chi^2 = 0.20, 1 df, P = 0.66) (Figure 6).

CT with rectal contrast enhancement

Estimates of sensitivity and specificity for CT with rectal contrast enhancement were available for nine independent study populations reported in nine studies. The median prevalence of appendicitis in these populations was 0.51, with interquartile range 0.45 to 0.56, and range 0.32 to 0.92. Estimates of sensitivity ranged from 0.82 to 1.0, and estimates of specificity from 0.67 to 1.0. The summary sensitivity was 0.97 (95% CI 0.93 to 0.99), and the summary specificity was 0.95 (95% CI 0.90 to 0.98).

In meta-regression analyses, summary sensitivity for CT with rectal contrast enhancement was statistically significantly higher than summary sensitivity for unenhanced CT (0.97, 95% CI 0.93 to 0.99 vs 0.90, 95% CI 0.87 to 0.93) (likelihood ratio test, Chi^2 = 5.78, 1 df, P = 0.02). There was no statistically significant difference for summary specificity (0.95, 95% CI 0.90 to 0.98 vs 0.94, 95% CI 0.90 to 0.96) (likelihood ratio test, Chi^2 = 0.27, 1 df, P = 0.61) (Figure 7).

CT with oral contrast enhancement

Estimates of sensitivity and specificity for CT with oral contrast enhancement were available for seven independent study populations reported in seven studies. One study provided results for standard-dose CT and low-dose CT in the same study population (Keyzer 2009), and we used the results for standard-dose CT for this analysis. The median prevalence of appendicitis in these populations was 0.24, with interquartile range 0.20 to 0.40, and range 0.15 to 0.43. Estimates of sensitivity ranged from 0.76 to 1.0, and estimates of specificity from 0.86 to 1.0. The summary sensitivity was 0.89 (95% CI 0.81 to 0.94), and the summary specificity was 0.94 (95% CI 0.90 to 0.97).

Meta-regression analyses showed no statistically significant difference between summary sensitivity or specificity for CT with oral contrast enhancement versus unenhanced CT (likelihood ratio test, Chi^2 = 0.46, 2 df, P = 0.80) (Figure 8).

CT with intravenous and oral contrast enhancement

Estimates of sensitivity and specificity for CT with IV and oral contrast enhancement were available for 15 independent study populations reported in 15 studies. Again, one study provided results for standard-dose CT and low-dose CT in the same study population (Keyzer 2009), and we used the results for standard-dose CT for this analysis. The median prevalence of appendicitis in these populations was 0.36, with interquartile range 0.30 to 0.51, and range 0.18 to 0.64. Estimates of sensitivity ranged from 0.80 to 1.0, and estimates of specificity from 0.83 to 0.99. The summary sensitivity was 0.96 (95% CI 0.93 to 0.98), and the summary specificity was 0.94 (95% CI 0.92 to 0.96).

In meta-regression analyses, summary sensitivity for CT with intravenous and oral contrast enhancement was statistically significantly higher than summary sensitivity for unenhanced CT (0.96, 95% CI 0.93 to 0.98 vs 0.90, 95% CI 0.87 to 0.93) (likelihood ratio test, Chi^2 = 6.85, 1 df, P = 0.01). There was no statistically significant difference for summary specificity (0.94, 95% CI 0.92 to 0.96 vs 0.94, 95% CI 0.90 to 0.96) (likelihood ratio test, Chi^2 = 0.23, 1 df, P = 0.63) (Figure 9).

Low-dose CT regardless of contrast enhancement

Estimates of sensitivity and specificity for low-dose CT were available for eight independent study populations reported in seven studies. The study that contributed two study populations was a randomised study that reported results for low-dose CT with no contrast and IV contrast enhancement in one group, and for oral contrast and oral+IV contrast enhancement in the other group. For this analysis, we selected intravenous contrast enhancement from the first group and oral contrast enhancement from the other. In the remaining six study populations, IV, oral, and no contrast enhancement were used in three, one, and two studies, respectively. The median prevalence of appendicitis in the eight populations was 0.38, with interquartile range 0.30 to 0.41, and range 0.20 to 0.53. Estimates of sensitivity ranged from 0.75 to 0.98, and estimates of specificity from 0.85 to 1.0. The summary sensitivity was 0.94 (95% CI 0.90 to 0.97), and the summary specificity was 0.94 (95% CI 0.91 to 0.96).

Meta-regression analyses showed no statistically significant difference between summary sensitivity or specificity for low-dose versus standard- or unspecified-dose CT (likelihood ratio test, Chi^2 = 0.21, 2 df, P = 0.90) (Figure 10).

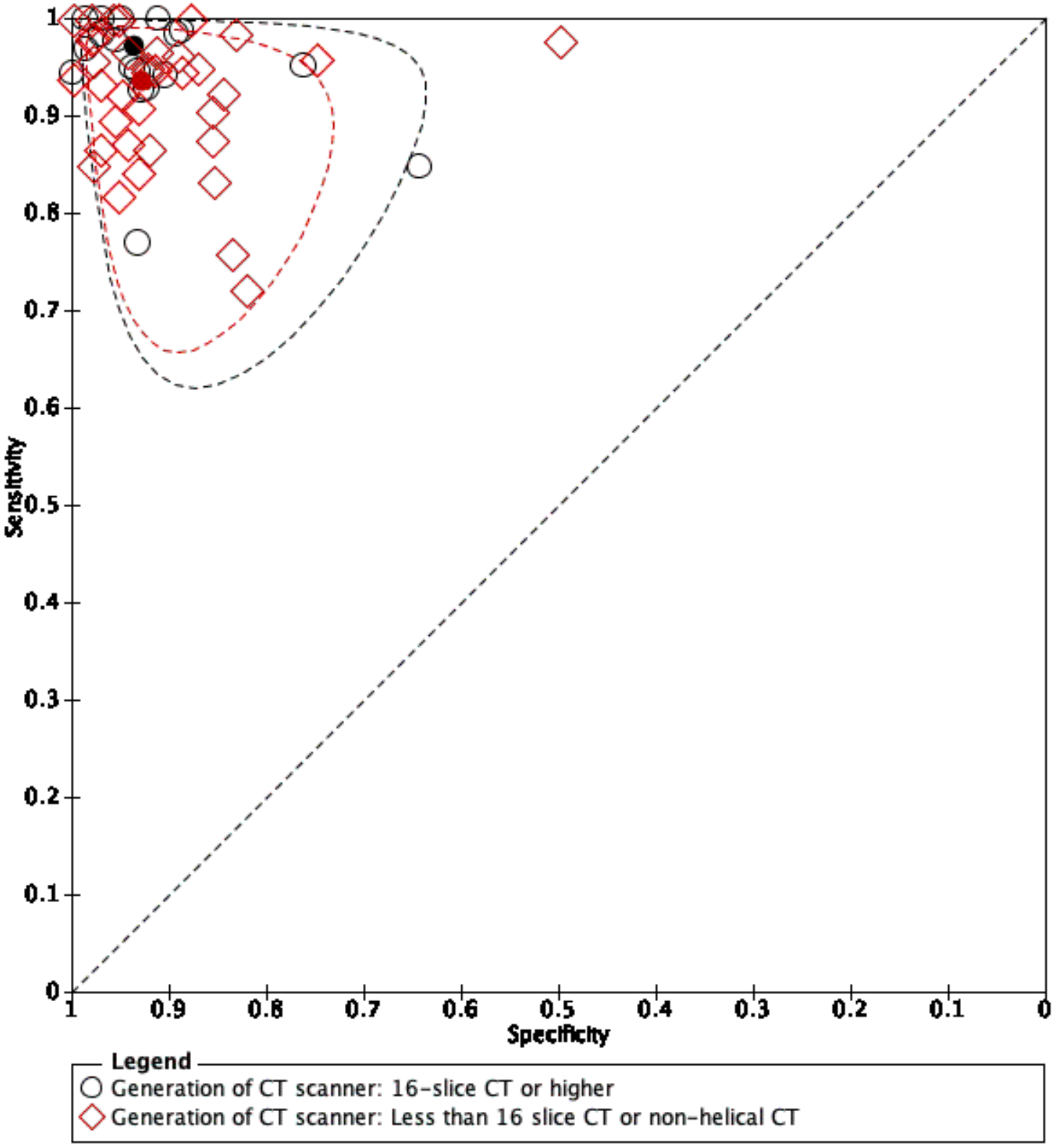
Post-test probabilities, summary likelihood ratios, and absolute differences in summary sensitivity and specificity for the subgroup analyses described above are presented in Summary of findings 1 and Table 4.

Investigation of heterogeneity

Influence of CT-scanner generation

A non-helical CT-scanner or a helical CT-scanner with less than 16-detector row technology was used in 32 studies (36 study populations), and summary sensitivity and specificity were 0.94 (95% CI 0.91 to 0.95) and 0.93 (95% CI 0.91 to 0.94), respectively. A helical CT-scanner with 16-detector row or higher technology was used in 15 studies (18 study populations), and summary sensitivity and specificity were 0.97 (95% CI 0.95 to 0.98) and 0.94 (95% CI 0.91 to 0.96), respectively. In meta-regression analyses, summary sensitivity was statistically significantly higher for the latter group than for the former (likelihood ratio test, 1 df, Chi^2 = 5.23, P = 0.02). There was no statistically significant difference for summary specificity between groups (likelihood ratio test, 1 df, Chi^2 = 0.24, P = 0.63) (Figure 11). The number of detector rows was not stated in 17 studies (17 study populations).

Figure 11. Exploration of heterogeneity: influence of CT-scanner generation (CT with 16 detector rows or higher vs CT with fewer than 16 detector rows). See the caption for Figure 5 for a description of symbols and lines.



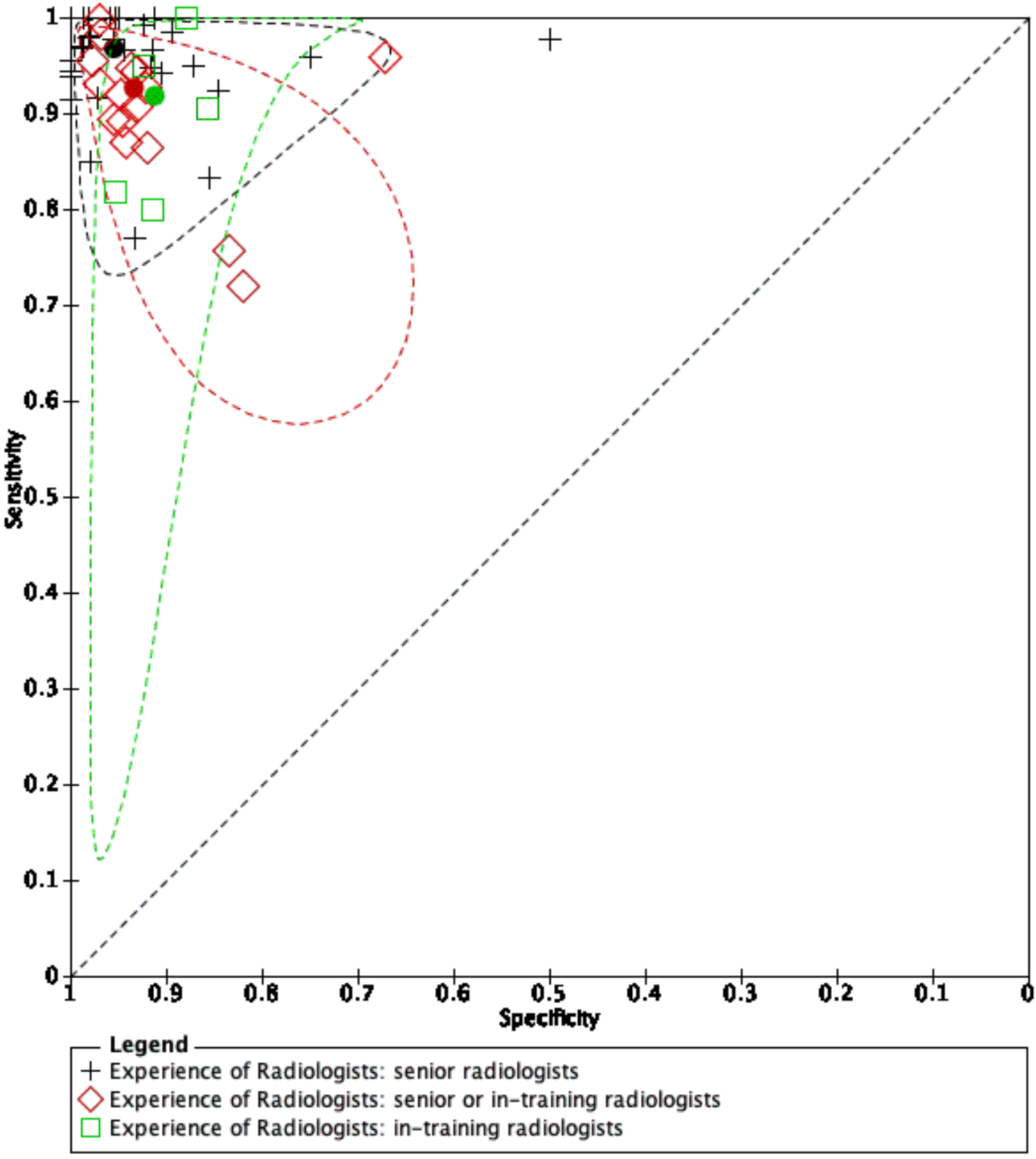
Influence of radiologists' experience

Senior radiologists evaluated CT-scans in 27 studies (31 study populations), in-training radiologists evaluated CT-scans in three studies (five study populations), and CT-scans were evaluated by senior or in-training radiologists in 15 studies (16 study populations). The radiologists' experience was not reported in 19 studies (19 study populations). Summary estimates of sensitivity and specificity were as follows for the three groups.

- Senior radiologists: 0.97 (95% CI 0.95 to 0.98) and 0.95 (95% CI 0.93 to 0.97), respectively.
- In-training radiologists: 0.92 (95% CI 0.80 to 0.97) and 0.91 (95% CI 0.86 to 0.94), respectively.
- Senior or in-training radiologists: 0.93 (95% CI 0.89 to 0.95) and 0.93 (95% CI 0.90 to 0.96), respectively.

In meta-regression analyses, we pooled in-training radiologists with senior or in-training radiologists. In these analyses, summary sensitivity was statistically significantly higher in study populations with senior radiologists' evaluations (likelihood ratio test, Chi^2 = 8.01, 1 df, P = 0.01). Summary specificity was also higher in study populations with senior radiologists' evaluations but was not significantly higher (likelihood ratio test, Chi^2 = 2.21, 1 df, P = 0.14) (Figure 12).

Figure 12. Exploration of heterogeneity: Influence of radiologists' experience. See the caption for Figure 5 for a description of symbols and lines.



Influence of pretest degree of suspicion of appendicitis

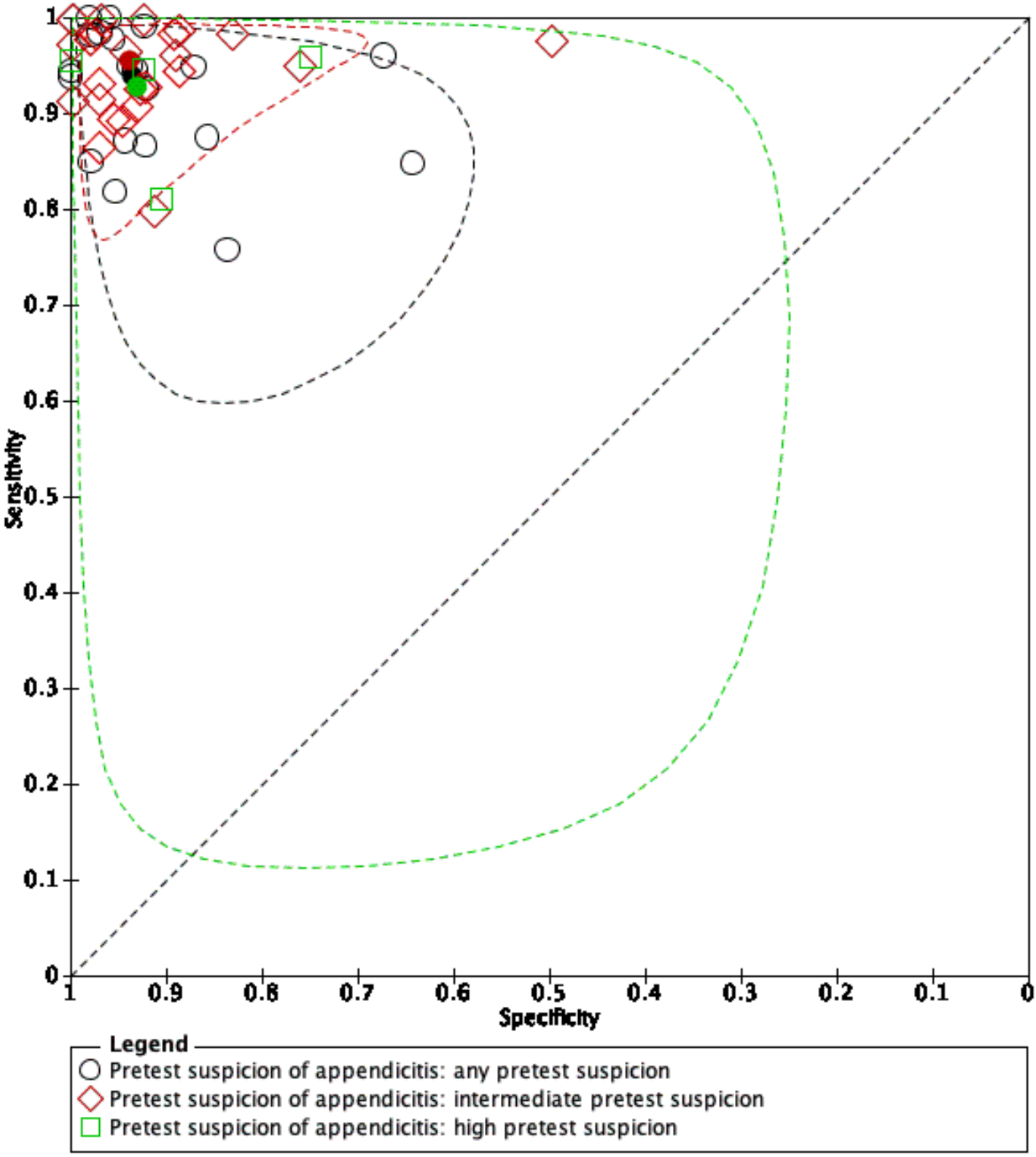
Participants with intermediate suspicion of appendicitis were recruited in 24 studies (25 study populations), participants with any suspicion were recruited in 18 studies (20 study populations), and participants with a high degree of suspicion were included in four studies (four study populations). The degree of suspicion was unclear in 18 studies (22 study populations). Summary estimates of sensitivity and specificity for the first two mentioned groups were as follows.

- Intermediate suspicion: 0.96 (95% CI 0.93 to 0.97) and 0.94 (95% CI 0.91 to 0.96), respectively.
- Any suspicion: 0.94 (95% CI 0.91 to 0.96) and 0.94 (95% CI 0.90 to 0.96), respectively.

There was no difference in the prevalence of appendicitis between studies recruiting participants with intermediate and any suspicion of appendicitis. Median and interquartile ranges were 0.47 (0.35 to 0.58) and 0.44 (0.34 to 0.64), respectively.

In meta-regression analyses, we found no statistical evidence of a difference in summary sensitivity or specificity between study populations including participants with intermediate and any suspicion of appendicitis (likelihood ratio test, Chi^2 = 1.78, 2 df, P = 0.41). This did not change when we included data from all study populations in the analysis and grouped studies with any, high, and unclear degree of suspicion (likelihood ratio test, Chi^2 = 1.08, 2 df, P = 0.58) (Figure 13).

Figure 13. Exploration of heterogeneity: influence of pre-test suspicion of appendicitis. See the caption for Figure 5 for a description of symbols and lines.



Sensitivity analyses

The analyses in this section differ from those planned in the protocol (see Differences between protocol and review).

Influence of methodological quality

Domains 1 and 2 (patient selection and index test)

Summary sensitivity and specificity for 18 study populations with low risk of bias for domain 1 were 0.94 (95% CI 0.91 to 0.96) and 0.94 (95% CI 0.91 to 0.96), respectively. Likewise, summary sensitivity and specificity for 50 study populations with low risk of bias for domain 2 were 0.94 (95% CI 0.92 to 0.96) and 0.95 (95% CI 0.93 to 0.96), respectively. These estimates were hardly different compared to the overall summary estimates of sensitivity (0.95) and specificity (0.94).

Domains 3 and 4 (reference standard and flow and timing)

Risk for bias was scored as low in two studies (three study populations) for domain 3 and in three studies (three study populations) for domain 4. This was insufficient for meta-analysis.

Other sensitivity analyses

In the overall meta-analysis, it was necessary to select one of two or more analyses from four paired studies including five study populations (Jacobs 2001; Keyzer 2004; Keyzer 2009; Platon 2009). These studies compared the accuracy of different doses or enhancement protocols in the same participants. We performed a sensitivity analysis to assess the influence of selecting other analyses from these studies and found that summary sensitivity and specificity did not change (Table 5). From the study that presented more than two analyses, we selected results from the standard-dose protocols. Likewise, two studies including three study populations compared the accuracy of two or more CT-protocols at low and standard doses (Keyzer 2004; Keyzer 2009). Results for the standard-dose protocols were selected in the subgroup meta-analyses. In sensitivity analyses, we used the low-dose protocol results from these studies instead and found no effects on summary estimates of sensitivity and specificity (Table 5).

We also explored the potential effects of including studies with a mix of paediatric and adult participants. Participants younger than 15 years of age were included in 26 studies with 28 study populations, and it was unclear if two other studies with two study populations included paediatric participants; summary sensitivity and specificity for these 30 study populations were 0.95 (95% CI 0.93 to 0.97) and 0.94 (95% CI 0.91 to 0.95), respectively. In contrast, all participants were adults in 36 studies with 41 study populations; summary sensitivity and specificity for this subgroup were 0.95 (95% CI 0.92 to 0.96) and 0.94 (95% CI 0.92 to 0.95), respectively. Hence, the inclusion of studies with a mix of adult and paediatric participants appears to have no effect on the summary estimates.

Finally, we explored whether inclusion of five studies that used laparoscopic findings as the reference standard influenced summary estimates of sensitivity and specificity (Gamanagatti 2007; in't Hof 2004; Jacobs 2001; Platon 2009; Poortman 2003). These estimates did not change when we repeated the overall meta-analysis and excluded results from the five studies.

DISCUSSION

Summary of main results

The main results of this review are presented in Summary of findings 1. We included 64 studies with results from 71 separate study populations. Summary sensitivity and specificity of computed tomography (CT) regardless of protocol were 0.95 (95% confidence interval (CI) 0.93 to 0.96) and 0.94 (95% CI 0.92 to 0.95), respectively. In subgroup analyses according to contrast enhancement, summary sensitivity was higher for CT with intravenous contrast (0.96, 95% CI 0.92 to 0.98), CT with rectal contrast (0.97, 95% CI 0.93 to 0.99), and CT with intravenous+oral contrast enhancement (0.96, 95% CI 0.93 to 0.98) as compared to unenhanced CT (0.91, 95% CI 0.87 to 0.93). Summary sensitivity of CT with oral contrast enhancement (0.89, 95% CI 0.81 to 0.94) was similar to summary sensitivity of unenhanced CT. Results showed no differences in summary specificity, which varied from 0.93 (95% CI 0.90 to 0.95) to 0.95 (95% CI 0.90 to 0.98) between subgroups. Summary sensitivity for low-dose CT (0.94, 95% 0.90 to 0.97) was similar to summary sensitivity for standard- or unspecified-dose CT (0.95, 95% 0.93 to 0.96). Summary specificity did not differ between low-dose and standard- or unspecified-dose CT.

In meta-regression analyses, summary sensitivity was statistically significantly higher in studies using CT-scanners with 16 or more detector rows, and in studies where CT-scans were evaluated by senior radiologists. Summary specificity did not differ significantly between groups in these analyses. Results showed no statistically significant differences in summary sensitivity or specificity between studies that recruited participants with an intermediate suspicion of acute appendicitis due to an equivocal presentation and studies that recruited participants with any suspicion of appendicitis. The methodological quality of the included studies was generally poor, particularly for the reference test and the flow and timing domains.

Strengths and weaknesses of the review

The major strengths of this review are that we adhered to recommended review methods and performed an extensive search of the literature without language restrictions and filters to target diagnostic test accuracy studies. We included data from 64 studies and produced a comprehensive review of the accuracy of CT for appendicitis in adults. Because of challenges related to differential and partial verification in this area, we focused on prospective studies to limit potential bias from retrospective studies with missing reference standard outcomes in participants who did not have surgery. In subgroup analyses, we explored the accuracy of different CT-protocols characterised by type of contrast enhancement and radiation dose. We also assessed the influence of CT-scanner generation, radiologists' experience, disease spectrum, and methodological quality on summary estimates of sensitivity and specificity.

We noted several limitations in the review process. In some study reports, the reporting quality made it difficult to assess whether data collection was conducted prospectively or retrospectively. In most of these situations, we contacted the corresponding author and excluded the study if we received no reply. However, for some studies, our judgements may have been too liberal. In general, we accepted studies as having prospective data collection if study authors used the term 'prospective' or 'consecutive' to characterise

the data collection, and if we found no clear-cut evidence to suggest the contrary (i.e. statements that participants were selected from databases or registries). As in previous systematic reviews in this and related areas, we decided to exclude studies using retrospective data collection from registers and hospital records to reduce potential bias from partial verification (Al-Khayal 2007; Ebell 2014; Terasawa 2004; van Randen 2008; Xiong 2015). Hospital records may not contain the necessary information, participants may be treated in other hospitals, and telephone follow-up after, say, 12 months is unlikely to be successful for all participants. However, the basis for this decision could be questioned due to the low standards of follow-up in the prospective studies included. Also, follow-up in the included studies was often based on reviews of hospital records for alternative diagnoses and a check that appendicectomy was not performed during the follow-up interval. Among the 71 studies that we excluded due to retrospective data collection, participants were selected following an appendicectomy and preoperative CT in 28 studies. The prevalence of appendicitis is high and the proportion with a negative CT outcome is correspondingly low in such studies; it follows that resulting estimates of specificity are unlikely to be applicable to CT-negatives in general. In another 38 of the retrospective studies, participants were selected from registries or databases. In most of these studies, follow-up of participants who did not have surgery was based on review of hospital records for alternative diagnoses and readmission; however, in a few studies, telephone interviews were also performed, but the response rate generally was not reported. In addition, our adaptation of Quality Assessment of Studies of Diagnostic Accuracy - Revised (QUADAS-2) included a definition for an adequate follow-up period, which lasted seven to 31 days. We admit this is arbitrary, but we maintain that length of follow-up is important for assessing the quality of follow-up. We believe that a follow-up period of seven to 31 days is sufficiently long to capture missed cases and is sufficiently short that new events are not captured.

Another limitation was that we did not distinguish between uncomplicated and complicated acute appendicitis as separate target conditions. This distinction is becoming increasingly relevant with emerging evidence of antibiotic therapy as an alternative to surgery in persons with uncomplicated acute appendicitis, because selection of persons for antibiotic therapy depends on the finding of uncomplicated acute appendicitis on CT (Salminen 2015; Vons 2011). Misclassification of complicated appendicitis as uncomplicated is a likely explanation for failure of antibiotic therapy.

Finally, it was not feasible to contact the authors of 28 studies including paediatric participants with a request for subgroup results for participants older than 14 years of age. Instead we decided to include these studies and perform a sensitivity analysis that revealed no difference in summary sensitivity and specificity between studies with and without paediatric participants.

The major limitation of the included studies was poor methodological quality. However, the impact of low methodological quality appears to be negligible for the patient selection domain and the index test domain as there was practically no difference in summary estimates between the overall meta-analysis and sensitivity analyses in studies with low risk of bias for these domains. Poor scorings in the reference standard domain and in the flow and timing domain were due to low quality of follow-up and partial verification. Differential verification appears to be inevitable in accuracy studies of CT for acute appendicitis, and this increases the demand for rigorous follow-up. In most studies, the majority of CT-positive participants had surgery and CT-negative participants generally had follow-up because it was considered unethical to expose CT-negative patients to surgery that was likely to be unnecessary. An important finding was the multitude of methods applied to perform follow-up, which ranged from checking hospital records for readmissions to using standardised regimens including telephone interviews or outpatient consultations within a predefined time frame. Accordingly, we considered follow-up as inadequate or insufficiently described in 42 studies. Another important piece of information that was often missing was the proportion of participants who had received follow-up as planned. We assumed that follow-up was complete when all participants were included in the 2×2 table, but this may be optimistic.

It could be argued that follow-up is irrelevant when an alternative diagnosis (e.g. diverticulitis, pelvis inflammatory disease, ureter stone) was made that explained participants' abdominal pain. The frequency of alternative diagnoses besides non-specific abdominal pain in participants without appendicitis was reported in 27 studies for 29 study populations. The median frequency was 0.56, with interquartile range 0.34 to 0.62 and range 0.13 to 0.94. It could be countered that although an alternative diagnosis rules out appendicitis in some cases, an alternative diagnosis may be less reliable in others; therefore it may not necessarily rule out appendicitis in all participants who do not have surgery.

In our view, the major problem incurred by low-quality follow-up and loss to follow-up is the partial verification that results. Partial verification has been associated with higher estimates of sensitivity in diagnostic accuracy studies in general (Whiting 2013), and we suspect that a similar association could exist in the studies that we reviewed. Unfortunately, it was not feasible to investigate if and to what extent low methodological quality in the reference standard domain and in the flow and timing domain impacted summary estimates due to the small number of studies with adequate and complete follow-up.

Another limitation of the included studies relates to the paucity of studies with direct comparisons of different CT-protocols using a paired or randomised design. We included nine such studies, but the number of primary analyses in these studies was too low for comparative meta-analyses to be performed to assess the influence of types of contrast enhancement and radiation dose. All comparisons that we made are indirect, and it is important to be aware that such comparisons may be confounded by factors such as differences in population characteristics, properties of the CT-scanner, radiologists' experience, and study methods. Nevertheless, our finding of similar accuracy for low-dose and standard-dose CT corresponds with results from a recent multi-centre study in which persons with suspected appendicitis were randomly allocated to low-dose and standard-dose CT (The Locat Group 2017). In addition, findings of lower sensitivity for unenhanced CT and no gain in accuracy from supplementing IV contrast with oral contrast enhancement are in line with the results from a retrospective study in 9047 adult persons who underwent appendicectomy in 56 hospitals in the USA (Drake 2014).

Applicability of findings to the review question

Participants in the included studies were predominantly adult or adolescent persons above 14 years of age with suspected appendicitis who were recruited in urban university hospitals. The suspicion of acute appendicitis was based on history, physical examination findings, and results of routine laboratory tests and urinalysis. Studies in persons who underwent ultrasonography before CT were excluded. We found no statistical evidence to show that summary estimates of accuracy differed between subgroups of studies that included persons with an intermediate suspicion of appendicitis due to an equivocal presentation and studies in persons recruited with any suspicion of appendicitis. Results from the primary studies cover a wide range of CT-scanners, CT-protocols, types of contrast enhancement, and radiation doses. Based on this, we believe that the findings presented in this review are applicable to most persons above 14 years of age with suspected appendicitis following initial evaluation. Our meta-regression analyses indicate that overall summary estimates of sensitivity may not be representative in two settings. In settings using newer CT-scanners (16 or more channels), sensitivity is likely to be higher. Conversely, in settings with in-training radiologists, sensitivity is likely to be lower. Again, these findings should be interpreted cautiously due to possible confounding by other factors.

Previous research

The results of our meta-analyses are consistent with the results from previous meta-analyses that are presented in Table 6.

AUTHORS' CONCLUSIONS

Implications for practice

Sensitivity and specificity of CT for diagnosing acute appendicitis in adults are high, hence the use of CT is likely to assist clinicians in treating persons with possible appendicitis. Unenhanced standard-dose CT appears to have lower sensitivity than standard-dose CT with IV, rectal, or oral and IV contrast enhancement. Use of different types of contrast enhancement or no enhancement does not appear to affect specificity. Differences in sensitivity and specificity between low-dose and standard-dose CT appear to be negligible. In adult persons, it seems that low-dose CT should be preferred over standard-dose CT as a first-line imaging test, with standard-dose CT reserved for persons with inconclusive findings on low-dose CT. To minimise radiation exposure, clinicians should critically assess whether additional information from CT imaging is needed for decision-making about surgery, watchful waiting, or discharge. Results of this review should be interpreted with caution for two reasons. First, the results are based on studies of low methodological quality. Second, the comparisons between types of contrast enhancement and radiation dose may be unreliable because they are based on indirect comparisons that may be confounded by other factors.

Implications for research

Future research should focus on low-dose CT and should corroborate the finding of equal accuracy between low-dose and standard-dose CT. Most existing studies have been performed in Asian populations (Chang 2016; Kim 2011; Kim 2012; Seo 2009; The Locat Group 2017; Yun 2016), three studies have been performed in European populations (Keyzer 2004; Keyzer 2009; Platon 2009), and two studies in paediatric populations have been performed in the USA (Callahan 2015; Didier 2015). Such studies should be designed as paired or randomised studies to minimise confounding from other factors that may influence accuracy. This research should also explore the influence of body mass index and whether contrast enhancement improves accuracy compared to unenhanced low-dose CT. Results from the recent LOCAT study indicate that intravenous contrast enhancement is not needed when low-dose CT is used (The Locat Group 2017).

The issue of contrast enhancement is also unsettled for standard-dose CT; we included five randomised trials and one paired study that compared the accuracy of different types of contrast enhancement. More such studies are needed to weigh up reliably estimated gains in sensitivity and specificity with risks and inconveniences related to intravenous, oral, and rectal contrast enhancement.

To minimise radiation expose and costs, future research should continue to explore the performance of existing clinical decision rules in identifying persons with suspected appendicitis that can be managed without the use of CT. Meta-analyses of the performance of the Alavarado Score have suggested that appendicitis can be ruled out in persons with low scores and ruled in among persons with high scores, but results were heterogeneous, and assessment of methodological quality demonstrated risk of verification bias (Ebell 2014; Ohle 2011). Several observational studies have explored consequences in terms of missed diagnoses and negative appendicectomies of limiting CT to persons with intermediate outcomes on the Alvarado Score (Coleman 2018; McKay 2007; Scott 2015), as well as the Adult Appendicitis Score (Sammalkorpi 2017). Results from a recent trial indicated that the need for imaging tests can be reduced even further. In this trial, persons with intermediate outcomes on the Appendicitis Inflammatory Response Score were randomly allocated to have mandatory or selective imaging (CT or ultrasonography (US)). There was no difference between groups in negative appendicectomy rate nor missed appendicitis rate at 30 days (Andersson 2017). This selective use of CT is supported by our finding that summary sensitivity and specificity for CT did not differ between study populations with intermediate suspicion due to an equivocal presentation and any suspicion of appendicitis.

In future systematic reviews in this area, study selection criteria require careful attention. All studies using retrospectively collected data to reduce potential bias from partial verification may exclude relevant information. Instead study authors should define minimum requirements for adequate follow-up, or, alternatively, should include all studies and explore whether the quality of follow-up affects summary estimates of sensitivity and specificity. A special caveat concerns studies in cohorts of persons selected following an appendicectomy and a CT-scan because clinically applicable estimates of specificity are unlikely to result from such studies.

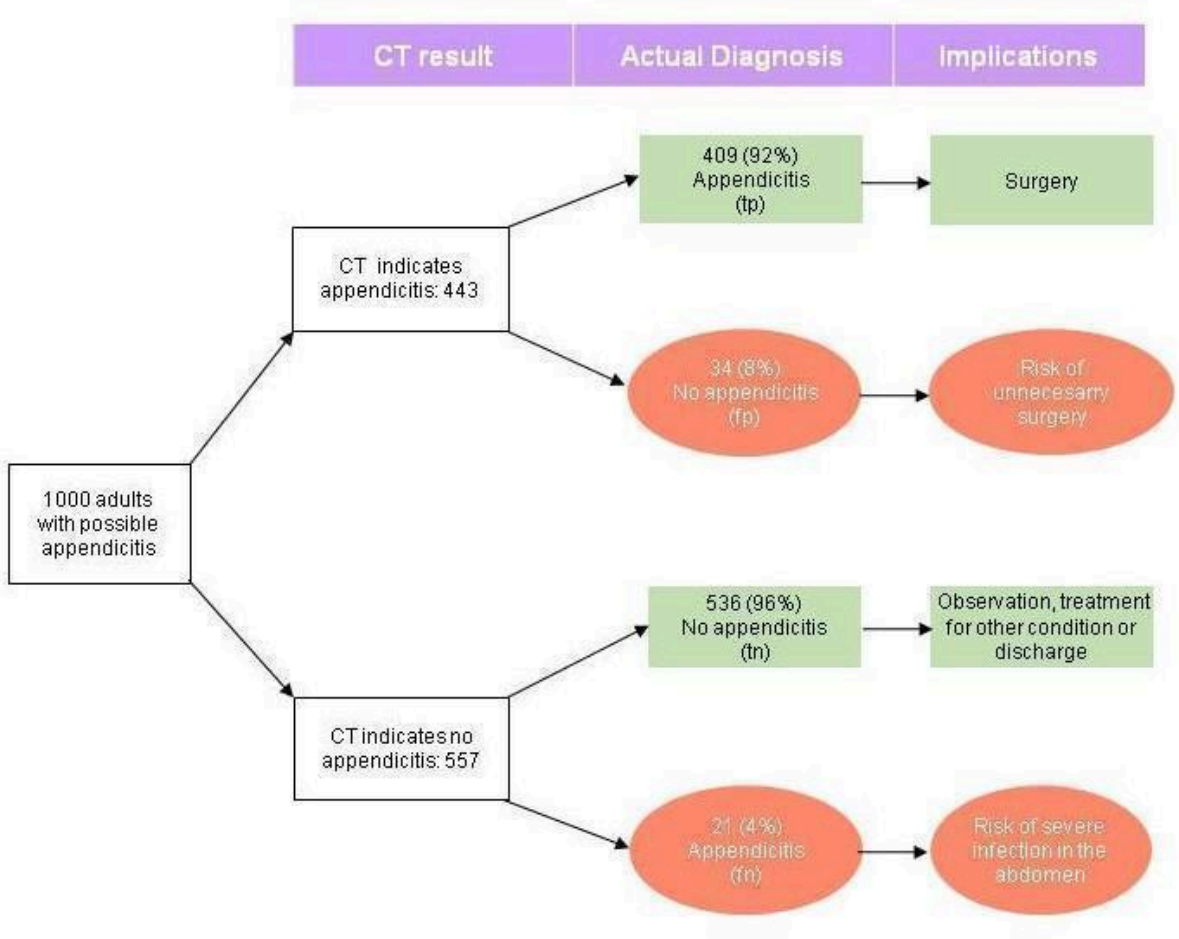
Future studies of the accuracy of CT for acute appendicitis should adhere to the updated STARD statement to improve the quality of reporting (Bossuyt 2015). Moreover, rigorous follow-up of participants who do not have surgery should receive special attention in the planning and conduct of such studies because differential verification appears to be inevitable in this area. In general, follow-up should be complete and careful, and should be of the right duration. In particular, follow-up should consist

of obtaining a reliable alternative diagnosis as well as contacting participants to check that symptoms have resolved, and that surgery or antibiotic therapy has not taken place elsewhere. Authors of future studies should consider strategies used to reduce loss to follow-up in other types of research such as clinical trials, surveys, and longitudinal studies. These methods include minimising inconvenience, providing monetary incentives, and collecting all available contact information from participants, family members, or other locators (Bower 2014; Brueton 2013; Woolard 2004). Despite all efforts, some participants will be lost to follow-up. It is important that the number of these participants is reported. Moreover, sensitivity analyses should be performed to assess the potential consequences of loss to follow-up for sensitivity and specificity.

Finally, the use of antibiotic therapy among participants in upcoming studies will add to the complexity of disease verification. A definitive reference standard would be available only for those who did not improve on antibiotics and underwent subsequent surgery.

Figure 14 presents a flow diagram for the plain language summary.

Figure 14. Plain language summary flowchart.



tp: true positive – test is positive (indicates appendicitis) and patient has appendicitis
fp: false positive – test is positive (indicates appendicitis) but patient does not have appendicitis
tn: true negative – test is negative (indicates no appendicitis) and patient does not have appendicitis
fn: false negative – test is negative (indicates no appendicitis) but patient has appendicitis

ACKNOWLEDGEMENTS

- Michael Wilhelmsen, MD, is acknowledged for his contribution to the literature search
- Julie Blockman, MD, is acknowledged for entering data
- Sys Johnsen, CCCG Trials Search Co-ordinator, is acknowledged for developing and updating the electronic search strategy
- Lydur Olufsson, MD, and Bent Carstenskjold-Seidenfarden, MD, are acknowledged for their contributions to the protocol
- Anna Aaresøn, MD, is acknowledged for extracting data from Nemsadze 2009, which was reported in Russian

- Fatma Kara, MD, is acknowledged for extracting data from Yüsekkaya 2004, which was reported in Turkish

No grants supported this study.

.

REFERENCES

References to studies included in this review

Antevil 2006 *{published data only}*

Antevil JL, Rivera L, Langenberg BJ, Hahm G, Favata MA, Brown CV. Computed tomography-based clinical diagnostic pathway for acute appendicitis: prospective validation. *Journal of American College of Surgeons* 2006;**203**(6):849-56.

Ashraf 2006 *{published data only}*

Ashraf K, Ashraf O, Bari V, Rafique MZ, Usman MU, Chisti I. Role of focused appendiceal computed tomography in clinically equivocal acute appendicitis. *Journal of the Pakistan Medical Association* 2006;**56**(5):200-3.

Atema 2015 *{published data only}*

Atema J, Stoker J, Boermeester M. Comparison of imaging strategies with conditional versus immediate contrast-enhanced CT in patients clinically suspected of having acute appendicitis. *Surgical Infections* 2014;**15**(3):A5.

* Atema JJ, Gans SL, Van Randen A, Lameris W, van Es HW, van Heesewijk JP, et al. Comparison of imaging strategies with conditional versus immediate contrast-enhanced computed tomography in patients with clinical suspicion of acute appendicitis. *European Radiology* 2015;**25**(8):2445-52.

Atema JJ, Gans SL, Van Randen A, .Lameris W, Van Es WH, Van Heesewijk JPM, et al. CT strategy for patients with suspected acute appendicitis; comparison of conditional and immediate CT [Dutch]. *Nederlands Tijdschrift voor Geneeskunde* 2016;**160**(9):A9603.

Balthazar 1991 *{published data only}*

Balthazar EJ, Megibow AJ, Siegel SE, Birnbaum BA. Appendicitis: prospective evaluation with high-resolution CT. *Radiology* 1991;**180**(1):21-4.

Balthazar 1994 *{published data only}*

Balthazar EJ, Birnbaum BA, Yee J, Megibow AJ, Roshkow J, Gray C. Acute appendicitis: CT and US correlation in 100 patients. *Radiology* 1994;**190**(1):31-5.

Bouillot 2001 *{published data only}*

Bouillot JL, Ruiz A, Alamowitch B, Capuano G, Aouad K, Fourmestraux J, et al. [Suspected acute appendicitis. Role of enhanced helical computed tomography. Prospective study of 100 patients]. *Annales de Chirugie* 2001;**126**(5):427-33.

Cakirer 2002 *{published data only}*

Cakirer S, Basak M, Colakoglu B, Bankaoglu M. Diagnosis of acute appendicitis with unenhanced helical CT: a study of 130 patients. *Emergency Radiology* 2002;**9**(3):155-61.

Christopher 2002 *{published data only}*

Christopher FL, Lane MJ, Ward JA, Morgan JA. Unenhanced helical CT scanning of the abdomen and pelvis changes disposition of patients presenting to the emergency department with possible acute appendicitis. *Journal of Emergency Medicine* 2002;**23**(1):1-7.

Cougard 2002 *{published data only}*

Cougard P, Robert-Valla C, Coudert M, Goudet P. [Helical computed tomography in the diagnosis of acute appendicitis]. *Annals de Chirugie* 2002;**127**(1):73-4.

del Cura 2000 *{published data only}*

* del Cura JL, Oleaga L, Grande D, Elizagaray E, Campo M. Rectal contrast-enhanced computed tomography in the diagnosis of acute appendicitis [Spanish]. *Radiologia* 2000;**42**(9):481-91.

del Cura JL, Oleaga L, Grande D, Farina MA, Isusi M. Comparison of ultrasonography and computed tomography in the diagnosis of acute appendicitis [Spanish]. *Radiologia* 2001;**43**(4):175-86.

del Cura JL, Oleaga L, Grande D, Vela AC, Ibanez AM. Reliability of diagnostic imaging techniques in suspected acute appendicitis: proposed diagnostic protocol [Spanish]. *Radiologia* 2001;**43**(10):478-89.

Funaki 1998 *{published data only}*

Funaki B, Grosskreutz SR, Funaki CN. Using unenhanced helical CT with enteric contrast material for suspected appendicitis in patients treated at a community hospital. *AJR American Journal of Roentgenology* 1998;**171**(4):997-1001.

Gamanagatti 2007 *{published data only}*

Gamanagatti S, Vashisht S, Kapoor A, Chumber S, Bal S. Comparison of graded compression ultrasonography and unenhanced spiral computed tomography in the diagnosis of acute appendicitis. *Singapore Medical Journal* 2007;**48**(1):80-7.

Hekimoglu 2011 *{published data only}*

Hekimoglu K, Yildirim UM, Karabulut E, Coskun M. Comparison of combined oral and i.v. contrast-enhanced versus single i.v. contrast-enhanced mdct for the detection of acute appendicitis. *JBR-BTR* 2011;**94**(5):278-82.

Hershko 2002 *{published data only}*

Hershko DD, Sroka G, Bahouth H, Ghersin E, Mahajna A, Krausz MM. The role of selective computed tomography in the diagnosis and management of suspected acute appendicitis. *American Surgeon* 2002;**68**(11):1003-7.

Hershko 2007 *{published data only}*

Hershko DD, Awad N, Fischer D, Mahajna A, Guralnik L, Israelit SH, 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. *Diseases of the Colon & Rectum* 2007;**50**(8):1223-9.

Holloway 2003 *{published data only}*

Holloway JA, Westerbuhr LM, Chain J, Forney GA, White TW, Hughes RJ, et al. Is appendiceal computed tomography in a community hospital useful?. *American Journal of Surgery* 2003;**186**(6):682-4.

Hong 2003 *{published data only}*

Hong JJ, Cohn SM, Ekeh AP, Newman M, Salama M, Leblang SD. A prospective randomized study of clinical assessment versus computed tomography for the diagnosis of acute appendicitis. *Surgical Infections* 2003;**4**(3):231-9.

Horton 2000 *{published data only}*

Horton MD, Counter SF, Florence MG, Hart MJ. A prospective trial of computed tomography and ultrasonography for diagnosing appendicitis in the atypical patient. *American Journal of Surgery* 2000;**179**(5):379-81.

in't Hof 2004 *{published data only}*

In't Hof KH, Krestin GP, Steijerberg EW, Bonjer HJ, Lange JF, Becking WB, et al. Interobserver variability in CT scan interpretation for suspected acute appendicitis. *Emergency Medicine Journal* 2009;**26**(2):92-4.

* In't Hof KH, Van Lankeren W, Krestin GP, Bonjer HJ, Lange JF, Becking WB, et al. Surgical validation of unenhanced helical computed tomography in acute appendicitis. *British Journal of Surgery* 2004;**91**(12):1641-5.

Jacobs 2001 *{published data only}*

Jacobs JE, Birnbaum BA, Macari M, Megibow AJ, Israel G, Maki DD, et al. Acute appendicitis: comparison of helical CT diagnosis - focused technique with oral contrast material versus nonfocused technique with oral and intravenous contrast material. *Radiology* 2001;**220**(3):683-90.

Jo 2010 *{published and unpublished data}*

* Jo YH, Kim K, Rhee JE, Kim TY, Lee JH, Kang SB, et al. The accuracy of emergency medicine and surgical residents in the diagnosis of acute appendicitis. *American Journal of Emergency Medicine* 2010;**28**(7):766-70.

Kan 2001 *{published data only}*

Kan JH, Fines BP, Funaki B. Conventional and hydrocolonic US of the appendix with CT correlation performed by on-call radiology residents. *Academic Radiology* 2001;**8**(12):1208-14.

Karabulut 2014 *{published data only}*

Karabulut N, Kiroglu Y, Herek D, Kocak TB, Erdur B. Feasibility of low-dose unenhanced multi-detector CT in patients with suspected acute appendicitis: comparison with sonography. *Clinical Imaging* 2014;**38**(3):296-301.

Kepner 2012 *{published data only}*

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. *American Journal of Emergency Medicine* 2012;**30**(9):1765-73.

Keyzer 2004 *{published data only}*

* Keyzer C, Tack D, De Maertelaer V, Bohy P, Gevenois PA, Van Gansbeke D. Acute appendicitis: comparison of low-dose and standard-dose unenhanced multi-detector row CT. *Radiology* 2004;**232**(1):164-72.

Keyzer C, Zalcman M, De Maertelaer V, Coppens E, Bali M-A, Gevenois PA, et al. Comparison of US and unenhanced multi-detector row CT in patients suspected of having acute appendicitis. *Radiology* 2005;**236**(2):527-34.

Keyzer 2009 *{published data only}*

Keyzer C, Cullus P, Tack D, De Maertelaer V, Bohy P, Gevenois PA. MDCT for suspected acute appendicitis in adults: impact of oral and IV contrast media at standard-dose and simulated low-dose techniques. *AJR American Journal of Roentgenology* 2009;**193**(5):1272-81.

Kim 2008 *{published data only}*

Kim K, Rhee JE, Lee CC, Kim KS, Shin JH, Kwak MJ, et al. Impact of helical computed tomography in clinically evident appendicitis. *Emergency Medical Journal* 2008;**25**(8):477-81.

Kim 2012 *{published data only}*

Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *New England Journal of Medicine* 2012;**366**(17):1596-605.

Lane 1999 *{published data only}*

Lane MJ, Liu DM, Huynh MD, Jeffrey RB Jr, Mindelzun RE, Katz DS. Suspected acute appendicitis: nonenhanced helical CT in 300 consecutive patients. *Radiology* 1999;**213**(2):341-6.

Lopez 2007 *{published data only}*

Lopez PP, Cohn SM, Popkin CA, Jackowski J, Michalek JE. The use of a computed tomography scan to rule out appendicitis in women of childbearing age is as accurate as clinical examination: a prospective randomized trial. *American Surgeon* 2007;**73**(12):1232-6.

Malone 1993 *{published data only}*

Malone AJ Jr, Wolf CR, Malmed AS, Melliere BF. Diagnosis of acute appendicitis: value of unenhanced CT. *AJR American Journal of Roentgenology* 1993;**160**(4):763-6.

Maluccio 2001 *{published data only}*

Maluccio MA, Covey AM, Weyant MJ, Eachempati SR, Hydo LJ, Barie PS. A prospective evaluation of the use of emergency department computed tomography for suspected acute appendicitis. *Surgical Infections (Larchmt.)* 2001;**2**(3):205-11.

Megibow 2002 *{published data only}*

Megibow AJ, Rusinek H, Lisi V, Bennett GL, Macari M, Israel GM, et al. Computed tomography diagnosis utilizing compressed image data: an ROC analysis using acute appendicitis as a model. *Journal of Digital Imaging* 2002;**15**:84-90.

Mittal 2004 *{published data only}*

Mittal VK, Goliath J, Sabir M, Patel R, Richards BF, Alkalay I, et al. Advantages of focused helical computed tomographic scanning with rectal contrast only vs triple contrast in the diagnosis of clinically uncertain acute appendicitis: a prospective randomized study. *Archives of Surgery* 2004;**139**(5):495-500.

Moteki 2009 *{published data only}*

Moteki T, Ohya N, Horikoshi H. Prospective examination of patients suspected of having appendicitis using new computed tomography criteria including "maximum depth of intraluminal

appendiceal fluid greater than 2.6 mm". *Journal of Computer Assisted Tomography* 2009;**33**(3):383-9.

Nathan 2008 *{published data only}*

Nathan RO, Blackmore CC, Jarvik JG. Therapeutic impact of CT of the appendix in a community hospital emergency department. *AJR American Journal of Roentgenology* 2008;**191**(4):1102-6.

Nemsadze 2009 *{published data only}*

Nemsadze GS, Urushadze OP, Tokhadze LT, Lomidze MN, Kipshidze NN. [The role and place of helical ct for preoperative diagnosis of acute appendicitis]. *Georgian Medical News* 2009;**174**(9):10-3.

Ozturk 2014 *{published and unpublished data}*

* Ozturk A, Bozkurtoglu H, Uckurt Y, Kaya C, Yananli ZD, Akinci OF. The effect of computed tomography on surgeon's decisions in suspected appendicitis cases. *Journal of Clinical and Analytical Medicine* 2015;**6**(6):No pagination.

Ozturk A, Yananli Z, Atalay T, Akinci OF. The comparison of the effectiveness of tomography and Alvarado scoring system in patients who underwent surgery with the diagnosis of appendicitis. *Turkish Journal of Surgery* 2016;**32**:111-4.

Pakaneh 2008 *{published data only}*

Pakaneh MA, Fazeli MS, Ghanaati H, Kaviani A, Aminian A, Zobeiri T, et al. Non-contrast spiral computed tomography in diagnosis of acute appendicitis. *Iranian Journal of Radiology* 2008;**5**(1):1-6.

Park 2016 *{published data only}*

Park JH, Kim B, Kim MS, Kim HJ, Ko Y, Ahn S, et al. Comparison of filtered back projection and iterative reconstruction in diagnosing appendicitis at 2-mSv CT. *Abdominal Radiology* 2016;**41**:1227-36.

Pickuth 2001 *{published data only}*

Pickuth D, Heywang-Kobrunner SH, Spielmann RP. Suspected acute appendicitis: is ultrasonography or computed tomography the preferred imaging technique?. *European Journal of Surgery* 2000;**166**(4):315-9.

* Pickuth D, Spielmann RP. Unenhanced spiral CT for evaluating acute appendicitis in daily routine. A prospective study. *Hepatogastroenterology* 2001;**48**(37):140-2.

Platon 2009 *{published data only}*

Platon A, Jlassi H, Rutschmann OT, Becker CD, Verdun FR, Gervaz P, et al. Evaluation of a low-dose CT protocol with oral contrast for assessment of acute appendicitis. *European Radiology* 2009;**19**(2):446-54.

Poortman 2003 *{published data only}*

Poortman P, Lohle PN, Schoemaker CM, Oostvogel HJ, Teepen HJ, Zwinderman KA, et al. Comparison of CT and sonography in the diagnosis of acute appendicitis: a blinded prospective study. *AJR American Journal of Roentgenology* 2003;**181**(5):1355-9.

Rao 1997 *{published data only}*

* Rao PM, Rhea JT, Novelline RA, Mccabe CJ, Lawrason JN, Berger DL, et al. Helical CT technique for the diagnosis of appendicitis: prospective evaluation of a focused appendix CT examination. *Radiology* 1997;**202**(1):139-44.

Rao PM, Wittenberg J, McDowell RK, Rhea JT, Novelline RA. Appendicitis: use of arrowhead sign for diagnosis at CT. *Radiology* 1997;**202**(2):363-6.

Rao 1998 *{published data only}*

Rao PM, Rhea JT, Novelline RA, Mostafavi AA, Lawrason JN, Mccabe CJ. Helical CT combined with contrast material administered only through the colon for imaging of suspected appendicitis. *AJR American Journal of Roentgenology* 1997;**169**(5):1275-80.

* Rao PM, Rhea JT, Novelline RA, Mostafavi AA, Mccabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *New England Journal of Medicine* 1998;**338**(3):141-6.

Rao 1999 *{published data only}*

Rao PM, Feltmate CM, Rhea JT, Schulick AH, Novelline RA. Helical computed tomography in differentiating appendicitis and acute gynecologic conditions. *Obstetetrics and Gynecology* 1999;**93**(3):417-21.

Repplinger 2015 *{published and unpublished data}*

Harringa J, Kinner S, Riedesel E, Gill K, Ziemlewicz T, Robbins J, et al. Prospective comparison of contrast-enhanced magnetic resonance imaging versus contrast-enhanced computed tomography for suspected appendicitis in children and young adults. *Annals of Emergency Medicine* 2016;**68 (4 Suppl 1)**:S147-8.

* Repplinger MD, Pickhardt P, Kitchin D, Robbins J, Ziemlewicz T, Reeder SB. Direct comparison of a magnetic resonance imaging protocol with contrast-enhanced computed tomography to diagnose appendicitis. *Annals of Emergency Medicine* 2014;**1**:S21.

Repplinger MD, Pickhardt P, Kitchin D, Robbins J, Ziemlewicz T, Reeder SB. Prospective evaluation of contrast-enhanced magnetic resonance imaging for suspected appendicitis. *Academic Emergency Medicine* 2015;**1**:S402-S403.

Sammalkorpi 2017 *{published data only}*

Sammalkorpi HE, Leppaniemi A, Lantto E, Mentula P. Performance of imaging studies in patients with suspected appendicitis after stratification with adult appendicitis score. *World Journal of Emergency Surgery* 2017;**12**:6.

Scott 2015 *{published data only (unpublished sought but not used)}*

Scott AJ, Mason SE, Arunakirinathan M, Reissis Y, Kinross JM, Smith JJ. Risk stratification by the Appendicitis Inflammatory Response score to guide decision-making in patients with suspected appendicitis. *British Journal of Surgery* 2015;**102**(5):563-72.

Sim 2013 *{published and unpublished data}*

Sim JY, Kim HJ, Yeon JW, Suh BS, Kim KH, Ha YR, et al. Added value of ultrasound re-evaluation for patients with equivocal CT findings of acute appendicitis: a preliminary study. *European Radiology* 2013;**23**(7):1882-90.

Stacher 1999 *{published data only}*

Stacher R, Portugaller H, Preidler KW, Ruppert-Kohlmayr AJ, Anegg U, Rabl H, et al. [Acute appendicitis in non-contrast spiral CT: a diagnostic luxury or benefit?]. *RoFo* 1999;**171**(1):26-31.

Tan 2015 *{published and unpublished data}*

* Tan WJ, Acharyya S, Goh YC, Chan WH, Wong WK, Ooi LL, et al. Prospective comparison of the Alvarado score and CT scan in the evaluation of suspected appendicitis: a proposed algorithm to guide CT use. *Journal of the American College of Surgeons* 2015;**220**(2):218-24.

Togawa 2005 *{published data only}*

Togawa A, Kimura F, Chiku T, Sano W, Tashiro T, Miyazaki M. Simple way to improve accuracy in diagnosis of quadrant inflammatory disease: how to avoid adverse laparotomy by using plain CT. *Hepato-Gastroenterology* 2005;**52**(61):135-8.

Torbati 2003 *{published data only}*

Torbati SS, Guss DA. Impact of helical computed tomography on the outcomes of emergency department patients with suspected appendicitis. *Academic Emergency Medicine* 2003;**10**(8):823-9.

Tsai 2001 *{published data only}*

Tsai SS, Coughlin BF, Hampf FE, Munshi IA, Wolfe JM. Diagnosing appendicitis with CT and ultrasound using prospective patient stratification by body mass index. *Emergency Radiology* 2001;**8**(5):267-71.

Uzunosmanoglu 2017 *{published and unpublished data}*

Uzunosmanoglu H, Cevik Y, Corbacioglu SK, Akinci E, Bulus H, Agladioglu K. Diagnostic value of appendicular Doppler ultrasonography in acute appendicitis. *Turkish Journal of Trauma & Emergency Surgery* 2017;**23**(3):188-92.

Walker 2000 *{published data only}*

Walker S, Haun W, Clark J, McMillin K, Zeren F, Gilliland T. The value of limited computed tomography with rectal contrast in the diagnosis of acute appendicitis. *American Journal of Surgery* 2000;**180**(6):450-4.

Wang 2012 *{published and unpublished data}*

Wang SY, Fang JF, Liao CH, Kuo IM, Yang CH, Yeh CN, et al. Prospective study of computed tomography in patients with suspected acute appendicitis and low Alvarado score. *American Journal of Emergency Medicine* 2011;**30**(8):1597-601.

Weltman 2000 *{published data only}*

Weltman DI, Yu J, Krumenacker J Jr, Huang S, Moh P. Diagnosis of acute appendicitis: comparison of 5- and 10-mm CT sections in the same patient. *Radiology* 2000;**216**(1):172-7.

Wijetunga 2001 *{published data only}*

Wijetunga R, Tan BS, Rouse JC, Bigg-Wither GW, Doust BD. Diagnostic accuracy of focused appendiceal CT in clinically equivocal cases of acute appendicitis. *Radiology* 2001;**221**(3):747-53.

Wilson 2001 *{published data only}*

Wilson EB, Cole JC, Nipper ML, Cooney DR, Smith RW. Computed tomography and ultrasonography in the diagnosis of appendicitis: when are they indicated?. *Archieves of Surgery* 2001;**136**(6):670-5.

Wise 2001 *{published data only}*

Wise SW, Labuski MR, Kasales CJ, Blebea JS, Meilstrup JW, Holley GP, et al. Comparative assessment of CT and sonographic techniques for appendiceal imaging. *AJR American Journal of Roentgenology* 2001;**176**(4):933-41.

Wong 2002 *{published data only}*

Wong SK, Chan LP, Yeo A. Helical CT imaging of clinically suspected appendicitis: correlation of CT and histological findings. *Clinical Radiology* 2002;**57**(8):741-5.

Yuksekkaya 2004 *{published data only}*

Yuksekkaya R, Akgul E, Inal M, Binokay F, Celiktas M, Aksungur E. [Unenhanced spiral CT in the diagnosis of acute appendicitis]. *Tanisal ve Girisimsel Radyoloji* 2004;**10**(2):131-9.

References to studies excluded from this review

Abo 2011 *{published data only}*

Abo A, Shannon M, Taylor G, Bachur R. The influence of body mass index on the accuracy of ultrasound and computed tomography in diagnosing appendicitis in children. *Pediatric Emergency Care* 2011;**27**(8):731-6.

Al-Faouri 2016 *{published data only}*

Al-Faouri AF, Ajarma KY, Al-Abbadi AM, Al-Omari AH, Almunaizel TS, Alzu'bi AA, et al. The Alvarado score versus computed tomography in the diagnosis of acute appendicitis: a prospective study. *Medical Journal Armed Forces India* 2016;**72**:332-7.

Ali 2017 *{published data only}*

Ali MT, Nabil DM. The role of 256-slice CT in differentiation between non-perforated and perforated appendicitis. *Egyptian Journal of Radiology and Nuclear Medicine* 2017;**48**:15-21.

Anderson 2009 *{published data only}*

Anderson SW, Rhea JT, Milch HN, Ozonoff A, Lucey BC, Soto JA. Influence of body habitus and use of oral contrast on reader confidence in patients with suspected acute appendicitis using 64 MDCT. *Emergency Radiology* 2010;**17**(6):445-53.

* Anderson SW, Soto JA, Lucey BC, Ozonoff A, Jordan JD, Ratevosian J, et al. Abdominal 64-MDCT for suspected appendicitis: the use of oral and IV contrast material versus IV contrast material only. *American Journal of Radiology* 2009;**193**:1282-8.

Antevil 2004 *{published data only}*

Antevil J, Rivera L, Langenberg B, Brown CV. The influence of age and gender on the utility of computed tomography to diagnose acute appendicitis. *American Surgeon* 2004;**70**(10):850-3.

Bachar 2013 *{published data only}*

Bachar I, Perry ZH, Dukhno L, Mizrahi S, Kirshtein B. Diagnostic value of laparoscopy, abdominal computed tomography, and ultrasonography in acute appendicitis. Journal of Laparoscopic & Advanced Surgical Techniques 2013; Vol. 23, issue 12:982-9.

Balthazar 1998 *{published data only}*

Balthazar EJ, Rofsky NM, Zucker R. Appendicitis: the impact of computed tomography imaging on negative appendectomy and perforation rates. *American Journal of Gastroenterology* 1998;**93**(5):768-71.

Bendeck 2002 *{published data only}*

Bendeck SE, Nino-Murcia M, Berry GJ, Jeffrey RB Jr. Imaging for suspected appendicitis: negative appendectomy and perforation rates. *Radiology* 2002;**225**(1):131-6.

Bixby 2006 *{published data only}*

Bixby SD, Lucey BC, Soto JA, Theyson JM, Ozonoff A, Varghese JC. Perforated versus nonperforated acute appendicitis: accuracy of multidetector CT detection. *Radiology* 2006;**241**(3):780-6.

Brandt 2003 *{published data only}*

Brandt MM, Wahl WL. Liberal use of CT scanning helps to diagnose appendicitis in adults. *American Surgeon* 2003;**69**(9):727-31.

Brontvein 2002 *{published data only}*

Brontvein AT, Egorov VI, Vit'ko NK, Nazarenko VA, Bogdanov AE, Gasilov DV. Radiodiagnosis of acute appendicitis [Russian]. *Khirurgiia* 2002;**12**:13-6.

Caglayan 2010 *{published data only}*

Caglayan K, Gunerhan Y, Koc A, Uzun MA, Altinli E, Koksal N. The role of computerized tomography in the diagnosis of acute appendicitis in patients with negative ultrasonography findings and a low Alvarado score [Turkish]. *Ulusal Travma ve Acil Cerrahi Dergisi* 2010;**16**(5):445-8.

Callahan 2015 *{published data only}*

Callahan MJ, Anandalwar SP, MacDougall RD, Stamoulis C, Kleinman PL, Rangel SJ, et al. Pediatric CT dose reduction for suspected appendicitis: a practice quality improvement project using artificial gaussian noise - part 2, clinical outcomes. *AJR American Journal of Roentgenology* 2015;**204**:636-44.

Callahan MJ, Kleinman PL, Strauss KJ, Bandos A, Taylor GA, Tsai A, et al. Pediatric CT dose reduction for suspected appendicitis: a practice quality improvement project using artificial Gaussian noise - part 1, computer simulations. *AJR American Journal of Roentgenology* 2015;**204**:W86-94.

Castro 2001 *{published data only}*

Castro MA, Shipp TD, Castro EE, Ouzounian J, Rao P. The use of helical computed tomography in pregnancy for the diagnosis of acute appendicitis. *American Journal of Obstetrics and Gynecology* 2001;**184**(5):954-7.

Ceydeli 2006 *{published data only}*

Ceydeli A, Lavotshkin S, Yu J, Wise L. When should we order a CT scan and when should we rely on the results to diagnose an acute appendicitis?. Current Surgery 2006; Vol. 63, issue 6:464-8.

Chang 2013 *{published data only}*

Chang CC, Wong YC, Wang LJ, Wu CH, Chen HW, Huang CC, et al. Computed tomography diagnostic values of acute appendicitis in different patient subgroups. *Chinese Journal of Radiology (Taiwan)* 2013;**38**:9-14.

Chang 2016 *{published data only}*

Chang CC, Wong YC, Wu CH, Chen HW, Wang LJ, Lee YH, et al. Diagnostic performance on low dose computed tomography for acute appendicitis among attending and resident radiologists. *Iranian Journal of Radiology* 2016;**13**:e33222.

Chen 2010 *{published data only}*

Chen JD, Chiu YH, Tiu CM, Chou YH, Chang CY. Reappraisal of CT criteria for acute appendicitis: maximum outer appendiceal diameter. *Emergency Radiology* 2010;**17 (6)**:520-1.

Chen 2016 *{published data only}*

Chen KC, Arad A, Chen KC, Storrar J, Christy AG. The clinical value of pathology tests and imaging study in the diagnosis of acute appendicitis. *Postgraduate Medical Journal* 2016;**92**:611-9.

Chiu 2012 *{published data only}*

Chiu YH, Chen JD, Wang SH, Tiu CM, How CK, Lai JI, et al. Whether intravenous contrast is necessary for CT diagnosis of acute appendicitis in adult ED patients?. Academic Radiology 2013; Vol. 20, issue 1:73-8.

Cho 1999 *{published data only}*

Cho CS, Buckingham JM, Pierce M, Hardman DT. Computed tomography in the diagnosis of equivocal appendicitis. *Australian and New Zealand Journal of Surgery* 1999;**69**(9):664-7.

Choi 1998 *{published data only}*

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. *Clinical Imaging* 1998;**22**(4):252-71.

Cuschieri 2008 *{published data only}*

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. *Annals of Surgery* 2008;**248**(4):557-63.

Davis 2017 *{published data only}*

Davis J, Roh AT, Petterson MB, Kopelman TR, Matz SL, Gridley DG, et al. Computed tomography localization of the

appendix in the pediatric population relative to the lumbar spine. *Pediatric Radiology* 2017;**47**(3):301-5.

Dearing 2008 *{published data only}*

Dearing DD, Recabaren JA, Alexander M. Can computed tomography scan be performed effectively in the diagnosis of acute appendicitis without the added morbidity of rectal contrast?. *American Surgeon* 2008;**74**(10):917-20.

Debnath 2015 *{published data only}*

Debnath J, Kumar R, Mathur A, Sharma P, Kumar N, Shridhar N, et al. On the role of ultrasonography and CT scan in the diagnosis of acute appendicitis. *Indian Journal of Surgery* 2015;**77**(Suppl 2):221-6.

Deneuville 1995 *{published data only}*

Deneuville M, Blum A, Delfau F, Tortuyaux JM, Bresler L, Regent D. The value of CT in the diagnosis and treatment of acute appendicitis in adults and the elderly [French]. *Revue d'Imagerie Medicale* 1995;**7**(2):63-72.

Dibble 2016 *{published data only}*

Dibble E, Swenson D, Cartagena C, Herliczek T. CT versus MRI: effectiveness following sonography in staged imaging algorithm for suspected pediatric appendicitis. *Pediatric Radiology* 2016;**46**:S124.

Didier 2015 *{published data only}*

Didier RA, Vajtai PL, Hopkins KL. Iterative reconstruction technique with reduced volume CT dose index: diagnostic accuracy in pediatric acute appendicitis. *Pediatric Radiology* 2015;**45**:181-7.

Dillman 2016 *{published data only}*

Dillman JR, Gadepalli S, Sroufe NS, Davenport MS, Smith EA, Chong ST, et al. Equivocal pediatric appendicitis: unenhanced MR imaging protocol for nonsedated children - a clinical effectiveness study. *Radiology* 2016;**279**:216-25.

Elikashvili 2014 *{published data only}*

Elikashvili I, Tay ET, Tsung JW. The effect of point-of-care ultrasonography on emergency department length of stay and CT utilization in children with suspected appendicitis. *Academic Emergency Medicine* 2013;**1**:S7-8.

* Elikashvili I, Tay ET, Tsung JW. The effect of point-of-care ultrasonography on emergency department length of stay and computed tomography utilization in children with suspected appendicitis. *Academic Emergency Medicine* 2014;**21**:163-70.

Fefferman 2001 *{published data only}*

Fefferman NR, Roche KJ, Pinkney LP, Ambrosino MM, Genieser NB. Suspected appendicitis in children: focused CT technique for evaluation. *Radiology* 2001;**220**(3):691-5.

Fefferman 2005 *{published data only}*

Fefferman NR, Bomsztyk E, Yim AM, Rivera R, Amodio JB, Pinkney LP, et al. Appendicitis in children: low-dose CT with a phantom-based simulation technique - initial observations. *Radiology* 2005;**237**(2):641-6.

Foley 2005 *{published data only}*

Foley TA, Earnest IVF, 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**(1):89-96.

Fraser 2010 *{published data only}*

Fraser JD, Aguayo P, Sharp SW, Snyder CL, Rivard DC, Cully BE, et al. Accuracy of computed tomography in predicting appendiceal perforation. *Journal of Pediatric Surgery* 2010;**45**(1):231-5.

Fuchs 2002 *{published data only}*

Fuchs JR, Schlamberg JS, Shortsleeve MJ, Schuler JG. Impact of abdominal CT imaging on the management of appendicitis: an update. *Journal of Surgical Research* 2002;**106**(1):131-6.

Gaitini 2008 *{published data only}*

Gaitini D, Beck-Razi N, Mor-Yosef D, Fischer D, Itzhak OB, Krausz MM, et al. Diagnosing acute appendicitis in adults: accuracy of color Doppler sonography and MDCT compared with surgery and clinical follow-up. *AJR American Journal of Roentgenology* 2008;**190**(5):1300-6.

Garcia Pena 1999 *{published data only}*

Garcia Pena BM, Mandl KD, Kraus SJ, Fischer AC, Fleisher GR, Lund DP, et al. Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children. *JAMA* 1999;**282**(11):1041-6.

Garcia Pena BM, Taylor GA, Fishman SJ, Mandl KD. Costs and effectiveness of ultrasonography and limited computed tomography for diagnosing appendicitis in children. *Pediatrics* 2000;**106**:672-6.

Gaskill 2016 *{published data only}*

Gaskill CE, Simianu VV, Carnell J, Hippe DS, Bhargava P, Flum DR, et al. Use of computed tomography to determine perforation in patients with acute appendicitis. *Current Problems in Diagnostic Radiology* 2016;**07**:07.

Giuliano 2004 *{published data only}*

Giuliano V, Giuliano C, Pinto F, Scaglione M. Rapid CT scan visualization of the appendix and early acute non-perforated appendicitis using an improved oral contrast method. *Emergency Radiology* 2004;**10**(5):235-7.

Gwynn 2001 *{published data only}*

Gwynn LK. The diagnosis of acute appendicitis: clinical assessment versus computed tomography evaluation. *Journal of Emergency Medicine* 2001;**21**(2):119-23.

Hill 2010 *{published data only}*

Hill BC, Johnson SC, Owens EK, Gerber JL, Senagore AJ. CT scan for suspected acute abdominal process: impact of combinations of IV, oral, and rectal contrast. *World Journal of Surgery* 2010;**34**(4):699-703.

Hoecker 2005 *{published data only}*

Hoecker CC, Billman GF. The utility of unenhanced computed tomography in appendicitis in children. *Journal of Emergency Medicine* 2005;**28**(4):415-21.

Hookman 2000 *{published data only}*

Hookman P, Barkin JS. Diagnosis of acute right lower quadrant pain and appendicitis by CT scan: do we still need the clinician?. *American Journal of Gastroenterology* 2000;**95**(5):1355-7.

Horrow 2003 *{published data only}*

Horrow MM, White DS, Horrow JC. Differentiation of perforated from nonperforated appendicitis at CT. *Radiology* 2003;**227**(1):46-51.

Huynh 2007 *{published data only}*

Huynh V, Lalezarzadeh F, Lawandy S, Wong DT, Joe VC. Abdominal computed tomography in the evaluation of acute and perforated appendicitis in the community setting. *American Surgeon* 2007;**73**(10):1002-5.

Ives 2008 *{published data only}*

Ives EP, Sung S, McCue P, Durrani H, Halpern EJ. Independent predictors of acute appendicitis on CT with pathologic correlation. *Academic Radiology* 2008;**15**(8):996-1003.

Iwahashi 2005 *{published data only}*

Iwahashi N, Kitagawa Y, Mayumi T, Kohno H. Intravenous contrast-enhanced computed tomography in the diagnosis of acute appendicitis. *World Journal of Surgery* 2005;**29**(1):83-7.

Iwama 2002 *{published data only}*

Iwama Y, Hamanaka A, Matsumoto S, Kawasaki R, Fushimi Y, Imagumbai T, et al. Detectability of the appendix with multidetector-row CT scanning and multiplanar reconstruction [Japanese]. *Japanese Journal of Clinical Radiology* 2002;**47**(6):779-84.

Jabra 1997 *{published data only}*

Jabra AA, Shalaby-Rana EI, Fishman EK. CT of appendicitis in children. *Journal of Computer Assisted Tomography* 1997;**21**(4):661-6.

Johansson 2007 *{published data only}*

Johansson EP, Rydh A, Riklund KA. Ultrasound, computed tomography, and laboratory findings in the diagnosis of appendicitis. *Acta Radiologica* 2007;**48**(3):267-73.

Kahn 2013 *{published data only}*

Kahn SA, Sleem R, Schiralli M, Stassen NA, Sangosanya A, Cheng JD, et al. Assessment of surgical residents' ability to predict appendicitis with computed tomography. *American Surgeon* 2013;**79**:1308-10.

Kailidou 2006 *{published data only}*

Kailidou E, Pikoulis E, Katsiva V, Karavokyros I, Felekouras E, Pavlakis E, et al. Spiral computerized tomography: a diagnostic aid in patients with atypical acute appendicitis. *Acta Chirurgica Belgica* 2006;**106**(1):47-51.

Kaiser 2002 *{published data only}*

Kaiser S, Frenckner B, Jorulf HK. Suspected appendicitis in children: US and CT - a prospective randomized study. *Radiology* 2002;**223**(3):633-8.

Kaiser 2004 *{published data only}*

Kaiser S, Finnbogason T, Jorulf HK, Soderman E, Frenckner B. Suspected appendicitis in children: diagnosis with contrast-enhanced versus nonenhanced helical CT. *Radiology* 2004;**231**(2):427-33.

Kamel 2000 *{published data only}*

Kamel IR, Goldberg SN, Keogan MT, Rosen MP, Raptopoulos V. Right lower quadrant pain and suspected appendicitis: nonfocused appendiceal CT - review of 100 cases. *Radiology* 2000;**217**(1):159-63.

Karakas 2000 *{published data only}*

Karakas SP, Guelfguat M, Leonidas JC, Springer S, Singh SP. Acute appendicitis in children: comparison of clinical diagnosis with ultrasound and CT imaging. *Pediatric Radiology* 2000;**30**(2):94-8.

Kene 2016 *{published data only}*

Kene M, Ballard DW, Carpenter D. Is ultrasound non-inferior to computed tomography for pediatric appendicitis?. *Academic Emergency Medicine* 2016;**23**:S47.

Kharbanda 2007 *{published data only}*

Kharbanda AB, Taylor GA, Bachur RG. Suspected appendicitis in children: rectal and intravenous contrast-enhanced versus intravenous contrast-enhanced CT. *Radiology* 2007;**243**(2):520-6.

Kilincer 2017 *{published data only}*

Kilincer A, Akpinar E, Erbil B, Unal E, Karaosmanoglu AD, Kaynaroglu V, et al. A new technique for the diagnosis of acute appendicitis: abdominal CT with compression to the right lower quadrant. *European Radiology* 2017;**27**:3317–25.

Kim 2008a *{published data only}*

Kim K, Lee CC, Song K-J, Kim W, Suh G, Singer AJ. The impact of helical computed tomography on the negative appendectomy rate: a multi-center comparison. *Journal of Emergency Medicine* 2008;**34**(1):3-6.

Kim 2011 *{published data only}*

Kim SY, Lee KH, Kim K, Kim TY, Lee HS, Hwang SS, et al. Acute appendicitis in young adults: low- versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. Radiology 2011; Vol. 260, issue 2:437–45.

Kimura 2016 *{published data only}*

Kimura S, Bashar K, Burke PE. Appendicitis: have we improved in the management of acute appendicitis?. *Irish Journal of Medical Science* 2016;**1**:S112.

Latifi 2011 *{published data only}*

Latifi A, Labruto F, Kaiser S, Ullberg U, Sundin A, Torkzad MR. Does enteral contrast increase the accuracy of appendicitis diagnosis?. *Radiologic Technology* 2011;**82**(4):294-9.

Lazarus 2007 *{published data only}*

Lazarus E, Mayo-Smith WW, Mainiero MB, Spencer PK. CT in the evaluation of nontraumatic abdominal pain in pregnant women. *Radiology* 2007;**244**(3):784-90.

Lee 2001 *{published data only}*

Lee SL, Walsh AJ, Ho HS, Schwesinger WH, Grosfeld JL, Kuhn J, et al. Computed tomography and ultrasonography do not improve and may delay the diagnosis and treatment of acute appendicitis. *Archives of Surgery* 2001;**136**(5):556-62.

Lee 2006 *{published data only}*

Lee KH, Kim YH, Hahn S, Lee KW, Lee HJ, Kim TJ, et al. Added value of coronal reformations for duty radiologists and for referring physicians or surgeons in the CT diagnosis of acute appendicitis. *Korean Journal of Radiology* 2006;**7**(2):87-96.

Lee 2016 *{published data only}*

* Lee JM, Yi DY, Lim IS. Availability of low-dose CT in the diagnosis of appendicitis in childhood and comparison of abdominal USG and standard dose CT. *Journal of Pediatric Gastroenterology and Nutrition* 2016;**63**:S383-4.

Lim IS, Lee KH, Yi DY. Availability of low dose CT in the diagnosis of appendicitis in childhood and comparison of abdominal USG and standard dose CT. *Journal of Pediatric Gastroenterology and Nutrition* 2016;**62**:307.

Leeuwenburgh 2013 *{published data only}*

Leeuwenburgh MMN, Wiarda BM, Marinus JW, Vrouenraets BC, Gratama JWC, Spilt A, et al. Comparison of imaging strategies with conditional contrast enhanced CT and unenhanced MR imaging in patients suspected of having appendicitis: a multicenter diagnostic performance study. *Radiology* 2013;**268**(1):135-43.

Lin 2016 *{published data only}*

Lin WC, Lin CH. Multidetector computed tomography in the evaluation of pediatric acute abdominal pain in the emergency department. *BioMedicine* 2016;**6**:10.

Liu 2015 *{published data only}*

Liu W, Wei Qiang J, Xun Sun R. Comparison of multislice computed tomography and clinical scores for diagnosing acute appendicitis. *Journal of International Medical Research* 2015;**43**:341-9.

Lowe 2000 *{published data only}*

Lowe LH, Penney MW, Scheker LE, Perez R Jr, Stein SM, Heller RM, et al. Appendicolith revealed on CT in children with suspected appendicitis: how specific is it in the diagnosis of appendicitis?. *American Journal of Roentgenology* 2000;**175**(4):981-4.

Lowe 2001 *{published data only}*

Lowe LH, Draud KS, Hernanz-Schulman M, Newton MR, Heller RM, Stein SM, et al. Nonenhanced limited CT in children suspected of having appendicitis: prospective comparison of attending and resident interpretations. *Radiology* 2001;**221**(3):755-9.

Lowe 2001a *{published data only}*

Lowe LH, Penney MW, Stein SM, Heller RM, Neblett WW, Shyr Y, et al. Unenhanced limited CT of the abdomen in the diagnosis of appendicitis in children: comparison with sonography. *AJR American Journal of Roentgenology* 2001;**176**(1):31-5.

Lu 2007 *{published data only}*

Lu J, Dai H-X, Tan G-X, Zhang A-Y, Wang J-B, Xiao Z-M. Value of spiral computed tomography multiple planar reconstruction in the diagnosis and differential diagnosis of acute appendicitis with atypical clinical features [Chinese]. *World Chinese Journal of Digestology* 2007;**15**(10):1165-70.

McDonough 2002 *{published data only}*

McDonough J, Stasik C 3rd, Piontkowsky D 4th, Treisman E. Can appendiceal CT scanning be utilized effectively in widespread general surgical practice?. *American Surgeon* 2002;**68**(10):917-21.

Miki 2005 *{published data only}*

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. *Radiation Medicine - Medical Imaging and Radiation Oncology* 2005;**23**(4):242-55.

Mizuo 1999 *{published data only}*

Mizuo H, Ito Y, Iwasaki T, Nobuta S, Hosokawa Y, Sasaki Y. Enhanced and 5mm collimation helical CT of acute appendicitis [Japanese]. *Japanese Journal of Clinical Radiology* 1999;**44**(9):1007-12.

Morris 2002 *{published data only}*

Morris KT, Kavanagh M, Hansen P, Whiteford MH, Deveney K, Standage B. The rational use of computed tomography scans in the diagnosis of appendicitis. *American Journal of Surgery* 2002;**183**(5):547-50.

Moteki 2007 *{published data only}*

Moteki T, Horikoshi H. New CT criterion for acute appendicitis: maximum depth of intraluminal appendiceal fluid. *American Journal of Roentgenology* 2007;**188**(5):1313-9.

Moteki 2011 *{published data only}*

Moteki T, Ohya N, Horikoshi H. Evaluation of the maximum depth of intraluminal appendiceal fluid to diagnose appendicitis with a 64-detector row CT scanner. Journal of Computer Assisted Tomography 2011; Vol. 35, issue 6:703-10.

Mullins 2001 *{published data only}*

Mullins ME, Rhea JT, Greene MF, Novelline RA. Diagnostic imaging of suspected appendicitis in pregnant women: comparison of CT to ultrasonography. *Emergency Radiology* 2001;**8**(5):262-6.

Mullins 2001a *{published data only}*

Mullins ME, Kircher MF, Ryan DP, Doody D, Mullins TC, Rhea JT, et al. Evaluation of suspected appendicitis in children using limited helical CT and colonic contrast material. *AJR American Journal of Roentgenology* 2001;**176**(1):37-41.

Mun 2006 *{published data only}*

Mun S, Ernst RD, Chen K, Oto A, Shah S, Mileski WJ. Rapid CT diagnosis of acute appendicitis with IV contrast material. *Emergency Radiology* 2006;**12**(3):99-102.

Naeger 2011 *{published data only}*

Naeger DM, Chang SD, Kolli P, Shah V, Huang W, Thoeni RF. Neutral vs positive oral contrast in diagnosing acute appendicitis with contrast-enhanced CT: sensitivity, specificity, reader confidence and interpretation time. *British Journal of Radiology* 2011;**84**(1001):418-26.

Naffaa 2005 *{published data only}*

Naffaa LN, Ishak GE, Haddad MC. The value of contrast-enhanced helical CT scan with rectal contrast enema in the diagnosis of acute appendicitis. *Clinical Imaging* 2005;**29**(4):255-8.

Naoum 2002 *{published data only}*

Naoum JJ, Mileski WJ, Daller JA, Gomez GA, Gore DC, Kimbrough TD, et al. The use of abdominal computed tomography scan decreases the frequency of misdiagnosis in cases of suspected appendicitis. *American Journal of Surgery* 2002;**184**(6):587-90.

Ng 2007 *{published data only}*

Ng S-P, Cheng S-M, Yang F-S, Tzen C-Y, Huang J-K. Hyperdense appendix on unenhanced CT: a sign of acute appendicitis. *Abdominal Imaging* 2007;**32**(6):701-4.

Oliak 1999 *{published data only}*

Oliak D, Sinow R, French S, Udani VM, Stamos MJ. Computed tomography scanning for the diagnosis of perforated appendicitis. *American Surgeon* 1999;**65**(10):959-64.

Ozkan 2014 *{published data only}*

Ozkan S, Duman A, Durukan P, Yildirim A, Ozbakan O. The accuracy rate of Alvarado score, ultrasonography, and computerized tomography scan in the diagnosis of acute appendicitis in our center. *Nigerian Journal of Clinical Practice* 2014;**17**:413-8.

Park 2013 *{published data only}*

Park JS, Jeong JH, Lee JI, Lee JH, Park JK, Moon HJ. Accuracies of diagnostic methods for acute appendicitis. *American Surgeon* 2013;**79**:101-6.

Partrick 2003 *{published data only}*

Partrick DA, Janik JE, Janik JS, Bensard DD, Karrer FM. Increased CT scan utilization does not improve the diagnostic accuracy of appendicitis in children. *Journal of Pediatric Surgery* 2003;**38**(5):659-62.

Paulson 2005 *{published data only}*

Paulson EK, Harris JP, Jaffe TA, Haugan PA, Nelson RC. Acute appendicitis: added diagnostic value of coronal reformations from isotropic voxels at multi-detector row CT. *Radiology* 2005;**235**(3):879-85.

Peck 2000 *{published data only}*

Peck J, Peck A, Peck C, Peck J. The clinical role of noncontrast helical computed tomography in the diagnosis of acute appendicitis. *American Journal of Surgery* 2000;**180**(2):133-6.

Pena 1999 *{published data only}*

Pena BM, Taylor GA, Lund DP, Mandl KD. Effect of computed tomography on patient management and costs in children with suspected appendicitis. *Pediatrics* 1999;**104**:440-6.

Pena 2002 *{published data only}*

Pena BM, Taylor GA, Fishman SJ, Mandl KD. Effect of an imaging protocol on clinical outcomes among pediatric patients with appendicitis. *Pediatrics* 2002;**110**(6):1088-93.

Perez 2003 *{published data only}*

Perez J, Barone JE, Wilbanks TO, Jorgensson D, Corvo PR. Liberal use of computed tomography scanning does not improve diagnostic accuracy in appendicitis. *American Journal of Surgery* 2003;**185**(3):194-7.

Pickhardt 2011 *{published data only}*

Pickhardt PJ, Lawrence EM, Pooler BD, Bruce RJ. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. *Annals of Internal Medicine* 2011;**154**(12):789-96.

Poh 2004 *{published data only}*

Poh ACC, Lin M, Teh HS, Tan AGS. The role of computed tomography in clinically-suspected but equivocal acute appendicitis. *Singapore Medical Journal* 2004;**45**(8):379-84.

Poletti 2011 *{published data only}*

Poletti P-A, Platon A, De Perrot T, Sarasin F, Andereggen E, Rutschmann O, et al. Acute appendicitis: prospective evaluation of a diagnostic algorithm integrating ultrasound and low-dose CT to reduce the need of standard CT. *European Radiology* 2011;**21**(12):2558-66.

Pooler 2012 *{published data only}*

Pooler BD, Lawrence EM, Pickhardt PJ. MDCT for suspected appendicitis in the elderly: diagnostic performance and patient outcome. *Emergency Radiology* 2012;**19**(1):27-33.

Poortman 2009 *{published data only}*

Poortman P, Lohle PNM, Schoemaker CM, Cuesta MA, Oostvogel HJM, de Lange-de Klerk ESM, et al. Improving the false-negative rate of CT in acute appendicitis - reassessment of CT images by body imaging radiologists: a blinded prospective study. *European Journal of Radiology* 2010;**74**(1):67-70.

Poortman P, Oostvogel HJ, Bosma E, Lohle PN, Cuesta MA, de Lange-de Klerk ES, et al. Improving diagnosis of acute appendicitis: results of a diagnostic pathway with standard use of ultrasonography followed by selective use of CT. *Journal of the American College Surgeon* 2009;**208**(3):434-41.

Ramalingam 2016 *{published data only}*

Ramalingam V, Bates DD, Buch K, Uyeda J, Zhao KM, Storer LA, et al. Diagnosing acute appendicitis using a nonoral contrast

CT protocol in patients with a BMI of less than 25. *Emergency Radiology* 2016;**23**:455-62.

Raman 2002 *{published data only}*

Raman SS, Kadell BM, Vodopich DJ, Sayer J, Cryer H, Lu DS. Patient gender-related performance of nonfocused helical computed tomography in the diagnosis of acute appendicitis. *Journal of Computer Assisted Tomography* 2003;**27**(4):583-9.

* Raman SS, Lu DS, Kadell BM, Vodopich DJ, Sayre J, Cryer H. Accuracy of nonfocused helical CT for the diagnosis of acute appendicitis: a 5-year review. *AJR American Journal of Roentgenology* 2002;**178**(6):1319-25.

Ramarajan 2009 *{published data only}*

Ramarajan N, Krishnamoorthi R, Barth R, Ghanouni P, Mueller C, Dannenburg B, et al. An interdisciplinary initiative to reduce radiation exposure: evaluation of appendicitis in a pediatric emergency department with clinical assessment supported by a staged ultrasound and computed tomography pathway. *Academic Emergency Medicine* 2009;**16**(11):1258-65.

Rao 1997b *{published data only}*

Rao PM, Rhea JT, Novelline RA. Sensitivity and specificity of the individual CT signs of appendicitis: experience with 200 helical appendiceal CT examinations. *Journal of Computer Assisted Tomography* 1997;**21**(5):686-92.

Reeve 2010 *{published data only}*

Reeve T, Ferguson C. Imaging appendicitis: diagnostic accuracy of pre-operative imaging and prevalence of negative appendicectomy in an Australian tertiary referral hospital. *Journal of Medical Imaging and Radiation Oncology* 2010;**54**(Suppl 1):A127.

Reich 2011 *{published data only}*

Reich B, Zalut T, Weiner SG. An international evaluation of ultrasound vs. computed tomography in the diagnosis of appendicitis. *International Journal of Emergency Medicine* 2011;**4**(1):68-74.

Rhea 2005 *{published data only}*

Rhea JT, Halpern EF, Ptak T, Lawrason JN, Sacknoff R, Novelline RA. The status of appendiceal CT in an urban medical center 5 years after its introduction: experience with 753 patients. *AJR American Journal of Roentgenology* 2005;**184**(6):1802-8.

Rosengren 2004 *{published data only}*

Rosengren D, Brown AFT, Chu K. Radiological imaging to improve the emergency department diagnosis of acute appendicitis. *EMA - Emergency Medicine Australasia* 2004;**16**(5-6):410-6.

Sakai 2007 *{published data only}*

Sakai T, Yagihashi N, Osawa T, Harada O, Ito H. Evaluation of the usefulness of computed tomography on diagnosis and management of acute abdomen [Japanese]. *Japanese Journal of Gastroenterological Surgery* 2007;**40**(1):15-25.

Schuler 1998 *{published data only}*

Schuler JG, Shortsleeve MJ, Goldenson RS, Perez-Rossello JM, Perlmutter RA, Thorsen A. Is there a role for abdominal computed tomographic scans in appendicitis?. *Archives of Surgery* 1998;**133**(4):373-7.

Seo 2009 *{published data only}*

Seo H, Lee KH, Kim HJ, Kim K, Kang SB, 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 American Journal of Roentgenology* 2009;**193**(1):96-105.

Siddiqui 2007 *{published data only}*

Siddiqui AH, Afzal S. Perforated appendicitis: accuracy of CT diagnosis and correlation of CT findings with the length of hospital stay. *Journal of the College of Physicians and Surgeons - Pakistan* 2007;**17**(12):721-5.

Sivit 2000 *{published data only}*

* Sivit CJ, Applegate KE, Stallion A, Dudgeon DL, Salvator A, Schluchter M, et al. Imaging evaluation of suspected appendicitis in a pediatric population: effectiveness of sonography versus CT. *AJR American Journal of Roentgenology* 2000;**175**(4):977-80.

Sivit CJ, Dudgeon DL, Applegate KE, Borisa VJ, Berlin SC, Morrison SC, et al. Evaluation of suspected appendicitis in children and young adults: helical CT. *Radiology* 2000;**216**(2):430-3.

Sovtsov 2017 *{published data only}*

Sovtsov SA, Revel-Muroz ZA, Shkarednykh VY. [The use of computed tomography in the diagnosis of acute appendicitis]. *Khirurgiia* 2017;**4**:41-5.

Srinivasan 2010 *{published data only}*

* Srinivasan AS, Servaes S, Pena A, Del Pozo G, Lavelle J, Darge K. Utility of CT after sonography for suspected acute appendicitis. *Pediatric Radiology* 2010;**40**(4):582.

Stabile 2010 *{published data only}*

Stabile Ianora AA, Moschetta M, Lorusso V, Scardapane A. Atypical appendicitis: diagnostic value of volume-rendered reconstructions obtained with 16-slice multidetector-row CT. *La Radiologia Medica* 2010;**115**(1):93-104.

Stephen 2003 *{published data only}*

Stephen AE, Segev DL, Ryan DP, Mullins ME, Kim SH, Schnitzer JJ, et al. The diagnosis of acute appendicitis in a pediatric population: to CT or not to CT. *Journal of Pediatric Surgery* 2003;**38**(3):367-71.

Stromberg 2007 *{published data only}*

Stromberg C, Johansson G, Adolfsson A. Acute abdominal pain: diagnostic impact of immediate CT scanning. *World Journal of Surgery* 2007;**31**(12):2347-54.

Styrud 2000 *{published data only}*

Styrud J, Josephson T, Eriksson S. Reducing negative appendectomy: evaluation of ultrasonography and computer

tomography in acute appendicitis. *International Journal of Quality Health Care* 2000;**12**(1):65-8.

Suh 2011 *{published data only}*

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. *Surgical Laparoscopy, Endoscopy and Percutaneous Techniques* 2011;**21**(2):72-5.

Suthikeeree 2010 *{published data only}*

Suthikeeree W, Lertdomrongdej L, Charoensak A. Diagnostic performance of CT findings in differentiation of perforated from nonperforated appendicitis. *Journal of the Medical Association of Thailand* 2010;**93**(12):1422-9.

Tamburrini 2007 *{published data only}*

Tamburrini S, Brunetti A, Brown M, Sirlin C, Casola G. Acute appendicitis: diagnostic value of nonenhanced CT with selective use of contrast in routine clinical settings. *European Radiology* 2007;**17**(8):2055-61.

Tan 2013 *{published data only}*

Tan JW, Wan P, Kabir T, Ong HS. Use of the Alvarado score to stratify patients for CT evaluation in suspected appendicitis - our proposed algorithm. *Proceedings of Singapore Healthcare* 2012;**21**:S140.

Tan W, Pek WS, Kabir T, Goh YC, Chan WH, Wong WK, et al. The Alvarado score - a clinical stratification for computed tomography evaluation in suspected appendicitis. *Annals of the Academy of Medicine Singapore* 2012;**2**:S20.

Tan WJ, Pek W, Kabir T, Goh YC, Chan WH, Wong WK, et al. Alvarado score: a guide to computed tomography utilization in appendicitis. *ANZ Journal of Surgery* 2013;**83**:748-52.

Tatar 2016 *{published data only}*

Tatar IG, Yilmaz KB, Sahin A, Aydin H, Akinci M, Hekimoglu B. Evaluation of clinical Alvarado scoring system and CT criteria in the diagnosis of acute appendicitis. *Radiology Research and Practice* 2016:Epub.

Teo 2000 *{published data only}*

Teo EL, Tan KP, Lam SL, Ong CL, Wong CS. Ultrasonography and computed tomography in a clinical algorithm for the evaluation of suspected acute appendicitis in children. *Singapore Medical Journal* 2000;**41**(8):387-92.

Toorenvliet 2010 *{published data only}*

Toorenvliet BR, Wiersma F, Bakker RF, Merkus JW, Breslau PJ, Hamming JF. Routine ultrasound and limited computed tomography for the diagnosis of acute appendicitis. *World Journal of Surgery* 2010;**34**(10):2278-85.

Ujiki 2002 *{published data only}*

Ujiki MB, Murayama KM, Cribbins AJ, Angelos P, Dawes L, Prystowsky JB, et al. CT scan in the management of acute appendicitis. *Journal of Surgical Research* 2002;**105**(2):119-22.

Unlu 2009 *{published data only}*

Unlu C, de Castro SMM, Tuynman JB, Wust AF, Steller EPh, van Wagensveld BA. Evaluating routine diagnostic imaging in acute appendicitis. *International Journal of Surgery* 2009;**7**:451-5.

Uyeda 2015 *{published data only}*

Uyeda JW, Yu H, Ramalingam V, Devalapalli AP, Soto JA, Anderson SW. Evaluation of acute abdominal pain in the emergency setting using computed tomography without oral contrast in patients with body mass index greater than 25. *Journal of Computer Assisted Tomography* 2015;**39**:681-6.

Vajtai 2013 *{published data only}*

Vajtai P, Hopkins K, Didier R. Accuracy in diagnosis of pediatric acute appendicitis: weight-based CT protocols with traditional filtered back projection versus 60% CTDIvol-reduced protocols with iterative reconstruction technique. *Pediatric Radiology* 2013;**43**:S555.

Van Randen 2011 *{published data only}*

Van Randen A, Lameris W, Van Es HW, Ten Hove W, Bouma WH, Van Leeuwen MS, et al. Profiles of US and CT imaging features with a high probability of appendicitis. *European Radiology* 2010;**20**(7):1657-66.

* Van Randen A, Lameris W, Van Es HW, Van Heesewijk HPM, Van Ramshorst B, Ten Hove W, et al. A comparison of the accuracy of ultrasound and computed tomography in common diagnoses causing acute abdominal pain. *European Radiology* 2011;**21**(7):1535-45.

Wadhwani 2016 *{published data only}*

Wadhwani A, Guo L, Saude E, Els H, Lang E, McRae A, et al. Intravenous and oral contrast vs intravenous contrast alone computed tomography for the visualization of appendix and diagnosis of appendicitis in adult emergency department patients. *Canadian Association of Radiologists Journal* 2016;**67**:234-41.

Wallace 2008 *{published data only}*

Wallace CA, Petrov MS, Soybel DI, Ferzoco SJ, Ashley SW, Tavakkolizadeh A. Influence of imaging on the negative appendectomy rate in pregnancy. *Journal of Gastrointestinal Surgery* 2008;**12**(1):46-50.

Westerland 2016 *{published data only}*

Westerland OA, Singh RK, Nagraj HC, Muthukumarasamy S, Ramesar K, Saunders MP, et al. Imaging in acute appendicitis: lessons learned from appendicectomy patients at two district general hospitals in the United Kingdom over a six-year period. *Hong Kong Journal of Radiology* 2016;**19**:118-25.

Weyant 2000 *{published data only}*

Weyant MJ, Eachempati SR, Maluccio MA, Rivadeneira DE, Grobmyer SR, Hydo LJ, et al. Interpretation of computed tomography does not correlate with laboratory or pathologic findings in surgically confirmed acute appendicitis. *Surgery* 2000;**128**(2):145-52.

Weyant 2001 *{published data only}*

Weyant MJ, Eachempati SR, Maluccio MA, Spigland N, Hydo LJ, Barie PS. The use of computed tomography for the diagnosis of acute appendicitis in children does not influence the overall rate of negative appendectomy or perforation. *Surgical Infections* 2001;**2**(1):19-23.

Yeoh 2016 *{published data only}*

Yeoh T. Evaluation of computed tomography in diagnosis of acute appendicitis. *Scottish Medical Journal* 2016;**61 (4)**:NP50.

Yi 2017 *{published data only}*

Yi DY, Lee KH, Park SB, Kim JT, Lee NM, Kim H, et al. Accuracy of low dose CT in the diagnosis of appendicitis in childhood and comparison with USG and standard dose CT. *Jornal de Pediatria* 2017;**23**:23.

Yoo 2009 *{published data only}*

Yoo JK, Kim J-E, Hyung SK, Hee YH. MDCT with coronal reconstruction: clinical benefit in evaluation of suspected acute appendicitis in pediatric patients. *AJR American Journal of Roentgenology* 2009;**192**(1):150-2.

Yun 2016 *{published data only}*

Yun SJ, Kim HC, Yang DM, Kim SW, Rhee SJ, Shin JS, et al. Diagnostic usefulness of low-dose nonenhanced computed tomography with coronal reformations in patients with suspected acute appendicitis: a comparison with standard-dose computed tomography. *Journal of Computer Assisted Tomography* 2016;**40**:485-92.

Zourob 2016 *{published data only}*

Zourob M, Board J, Baithun S, Cunningham FO, Saeed MF, Corbally MT, et al. A clinicopathological study of emergency appendectomy. *Bahrain Medical Bulletin* 2016;**38**:219-22.

References to studies awaiting assessment

D'lppolito 1998 *{published data only}*

D'lppolito G, de Mello GG, Szejnfeld J. The value of unenhanced CT in the diagnosis of acute appendicitis. *Sao Paulo Medical Journal* 1998;**116**(6):1838-45.

Ege 2002 *{published data only}*

Ege G, Akman H, Sahin A, Bugra D, Kuzucu K. Diagnostic value of unenhanced helical CT in adult patients with suspected acute appendicitis. *British Journal of Radiology* 2002;**75**(897):721-5.

Elghany 2011 *{published data only}*

Elghany EA, Ali GG. Multidetector row helical CT and US in diagnosing appendicitis. *Egyptian Journal of Radiology and Nuclear Medicine* 2011;**42**(2):139-45.

Lane 1997 *{published data only}*

Lane MJ, Katz DS, Ross BA, Clautice-Engle TL, Mindelzun RE, Jeffrey RB Jr. Unenhanced helical CT for suspected acute appendicitis. *AJR American Journal of Roentgenology* 1997;**168**(2):405-9.

Lietzen 2018 *{published data only}*

Lietzen E, Salminen P, Rinta-Kiikka I, Paajanen H, Rautio T, Nordstrom P, et al. The accuracy of the computed tomography diagnosis of acute appendicitis: does the experience of the radiologist matter?. *Scandinavian Journal of Surgery* 2018;**107**:43-7.

LOCAT Group 2018 *{published data only}*

Locat Group. Low-dose CT for the diagnosis of appendicitis in adolescents and young adults (LOCAT): a pragmatic, multicentre, randomised controlled non-inferiority trial. *The Lancet Gastroenterology & Hepatology* 2017;**2**:793-804.

Park 2018 *{published data only}*

Park JH, Jeon J-J, Lee SS, Dhanantwari AC, Sim JY, Kim Hae Y, et al. Can we perform CT of the appendix with less than 1 mSv? A de-escalating dose-simulation study. *European Radiology* 2018;**28**:1826-34.

Rao 1996 *{published data only}*

Rao PM, Rhea JT, Novelline RA, Mccabe CJ, Lawrason JN, Sacknoff R. A new helical computed tomographic technique for appendiceal imaging: preliminary experience with the focused appendix computed tomographic examination. *Emergency Radiology* 1996;**3**(5):241-6.

Singh 2018 *{published data only}*

Singh NR, Luwang NT, Priyabarta Y, Singh CG, Singh WJ. Low-dose noncontrast computed tomography in adults with acute appendicitis. *Journal of the Medical Society* 2018;**32**:111-7.

Sippola 2018 *{published data only}*

Sippola S, Virtanen J, Tammilehto V, Gronroos J, Hurme S, Niiniviita H, et al. The accuracy of low-dose computed tomography protocol in patients with suspected acute appendicitis: the OPTICAP study. *Annals of Surgery* 2018;**published online ahead of print**.

Stroman 1999 *{published data only}*

Stroman DL, Bayouth CV, Kuhn JA, Westmoreland M, Jones RC, Fisher TL, et al. The role of computed tomography in the diagnosis of acute appendicitis. *American Journal of Surgery* 1999;**178**(6):485-9.

Yang 2016 *{published data only}*

Yang F, Zhou J, Zhao C. Feasibility of low-dose CT based on abdominal cross-sectional area in diagnosis of acute appendicitis [Chinese]. *Chinese Journal of Interventional Imaging and Therapy* 2016;**13**:302-5.

Yetkin 2002 *{published data only}*

Yetkin G, Basak M, Isgor A, Kebudi A, Akgun I. Can negative appendectomy rate be decreased by using spiral computed tomography without contrast material?. *Acta Chirurgica Belgica* 2002;**102**(5):334-7.

Yildirim 2008 *{published data only}*

Yildirim E, Karagulle E, Kirbas I, Turk E, Hasdogan B, Teksam M, et al. Alvarado scores and pain onset in relation to multislice CT findings in acute appendicitis. *Diagnostic and Interventional Radiology* 2008;**14**(1):14-8.

Additional references

Al-Khayal 2007

Al-Khayal KA, Al-Omran MA. Computed tomography and ultrasonography in the diagnosis of equivocal acute appendicitis. A meta-analysis. *Saudi Medical Journal* 2007;**28**:173-80.

Alvarado 1986

Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Annals of Emergency Medicine* 1986;**15**:557-64.

Aly 2016

Aly NE, McAteer D, Aly EH. Low vs. standard dose computed tomography in suspected acute appendicitis: is it time for a change?. *International Journal of Surgery* 2016;**31**:71-9.

Anderson 2005

Anderson BA, Salem L, Flum DR. A systematic review of whether oral contrast is necessary for the computed tomography diagnosis of appendicitis in adults. *American Journal of Surgery* 2005;**190**:474-8.

Anderson 2012

Anderson JE, Bickler SW, Chang DC, Talamini MA. Examining a common disease with unknown etiology: trends in epidemiology and surgical management of appendicitis in California, 1995–2009. *World Journal of Surgery* 2012;**36**(12):2787-94.

Andersson 2004

Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *British Journal of Surgery* 2004;**91**:28-37.

Andersson 2007

Andersson RE, Petzold MG. Nonsurgical treatment of appendiceal abscess or phlegmon: a systematic review and meta-analysis. *Annals of Surgery* 2007;**246**:741-8.

Andersson 2008

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**:1843-9.

Andersson 2017

Andersson M, Kolodziej B, Andersson RE, Group Strappscore Study. Randomized clinical trial of Appendicitis Inflammatory Response score-based management of patients with suspected appendicitis. *British Journal of Surgery* 2017;**104**:1451-61.

Balthazar 1986

Balthazar EJ, Megibow AJ, Hulnick D, Gordon RB, Naidich DP, Beranbaum ER. CT of appendicitis. *AJR American Journal of Roentgenology* 1986;**147**:705-10.

Basaran 2009

Basaran A, Basaran M. Diagnosis of acute appendicitis during pregnancy: a systematic review. *Obstetrical & Gynecological Survey* 2009;**64**:481-8.

Bhangu 2015

Bhangu A, Soreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet* 2015;**386**:1278-87.

Bijnen 2003

Bijnen CL, van den Broek WT, Bijnen AB, de Ruiter P, Gouma DJ. Implications of removing a normal appendix. *Digestive Surgery* 2003;**20**:215-9.

Birnbaum 2000

Birnbaum BA, Wilson SR. Appendicitis at the millennium. *Radiology* 2000;**215**:337-48.

Blomqvist 2001

Blomqvist PG, Andersson RE, Granath F, Lambe MP, Ekbom AR. Mortality after appendectomy in Sweden 1987-1996. *Annals of Surgery* 2001;**233**(4):455-60.

Bossuyt 2015

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;**351**:h5527.

Bower 2014

Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. *Trials* 2014;**15**:399.

Brenner 2007

Brenner DJ, Hall EJ. Computed tomography - an increasing source of radiation exposure. *New England Journal of Medicine* 2007;**357**:2277-84.

Brown 2008

Brown MA. Imaging acute appendicitis. *Seminars in Ultrasound, CT and MRI* 2008;**29**:293-307.

Brueton 2013

Brueton VC, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. *Cochrane Database of Systematic Reviews* 2013, Issue 12. [doi: 10.1002/14651858.MR000032.pub2]

Chong 2010

Chong CF, Adi MI, Thien A, Suyoi A, Mackie AJ, Tin AS, et al. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. *Singapore Medical Journal* 2010;**51**:220-5.

Coleman 2018

Coleman JJ, Carr BW, Rogers T, Field MS, Zarzaur BL, Savage SA, et al. The Alvarado score should be used to reduce emergency department length of stay and radiation exposure in select

patients with abdominal pain. *Journal of Trauma and Acute Care Surgery* 2018;**84**:946-50.

Coursey 2010

Coursey CA, Nelson RC, Patel MB, Cochran C, Dodd LG, Delong DM, et al. Making the diagnosis of acute appendicitis: do more preoperative CT scans mean fewer negative appendectomies? A 10-year study. *Radiology* 2010;**254**:460-8.

D'Souza 2015

D'Souza N, D'Souza C, Grant D, Royston E, Farouk M. The value of ultrasonography in the diagnosis of appendicitis. *International Journal of Surgery* 2015;**13**:165-9.

Dahabreh 2015

Dahabreh IJ, Adam GP, Halladay CW, Steele DW, Daiello LA, Wieland LS, et al. [Diagnosis of Right Lower Quadrant Pain and Suspected Acute Appendicitis. Comparative Effectiveness Review No. 157 (prepared by the Brown Evidence-based Practice Center under Contract No. 290-2012-00012-I)]. AHRQ Publication No. 15(16)-EHC025-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2015, issue www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Di Saverio 2016

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 Journal of Emergency Surgery* 2016;**11**:34.

Doria 2006

Doria AS, Moineddin R, Kellenberger CJ, Epelman M, Beyene J, Schuh S, et al. A meta-analysis. *Radiology* 2006;**241**:83-94.

Drake 2014

Drake FT, Alfonso R, Bhargava P, Cuevas C, Dighe MK, Florence MG, et al. Enteral contrast in the computed tomography diagnosis of appendicitis: comparative effectiveness in a prospective surgical cohort. *Annals of Surgery* 2014;**260**:311-6.

Duke 2016

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 American Journal of Roentgenology* 2016;**206**:508-17.

Ebell 2014

Ebell MH, Shinholser J. What are the most clinically useful cutoffs for the Alvarado and Pediatric Appendicitis Scores? A systematic review. *Annals of Emergency Medicine* 2014;**64**:365-72 e2.

Frush 2009

Frush DP, Frush KS, Oldham KT. A North American perspective. *Pediatric Radiology* 2009;**39**:500-5.

Grunewald 1993

Grunewald B, Keating J. Should the 'normal' appendix be removed at operation for appendicitis?. *Journal of the Royal College of Surgeons of Edinburgh* 1993;**38**:158-60.

Hall 2010

Hall MJ, DeFrances CJ, Williams SN, Golosinskiy A, Schwartzma A. National Hospital Discharge Survey: 2007 Summary. National Health Statistics Reports 2010, issue 27.

Hernanz-Schulman 2010

Hernanz-Schulman M. CT and US in the diagnosis of appendicitis: an argument for CT. *Radiology* 2010;**255**:3-7.

Hlibczuk 2010

Hlibczuk V, Dattaro JA, Jin Z, Falzon L, Brown MD. Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review. *Annals of Emergency Medicine* 2010;**55**:51-9.

Hospital Episode Statistics 2015

Hospital Episode Statistics. Admitted Patient Care–England 2014–2015. Total Procedures and Interventions. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/hospital-episode-statistics-admitted-patient-care-england-2014-15. Accessed 2 June 2018.

Humes 2006

Humes DJ, Simpson J. Acute appendicitis. *British Medical Journal* 2006;**333**:530-4.

Jaunoo 2012

Jaunoo SS, Hale AL, Masters JP, Jaunoo SR. An international survey of opinion regarding investigation of possible appendicitis and laparoscopic management of a macroscopically normal appendix. *Annals of the Royal College of Surgeons England* 2012;**94**:476-80.

Krajewski 2011

Krajewski S, Brown J, Phang PT, Raval M, Brown CJ. Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. *Canadian Journal of Surgery* 2011;**54**:43-53.

Lameris 2009

Lameris W, van Randen A, Go PM, Bouma WH, Donkervoort SC, Bossuyt PM, et al. Single and combined diagnostic value of clinical features and laboratory tests in acute appendicitis. *Academic Emergency Medicine* 2009;**16**:835-42.

Leeuwenburgh 2012

Leeuwenburgh MM, Wiarda BM, Bipat S, Nio CY, Bollen TL, Kardux JJ, et al. Acute appendicitis on abdominal MR images: training readers to improve diagnostic accuracy. *Radiology* 2012;**264**:455-63.

Lewis 1975

Lewis FR, Holcroft JW, Boey J, Dunphy E. Appendicitis. A critical review of diagnosis and treatment in 1000 cases. *Archives of Surgery* 1975;**110**:677-84.

McKay 2007

McKay R, Shepherd J. The use of the clinical scoring system by Alvarado in the decision to perform computed tomography for acute appendicitis in the ED. *American Journal of Emergency Medicine* 2007;**25**:489-93.

National Surgical Research Collaborative 2013

National Surgical Research Collaborative. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. *British Journal of Surgery* 2013;**100**:1240-52.

Neville 2009

Neville AM, Paulson EK. MDCT of appendicitis: value of coronal reformations. *Abdominal Imaging* 2009;**34**:42-8.

Ohle 2011

Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Medicine* 2011;**9**:139.

Paulson 2003

Paulson EK, Kalady MF, Pappas TN. Clinical practice. Suspected appendicitis. *New England Journal of Medicine* 2003;**348**(3):236-42.

Raja 2010

Raja AS, Wright C, Sodickson AD, Zane RD, Schiff GD, Hanson R, et al. Negative appendectomy rate in the era of CT: an 18-year perspective. *Radiology* 2010;**256**:460-5.

Raman 2008

Raman SS, Osuagwu FC, Kadell B, Cryer H, Sayre J, Lu DS. Effect of CT on false positive diagnosis of appendicitis and perforation. *New England Journal of Medicine* 2008;**358**:972-3.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**:982-90.

Repplinger 2018

Repplinger MD, Pickhardt PJ, Robbins JB, Kitchin DR, Ziemlewicz TJ, Hetzel SJ, et al. Prospective comparison of the diagnostic accuracy of MR imaging versus CT for acute appendicitis. *Radiology* 2018;**288**(2):467–75.

Rybkin 2007

Rybkin AV, Thoeni RF. Current concepts in imaging of appendicitis. *Radiologic Clinics of North America* 2007;**45**:411-22.

Sadr 2009

Sadr Azodi O, Andren-Sandberg A, Larsson H. Genetic and environmental influences on the risk of acute appendicitis in twins. *British Journal of Surgery* 2009;**96**:1336-40.

Salminen 2015

Salminen P, Paajanen H, Rautio T, Nordstrom 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**:2340-8.

Sammalkorpi 2014

Sammalkorpi HE, Mentula P, Leppaniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis - a prospective study. *BMC Gastroenterology* 2014;**14**:114.

Smith-Bindman 2009

Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Archieves of Internal Medicine* 2009;**169**:2078-86.

Strong 2015

Strong S, Blencowe N, Bhangu A, National Surgical Research Collaborative. How good are surgeons at identifying appendicitis? Results from a multi-centre cohort study. *International Journal of Surgery* 2015;**15**:107-12.

Strouse 2010

Strouse PJ. Pediatric appendicitis: an argument for US. *Radiology* 2010;**255**:8-13.

Tan 2017

Tan VF, Patlas MN, Katz DS. Imaging trends in suspected appendicitis - a Canadian perspective. *Emergency Radiology* 2017;**24**:249-54.

Teh 2000

Teh SH, O'Ceallaigh S, Mckeon JG, O'Donohoe MK, Tanner WA, Keane FB. Should an appendix that looks 'normal' be removed at diagnostic laparoscopy for acute right iliac fossa pain?. *European Journal of Surgery* 2000;**166**:388-9.

Terasawa 2004

Terasawa T, Blackmore CC, Bent S, Kohlwes RJ. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Annals of Internal Medicine* 2004;**141**:537-46.

The Locat Group 2017

The Locat Group. Low-dose CT for the diagnosis of appendicitis in adolescents and young adults (LOCAT): a pragmatic, multicentre, randomised controlled non-inferiority trial. *Lancet Gastroenterology & Hepatology* 2017;**2**:793-804.

van den Broek 2001

van den Broek WT, Bijnen AB, de Ruiter P, Gouma DJ. A normal appendix found during diagnostic laparoscopy should not be removed. *British Journal of Surgery* 2001;**88**:251-4.

van Randen 2008

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.

van Rossem 2016

van Rossem CC, Bolmers MD, Schreinemacher MH, van Geloven AA, Bemelman WA, Snapshot Appendicitis Collaborative Study Group. Prospective nationwide outcome audit of surgery for suspected acute appendicitis. *British Journal of Surgery* 2016;**103**:144-51.

Velanovich 1992

Velanovich V, Satava R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. *American Surgeon* 1992;**58**:264-9.

Vons 2011

Vons C, Barry C, Maitre S, Pautrat K, Leconte M, Costaglioli B, et al. Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: an open-label, non-inferiority, randomised controlled trial. *Lancet* 2011;**377**:1573-9.

Wagner 2008

Wagner PL, Eachempati SR, Soe K, Pieracci FM, Shou J, Barie PS. Defining the current negative appendectomy rate: for whom is preoperative computed tomography making an impact?. *Surgery* 2008;**144**:276-82.

Wagner 2009

Wagner J, Shojania K. Update: appendicitis, adults. In: Simel DL, Rennie D editor(s). The Rational Clinical Examination: Evidence-Based Clinical Diagnosis. First Edition. New York: McGraw-Hill, 2009:61-3.

Weiss 2014

Weiss AJ, Elixhauser A, Andrews RM. Characteristics of Operating Room Procedures in U.S. Hospitals, 2011: HCUP Statistical Brief #170. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville, MD: Agency for Healthcare Research and Quality, February 2014.

Weston 2005

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**:368-79.

Whiting 2013

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Group Quadas-Steering. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *Journal of Clinical Epidemiology* 2013;**66**:1093-104.

Woolard 2004

Woolard RH, Carty K, Wirtz P, Longabaugh R, Nirenberg TD, Minugh PA, et al. Research fundamentals: follow-up of subjects in clinical trials: addressing subject attrition. *Academic Emergency Medicine* 2004;**11**:859-66.

Xiong 2015

Xiong B, Zhong B, Li Z, Zhou F, Hu R, Feng Z, et al. Diagnostic accuracy of noncontrast CT in detecting acute appendicitis: a meta-analysis of prospective studies. *American Surgeon* 2015;**81**:626-9.

Yun 2017

Yun SJ, Ryu CW, Choi NY, Kim HC, Oh JM, Yang DM. Comparison of low- and standard-dose CT for the diagnosis of acute appendicitis: a meta-analysis. *AJR American Journal of Roentgenology* 2017;**208**:W198-207.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Antevil 2006

Study characteristics	
Patient sampling	All patients were evaluated for appendicitis according to a multi-disciplinary diagnostic pathway. Recruitment period: August 2004 to August 2005
Patient characteristics and setting	Age range (mean): 15 to 90 years (32); 60% women. Pregnant women and patients younger than 15 years were excluded Emergency department in San Diego, California, USA. Single-centre study Disease spectrum: women were included regardless of appendicitis risk. Men with intermediate risk of appendicitis were included. Men with characteristic symptoms and signs had surgery without preceding CT Disease spectrum: women with any degree of suspicion and men with intermediate suspicion of appendicitis were included. Men with charac-

Antevil 2006 *(Continued)*

	teristic symptoms and signs (high suspicion) had surgery without preceding CT
Index tests	Helical CT of the entire abdomen with oral and intravenous contrast enhancement. Number of slices, model and manufacturer of CT device: not reported. Slice thickness: 5 mm. Slice interval, voltage, mAs product: not reported
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy. Patients who did not have appendectomy were assumed to not have appendicitis
Flow and timing	609 patients were included, and 529 had abdominal CT. Of these, 95 had appendicitis confirmed by histological evaluation of removed appendices. The number of patients having surgery is not reported
Comparative	
Criteria for CT diagnosis of appendicitis	Not stated
Assessors of the CT-scan	An attending body imaging radiologist
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		High	**High**
DOMAIN 3: Reference Standard			

Antevil 2006 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Ashraf 2006

Study characteristics	
Patient sampling	Patients with clinically equivocal symptoms and signs of appendicitis were referred for focused appendiceal CT during a 1-year period. Inclusion criteria for clinically equivocal appendicitis were based on the clinical judgement of the referring surgeon or emergency care physician. All patients were included regardless of age Recruitment period: not stated
Patient characteristics and setting	Age range (mean): 9 to 67 years (24.5). The proportion of patients younger than 15 years of age and the gender distribution are not reported Radiology Department at a university hospital, Karachi, Pakistan. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Unenhanced, helical CT of the lower abdomen (HiSpeed Advantage, General Electric Medical Systems). Slice thickness and interval: 5 mm. Voltage: 120 to 140 kV. mAs product: 220 to 250 mAs
Target condition and reference standard(s)	Appendicitis. Surgical reports and histopathological reports were prepared for patients who had surgery with or without appendectomy. All patients who did not undergo surgery were followed up for 2 months. Follow-up is not described in further detail
Flow and timing	63 patients were included; 23 had appendicitis. The number of patients who had surgery is not stated, and no account is provided of the completeness and

Ashraf 2006 *(Continued)*

	the outcome of follow-up in patients who did not have surgery. Two patients were excluded from the analysis of accuracy, which included 61 patients. Reasons for exclusion are not reported		
Comparative			
Criteria for CT diagnosis of appendicitis	An appendix > 6 mm in transverse diameter was considered abnormal. Additional secondary criteria were periappendiceal inflammatory changes. Presence of an appendicolith in the absence of other primary and secondary criteria was not interpreted as acute appendicitis		
Assessors of the CT-scan	One faculty radiologist		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			

Ashraf 2006 *(Continued)*

Did all patients receive a reference standard?	Yes
Did all patients receive the same reference standard?	No
Did all patients with a positive CT-scan have surgery?	Yes
Did all patients with a negative CT-scan have clinical follow-up?	No
Was the choice of reference standard independent of the result of the index test?	No
Were all patients included in the analyses?	Yes
	High

Atema 2015

Study characteristics	
Patient sampling	Adult patients (≥ 18 years) with suspected appendicitis based on medical history, physical examination findings, and laboratory test results were included. Pregnant women were excluded. Patients were not included during nighttime. Recruitment period: March 2005 to November 2006
Patient characteristics and setting	Age range (mean): 19 to 89 (40). 54% women Clinical setting: 6 Emergency Departments in The Netherlands. Multi-centre study Disease spectrum: any suspicion of appendicitis
Index tests	4- to 16-slice helical CT of the entire abdomen (Somatom Plus, Somatom Sensation 16; Siemens Medical Systems and Tomoscan AV; Philips Medical Systems). Enhancement with intravenous contrast material. Slice thickness: 1.5 to 6.5 mm. Slice interval: not stated. Voltage: 120 to 140 kV. mAs product: 165 to 200 mAs
Target condition and reference standard(s)	Appendicitis. An expert panel assigned a final diagnosis based on all available information after at least 6 months of follow-up. A final diagnosis of appendicitis was based predominantly on surgical findings, histopathology, and follow-up data. The expert panel consisted of 2 experienced gastrointestinal surgeons and an experienced abdominal radiologist
Flow and timing	422 patients were included; all had CT and 251 had appendicitis. The number who had surgery and the number intended for clinical follow-up are not reported. All patients had a final diagnosis assigned by the expert panel
Comparative	
Criteria for CT diagnosis of appendicitis	No criteria were prespecified for the CT diagnosis of appendicitis. The diagnosis was left to the discretion of the reader who evaluated the images
Assessors of the CT-scan	CT-scans were evaluated by staff radiologists or by residents supervised by staff radiologists. CT done after office hours was reevaluated the next day

Atema 2015 *(Continued)*

Notes	Atema 2015 is a substudy of the OPTIMA study, which compared 11 imaging strategies for detecting urgent conditions in patients with acute abdominal pain. The OPTIMA study included 1101 adult patients (≥ 18 years) with non-traumatic abdominal pain for longer than 2 hours and less than 5 days. All patients had plain radiographs (upright chest and supine abdominal), abdominal ultrasonography, and abdominal CT in a paired design. An expert panel assigned a final diagnosis based on all available information after 6 months. Analyses in Atema 2015 are restricted to patients with suspected appendicitis based on clinical evaluation and laboratory tests

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			

Atema 2015 *(Continued)*

Did all patients receive a reference standard?	Yes
Did all patients receive the same reference standard?	No
Did all patients with a positive CT-scan have surgery?	Unclear
Did all patients with a negative CT-scan have clinical follow-up?	Unclear
Was the choice of reference standard independent of the result of the index test?	No
Were all patients included in the analyses?	Yes
	High

Balthazar 1991

Study characteristics	
Patient sampling	Consecutive patients referred for abdominal CT due to atypical symptoms and signs of appendicitis (lower abdominal pain and tenderness, but no nausea, vomiting, low-grade fever, or leukocytosis). No exclusion criteria reported
Patient characteristics and setting	Age range (mean): 9 to 87 years (42). The proportion of patients younger than 15 years is not reported. 48% women Department of Radiology, Philadelphia, Pennsylvania, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Non-helical CT of the entire abdomen with oral and intravenous contrast enhancement (9800, General Electric Medical Systems). Slice thickness and slice interval: 5 to 8 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy. Discharge diagnosis and follow-up were provided for patients who did not have surgery. It is not stated how follow-up was performed
Flow and timing	100 patients were included. Surgery with appendectomy was performed in 74 patients; 64 had appendicitis confirmed histologically. No account is reported on completeness and outcomes of follow-up in the 36 patients who did not have surgery
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 3 mm, periappendiceal inflammatory changes, appendix wall hyperenhancement, thickened appendix wall, abscess or phlegmon in the right iliac fossa, appendicolith

Balthazar 1991 *(Continued)*

Assessors of the CT-scan	Not stated
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		

Balthazar 1991 *(Continued)*

Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Balthazar 1994

Study characteristics			
Patient sampling	Consecutive patients were suspected of having appendicitis. Selected group of patients had suggestive but not typical clinical and laboratory findings of appendicitis. No exclusion criteria were reported Recruitment period: not stated		
Patient characteristics and setting	Age range (mean): 15 to 82 (38) years. 49% women Department of Radiology, New York, New York, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis		
Index tests	Non-helical CT of the entire abdomen with oral and intravenous contrast enhancement (9800 HiLight, General Electric Medical Systems). Slice thickness and slice interval: 5 to 8 mm. Voltage and mAs product: not stated		
Target condition and reference standard(s)	Appendicitis. Histopathological examination was performed in patients who had surgery with appendectomy. The reference test is not reported for patients who did not have surgery		
Flow and timing	100 patients were included; 69 had surgery and 54 had appendicitis confirmed histologically. There is no report of follow-up in patients who did not have surgery. Calculation of sensitivity and specificity was based on a 2×2 table that comprised all included patients		
Comparative			
Criteria for CT diagnosis of appendicitis	Abnormal appendix (a specific diameter criterion is not reported) or presence of pericaecal inflammation and/or abscess associated with an appendicolith was noted. Pericaecal inflammatory changes, phlegmon, or an abscess without visualisation of an abnormal appendix or appendicolith was considered suggestive but not specific for appendicitis. In statistical analyses, patients with these non-specific findings were counted as appendicitis positive		
Assessors of the CT-scan	Not reported		
Notes	This was a comparative accuracy study evaluating CT and ultrasonography in the same patients		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Balthazar 1994 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Bouillot 2001

Study characteristics	
Patient sampling	During a 6-month period, consecutive adult patients admitted with suspected appendicitis were included. Patients with septic shock, allergy to IV contrast material, renal failure, or tablet-treated diabetes were excluded Recruitment period: not stated
Patient characteristics and setting	Age range (median): 17 to 91 (30). 43% women Department of Surgery in Paris, France. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Single-slice helical CT with IV contrast enhancement. It is unclear whether the entire abdomen was included in the scan. Model name and manufacturer of the CT-scanner: not stated. Slice thickness: 5.5 mm. Slice interval: 2.5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Alternative intraoperative findings were noted in patients who had a laparoscopy without appendectomy. Patients who did not have a laparoscopy were followed up clinically or by telephone calls for a year or longer
Flow and timing	100 patients were included - it is unclear if any patients were excluded. A laparoscopy was performed in 81 patients; 78 had an appendectomy and 72 had histological evidence of appendicitis. Of the 19 patients who did not have a laparoscopy, 2 were lost to follow-up. These 2 patients were classified as 'not appendicitis'
Comparative	
Criteria for CT diagnosis of appendicitis	One or more of the following 4 criteria were met: appendix diameter > 6 mm, signs of right lower quadrant inflammation, appendicolith, periappendicular collection
Assessors of the CT-scan	A single radiologist evaluated CT-scans for all patients at a later point in time than the scan date
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**

Bouillot 2001 *(Continued)*

DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	Yes		
		High	

Cakirer 2002

Study characteristics	
Patient sampling	Consecutive patients with clinically suspected appendicitis were referred for CT. Recruitment period: January 1999 to June 2000 Pregnant women and patients with prior appendectomy were excluded

Cakirer 2002 *(Continued)*

Patient characteristics and setting	Age range (mean): 16 to 67 years (34); 48% women. Pregnant women were excluded Radiology department - patients were referred through the Department of Emergency Surgery. Istanbul, Turkey. Single-centre study Disease spectrum: unclear
Index tests	Helical CT of the lower abdomen without contrast enhancement (SR 950 W, Hitachi). Slice thickness: 5 mm. Slice interval: 5 mm. Voltage: 120 kV. mAs product: 220 to 270 mAs
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendicectomy; follow-up was provided for patients who did not
Flow and timing	Of 130 included patients, 103 had surgery, 94 had appendicitis confirmed histologically, and 27 were followed up clinically. No details about follow-up were reported
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm and periappendicular inflammatory changes
Assessors of the CT-scan	3 experienced radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		

Cakirer 2002 *(Continued)*

Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Christopher 2002

Study characteristics	
Patient sampling	A convenience sample of patients presenting with signs and symptoms was considered by the examining physician as possibly having appendicitis Exclusion criteria: first trimester of pregnancy; obvious requirement for surgical intervention due to presence of a rigid abdomen, hypotension, or other signs of instability Recruitment periods: April 1998 to December 1998; April 2000 to October 2000
Patient characteristics and setting	Age range (mean): 5 to 77 years (32). 52% women. The proportion of patients younger than 10 years was 3% Emergency Department in an urban teaching hospital in Houston, Texas, USA. Single-centre study Disease spectrum: any suspicion of appendicitis

Christopher 2002 *(Continued)*

Index tests	Unenhanced helical CT of the lower abdomen (Picker PQ6000, Picker International; MX8000, Marconi Medical Systems). Slice thickness: 5 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: 250 to 300 mAs
Target condition and reference standard(s)	Appendicitis. Intraoperative findings and histopathological reports in patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up with telephone calls 6 to 8 weeks after presentation to the Emergency Department
Flow and timing	107 patients were included. Of these, 40 had surgery and 31 had appendicitis confirmed histologically. Six patients were lost to follow-up, and 1 patient withdrew consent before the CT-scan was obtained, hence 101 patients were included in the analyses
Comparative	
Criteria for CT diagnosis of appendicitis	Not reported
Assessors of the CT-scan	Attending general radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		High	**Unclear**
DOMAIN 3: Reference Standard			

Christopher 2002 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Cougard 2002

Study characteristics	
Patient sampling	Patients had been admitted with suspected appendicitis. No exclusion criteria were reported Recruitment period: February 1998 to February 2000
Patient characteristics and setting	Age range: not reported; mean age 33.9 years. The proportion of patients younger than 16 years is unclear. 61% women General hospital in Dijon, France. Single-centre study Disease spectrum: unclear
Index tests	Helical CT of the abdomen with IV contrast enhancement. Number of slices, slice thickness, slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histopathological findings were reported in patients who had surgery with appendectomy. Intraoperative findings were noted for patients who had surgery without appendectomy. Follow-up was 2 months for patients who did not have surgery
Flow and timing	89 patients were included. It is unclear whether any patients were excluded or dropped out. 60 patients had surgery; 44 had appendicitis; 29 were followed up for 2 months; none of these had an appendectomy

Cougard 2002 *(Continued)*

Comparative
Criteria for CT diagnosis of appendicitis: Appendix diameter > 5 mm, appendicolith, appendix wall thickening with hyperenhancement, periappendiceal or pericaecal fat stranding, fluid collection around the appendix or in the pouch of Douglas
Assessors of the CT-scan: Not stated
Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		

Cougard 2002 *(Continued)*

Did all patients with a positive CT-scan have surgery?	Yes	
Did all patients with a negative CT-scan have clinical follow-up?	No	
Was the choice of reference standard independent of the result of the index test?	No	
Were all patients included in the analyses?	Yes	
		High

del Cura 2000

Study characteristics	
Patient sampling	Patients presenting with clinical signs of appendicitis in daytime on labor days during a 1-year period. Pregnant women were excluded. Recruitment period: May 1997 to May 1998
Patient characteristics and setting	Age range (median): 4 to 92 years (31.5) - 12% were younger than 15 years of age. 56% women Emergency Department, Bilbao, Spain. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Non-helical focused CT of the lower abdomen (Somatom HiQ, Siemens in 137 patients; Excel 2400 Elite, Elscint in 15 patients). Enhancement with rectal contrast material was provided. Slice thickness: 5 mm. Slice interval: 5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed for patients who had an appendectomy. Follow-up was provided for 72 patients who did not have surgery (clinical control or telephone call 6 to 18 months after CT-scan)
Flow and timing	152 patients were recruited. All had CT of the lower abdomen, 80 had surgery, and 66 had appendicitis confirmed histologically. 72 patients were followed up with clinical control or telephone calls. No patients were lost to follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 cm or an appendicolith. Presence of gas or contrast material in the appendiceal lumen was considered evidence against appendicitis
Assessors of the CT-scan	One radiologist evaluated the CT-scan just after it was performed. A second radiologist evaluated CT-scans from all patients after completion of recruitment in June 1998. Evaluations coincided in 134 patients. Consensus evaluations from the 2 radiologists were used in analyses for the remaining 18 patients
Notes	Supplementary information and additional results from this study have been published in Radiologia 2001;43:175-186; and in Radiologia 2001;43:478-489

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

del Cura 2000 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Funaki 1998

Study characteristics	
Patient sampling	Patients presenting to an Emergency Department with equivocal symptoms and signs of appendicitis were referred for CT examination. Unequivocal cases of appendicitis underwent immediate laparotomy. Entrance criteria were based on the clinical judgement of the Emergency Department physician. No exclusion criteria were stated Recruitment period: May 1997 to January 1998
Patient characteristics and setting	Age range (mean): 6 to 71 years (ns). 63% women Radiology Department in Hilo, Hawaii, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Single-slice helical CT of the lower abdomen with oral (in 95%) and rectal (in 100%) contrast material (PQ 5000, Picker International). Slice thickness: 5 mm. Slice interval: 2.5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Surgical and histopathological findings were reported for patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up clinically for at least 2 months. No further description of follow-up was provided
Flow and timing	100 patients were included. It is unclear whether some were excluded from participation. 45 patients had surgery; 30 had appendicitis. Patients who did not have surgery were followed up clinically and were free of symptoms for at least 2 months after CT examination
Comparative	
Criteria for CT diagnosis of appendicitis	A non-opacified and enlarged (> 6 mm in diameter) appendix was noted. Ancillary signs of appendicitis, including right lower quadrant inflammation, appendicoliths. lymphadenopathy, and caecal apical changes (caecal bar or arrowhead sign), were also recorded. Findings were interpreted as negative if the appendix was visualised with intraluminal air or contrast material extending to its tip
Assessors of the CT-scan	11 board certified general radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		

Funaki 1998 *(Continued)*

		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Gamanagatti 2007

Study characteristics	
Patient sampling	Patients with suspected appendicitis based on history, clinical examination findings, and laboratory test results were recruited from an

Gamanagatti 2007 *(Continued)*

	Emergency Department. Pregnant women were excluded. Recruitment period: November 1999 to October 2001
Patient characteristics and setting	Age range (mean): 12 to 74 (25) years, 90% of participants were adults. 31% women Hospital in New Delhi, India. Single-centre study Disease spectrum: unclear
Index tests	Unenhanced single-slice or 4-slice helical CT of the lower abdomen (AR-Star or Somatom Plus 4, Siemens). Slice thickness and slice interval: 5 mm. Voltage: 120 kV. mAs product: 220 to 230 mAs in patients 15 years or older and 63 mAs in patients younger than 15 years
Target condition and reference standard(s)	Appendicitis. Finding consisted of a macroscopically inflamed appendix during surgery
Flow and timing	58 patients were included. 52 had surgery; 48 had appendicitis diagnosed at surgery. The 6 patients who did not have surgery were excluded from analyses
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, appendicolith, pericaecal or periappendiceal inflammation, fluid collection, abscess, or lymphadenopathy. CT findings were interpreted as negative if the appendix was visualised with intraluminal air
Assessors of the CT-scan	CT-scans were initially evaluated by attending resident radiologists. Two consultant radiologists reevaluated CT images at a later date. The latter evaluations were used in the analyses
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Gamanagatti 2007 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Unclear		
Were all patients included in the analyses?	No		
		High	

Hekimoglu 2011

Study characteristics	
Patient sampling	Patients presented with acute, non-traumatic abdominal pain clinically suspected to be secondary to acute appendicitis. Patients with possible contrast allergy, pregnant women, and patients with abdominal trauma were excluded Recruitment period: March 2008 to October 2010 Patients were randomly allocated to receive either CT with intravenous contrast enhancement or CT with intravenous and oral contrast enhancement
Patient characteristics and setting	Intravenous contrast group: age range (mean): 20 to 66 years (42); 38% women Intravenous and oral contrast group: age range (mean): 18 to 74 years (38); 42% women Pregnant women and patients with possible contrast allergy were excluded Emergency Department in Ankara, Turkey. Single-centre study Disease spectrum: unclear

Hekimoglu 2011 *(Continued)*

Index tests	16-slice CT of the entire abdomen with intravenous contrast enhancement (Sensation 16, Siemens Medical Solutions). Multi-planar reconstructions were used Slice thickness: 5 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: not stated Patients were randomly allocated to receive either CT with intravenous contrast enhancement or CT with intravenous and oral contrast enhancement
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy and follow-up was provided for patients who did not have surgery (review of medical records and telephone interviews 1 day and 1 week after discharge)
Flow and timing	Intravenous contrast group: 100 patients were included. All had CT; 26 had appendicitis confirmed histologically Intravenous and oral contrast group: 100 patients were included. All had CT; 32 had appendicitis confirmed histologically The number who had surgery and the number receiving follow-up are not reported for any of the groups
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, thickened appendix wall, appendix wall hyperenhancement, periappendiceal fat stranding. In patients who had oral contrast enhancement, the absence of contrast filling of the appendix was considered an additional criterion in favour of appendicitis Radiologists used a 5-point Likert scale to rate their confidence in the radiological diagnosis of appendicitis (1: definitely absent, 2: probably absent, 3: intermediate, 4: probably present, 5: definitely present) Patients rated 1 or 2 were considered CT negative; patients rated 4 or 5 were considered CT-positive. Patients rated 3 were counted as CT negative when appendicitis was confirmed, and as CT-positive when appendicitis was not confirmed (worst-case scenario)
Assessors of the CT-scan	All CT-scans were evaluated by 2 radiologists with over 5 years' experience in interpreting abdominal CT-scans Sensitivity and specificity were reported for each of the 2 radiologists. No consensus evaluation is available. Rounded mean numbers of true-positives, false-positives, false-negatives, and true negatives were used in meta-analyses
Notes	Patients were recruited for a randomised trial comparing CT with intravenous contrast enhancement vs CT with oral and intravenous contrast enhancement. The 2 groups are considered as individual studies in the meta-analyses

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		

Hekimoglu 2011 *(Continued)*

		High	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Hershko 2002

Study characteristics			
Patient sampling	Consecutive patients with suspected acute appendicitis were recruited. Pregnant women and patients with low (0 to 19%) or high (80% to 100%) clinical likelihood of acute appendicitis were excluded. Clinical likelihood was evaluated by the attending surgeon. Recruitment period: 1999 to 2001		
Patient characteristics and setting	Age range (mean): 15 to 83 years (31); 49% women. Patients with high (> 80%) and low (< 20%) risk of appendicitis based on symptoms and findings were excluded. Pregnant women were also excluded Emergency Department in Haifa, Israel. Single-centre study Disease spectrum: intermediate suspicion of appendicitis		
Index tests	Helical CT of the entire abdomen with oral and intravenous contrast enhancement (Twin RTS, Elscint CT). Slice thickness: 8 mm. Slice interval, voltage, and mAs product: not stated		
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery		
Flow and timing	206 patients were included. All had CT; 75 had appendicitis confirmed histologically. No details of clinical follow-up are reported		
Comparative			
Criteria for CT diagnosis of appendicitis	Not stated		
Assessors of the CT-scan	In-training and senior radiologists		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Hershko 2002 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Hershko 2007

Study characteristics	
Patient sampling	Consecutive patients with suspected acute appendicitis were randomly assigned to 1 of 3 CT-protocols. Patients who were pregnant or who had contraindications to intravenous contrast material, severe asthma, or chronic renal failure were excluded Recruitment period: June 2002 to January 2005
Patient characteristics and setting	Age range (mean): 16 to 83 years (30). 54% women Department of Surgery in Haifa, Israel. Single-centre study Disease spectrum: unclear
Index tests	Included patients were randomly allocated to 1 of the following 3 helical CT-protocols: • Unenhanced CT of the lower abdomen

Hershko 2007 *(Continued)*

	• CT of the lower abdomen with rectal contrast material • CT of the lower abdomen with oral and intravenous contrast material All CT-scans were performed with a multi-slice CT-scanner (NIX8000 - IDT TM, Philips). Slice thickness: 2.5 mm. Voltage: 120 kV. Slice interval and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Surgical and histopathological findings were reported for patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up clinically. Follow-up procedures and timing are not described
Flow and timing	• Unenhanced CT of the lower abdomen 70 patients were allocated. 14 were excluded due to inconclusive CT findings. 21 of the remaining 56 patients had appendicitis. Numbers who had surgery and follow-up were not reported • CT of the lower abdomen with rectal contrast material 78 patients were allocated. There were no inconclusive CT-scans. Appendicitis was found in 39 patients. Numbers who had surgery and follow-up were not reported • CT of the lower abdomen with oral and intravenous contrast material 84 patients were allocated. There were no inconclusive CT-scans. Surgery was performed in 48 patients with positive CT-scans; 43 had appendicitis confirmed histologically. All patients with negative CT-scans had uneventful follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Findings were interpreted as acute appendicitis if the appendix was > 6 mm in diameter and/or had surrounding signs of inflammation The appendix was interpreted as normal if it was < 7 mm in diameter or was filled to the tip with contrast material or air. Similarly, when the appendix was not visualised, the scan was interpreted as normal
Assessors of the CT-scan	6 radiology residents who were at least 2 years into their training programmes
Notes	3 randomised groups were considered as individual studies in the meta-analyses

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			

Hershko 2007 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Yes		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Holloway 2003

Study characteristics	
Patient sampling	Patients with right lower quadrant pain consistent with acute appendicitis were included

Holloway 2003 *(Continued)*

	Pregnant women, patients with unequivocal symptoms and signs of appendicitis, and patients with contraindications to the instillation of contrast material into the colon were excluded Recruitment period: January 1998 to July 2002
Patient characteristics and setting	Age range (mean): not stated. 60% women Community hospital in Scottsbluff, Nebraska, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	CT of the lower abdomen with rectal contrast enhancement. The model name and the manufacturer of the CT-scanner used in the study were not reported. Slice thickness: 5 mm. Slice interval: 5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histopathology was performed in patients who had an appendectomy, and follow-up was provided as needed for patients who did not have surgery
Flow and timing	423 patients were included; all had a CT-scan. Of these, 188 had appendicitis. The total number of patients who had surgery and the number who completed follow-up were not reported
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with surrounding signs of inflammation such as fat stranding, free fluid, abscess, phlegmon, appendicolith, or thickening of the adjacent caecal wall
Assessors of the CT-scan	Not reported
Notes	In an email correspondence, Dr. Jeffrey A Holloway confirmed that data were collected prospectively according to a protocol

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Holloway 2003 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Hong 2003

Study characteristics	
Patient sampling	All patients presenting to the Emergency Department with possible appendicitis and an Alvarado score of 2 to 8 were included Included patients were randomly allocated to clinical assessment with abdominal CT or clinical assessment alone Exclusion criteria: age younger than 18 years, inability to receive intravenous contrast, pregnancy, HIV-positive, patients awaiting interval appendectomy, unreliable clinical examination (steroid administration, known inflammatory bowel disease, sickle cell disease) Recruitment period: November 1998 to October 1999
Patient characteristics and setting	Mean age (SD): 34 (11). Lower age range: 18 years. Upper age range: not reported. 49% women Emergency Department in Miami, Florida, USA. Single-centre study

Hong 2003 *(Continued)*

	Disease spectrum: intermediate suspicion of appendicitis
Index tests	Single-slice helical CT of the entire abdomen and pelvis with intravenous and oral contrast enhancement (HiSpeed Advantage, General Electric). Slice thickness: 7 mm. Slice interval, voltage, and mAs product: not reported
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Follow-up with telephone interview after 1 week was planned for patients who did not have an appendectomy
Flow and timing	316 patients were evaluated for inclusion; 134 were excluded because the Alvarado score was outside the range of 2 to 8. One patient was excluded due to HIV positivity. Of the remaining patients, 97 were allocated to CT. 19 patients were excluded before the CT-scan was performed because they went directly to the operating room or withdrew consent. Of the 78 patients who had CT, 44 had surgery and 33 had appendicitis. Follow-up with telephone interviews was attempted in 34 patients; 28 could not be reached and 6 reported no new abdominal pain. The 28 patients who could not be reached were counted as appendicitis negative
Comparative	
Criteria for CT diagnosis of appendicitis	A dilated appendix (> 6 mm) with an enhancing rim or pericaecal soft tissue prominence
Assessors of the CT-scan	Radiology residents (third or fourth postgraduate year) and attending radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		

Hong 2003 *(Continued)*

		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Horton 2000

Study characteristics	
Patient sampling	Patients 18 to 65 years old presenting to the Emergency Department with equivocal symptoms and signs of appendicitis Patients with unequivocal symptoms and signs of appendicitis (symptom duration < 48 hours, migration of pain to right lower quadrant, rebound tenderness, anorexia, and white blood cell count > 10,000) were excluded Included patients were randomised to have ultrasonography or unenhanced CT of the abdomen Recruitment period: May 1997 to May 1999
Patient characteristics and setting	Age range (mean): not stated - inclusion criterion age 18 to 65 years Gender distribution: not stated for the CT group. Overall 54% women Emergency Department in Seattle, Washington, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis

Horton 2000 *(Continued)*

Index tests	Unenhanced CT of the lower abdomen. Model name and manufacturer of the CT-scanner were not stated. Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. It is unclear whether the reference standard is the intraoperative finding of an inflamed appendix or histological assessment of the removed appendix. Patients who did not have surgery were followed up
Flow and timing	106 patients with equivocal signs of appendicitis were included. 17 of these were withdrawn because the admitting surgeon believed that presentation was typical; these patients were admitted for surgery without diagnostic imaging. Of the remaining 89 patients, 49 were randomly allocated to CT. Among these, 38 had appendicitis and 37 had surgery with appendectomy. Follow-up was uneventful in 9 patients, other diagnoses were made in 2 patients, and 1 patient was treated for appendicitis with antibiotics
Comparative	
Criteria for CT diagnosis of appendicitis	One or more of the following: appendix diameter > 6 mm, appendicolith, inflamed pericaecal fat, pericaecal free fluid with or without gas bubbles
Assessors of the CT-scan	Not stated
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**Unclear**

Horton 2000 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

in't Hof 2004

Study characteristics	
Patient sampling	Patients presenting to the Emergency Department with suspected appendicitis were included Exclusion criteria: signs of acute bowel obstruction, contraindication to laparoscopy, contraindication to general anaesthesia or pneumoperitoneum, younger than 16 years of age, pregnancy, and sepsis. Signs of acute pancreatitis or acute aneurysm of the abdominal aorta or iliac arteries on CT were considered to be stopping points Recruitment period: December 1999 to November 2001
Patient characteristics and setting	Age range (median): 16 to 82 years (36). 38% women Emergency Departments in University Hospital Rotterdam and in Medical Centre Rijnmond-Zuid, the Netherlands Disease spectrum: high suspicion of appendicitis (all included patients were scheduled for laparoscopy due to suspected appendicitis)
Index tests	Unenhanced CT of the entire abdomen (LightSpeed Advantage, General Electric Medical Systems). Slice thickness: 5 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: 190 mAs

in't Hof 2004 *(Continued)*

Target condition and reference standard(s)	Appendicitis. The finding of an inflamed appendix on laparoscopy was considered the reference test for appendicitis. All patients had laparoscopy Non-inflammed appendices were not removed. Removed appendices were sent for pathological examination
Flow and timing	103 patients were included. All had laparoscopy and CT. 87 had an appendectomy because appendicitis was confirmed by laparoscopy. All removed appendices were inflamed on microscopic examination
Comparative	
Criteria for CT diagnosis of appendicitis	Transverse appendix diameter > 6 mm, periappendiceal infiltration, thickening of the caecal wall, presence of an appendicolith, periappendiceal phlegmon or abscess, and adenopathy
Assessors of the CT-scan	At completion of the study, all scans were reviewed by an expert radiologist who was blinded to clinical history and to surgical findings
Notes	This study is reported in 2 publications (in't Hof 2004 and in't Hof 2009). In in't Hof 2004, sensitivity and specificity are reported for 1 expert radiologist. In in't Hof 2009, sensitivity and specificity are reported for 3 observers: a resident radiologist, an on-call radiologist, and an expert abdominal radiologist. We would have preferred to extract a 2×2 table for performance of the on-call radiologist reported in in't Hof 2009. Unfortunately, the reported information is inconsistent. We contacted study authors by email, but we have not received a reply to our enquiry. Therefore, we extracted 2×2 tables from the information reported in in't Hof 2004

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		

in't Hof 2004 *(Continued)*

Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	Yes		
		Low	

Jacobs 2001

Study characteristics	
Patient sampling	All patients with right lower quadrant pain in whom a CT examination was requested to evaluate for acute appendicitis were asked to participate Exclusion criteria: prior appendectomy, Crohn's disease, inability to receive oral or intravenous contrast material Recruitment period: August 1997 to April 1999 Included patients initially had focused CT of the lower abdomen with oral contrast material. Immediately thereafter, an intravenous contrast enhanced CT-scan of the entire abdomen and pelvis was performed
Patient characteristics and setting	Age range (mean): 13 to 87 years (32). The proportion of patients younger than 15 years of age is not reported. 64% women

Jacobs 2001 *(Continued)*

	Department of Radiology, Philadelphia, Pennsylvania, USA. Single-centre study Disease spectrum: unclear
Index tests	Included patients initially had focused CT of the lower abdomen with oral contrast material. Immediately thereafter, an intravenous contrast enhanced CT-scan of the entire abdomen and pelvis was performed. Hence, each patient was examined with 2 different CT-protocols CT examination was performed with single-slice CT-scanners (CTi or HiSpeed Advantage; General Electric Medical Systems) Slice thickness: 5 mm. Slice interval and voltage: not stated. mAs product: 200 to 220
Target condition and reference standard(s)	Appendicitis. The reference standard had 2 components: intraoperative findings in patients who had surgery, and follow-up in patients who did not have surgery. What follow-up consisted of is not reported
Flow and timing	228 patients were included; all were examined with both CT-protocols. 58 patients had surgery; 152 had clinical follow-up. 18 patients were lost to follow-up and were excluded from the analyses, hence 210 patients were included in the analyses. 51 patients had appendicitis
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, abscess or phlegmon in the right iliac fossa, appendicolith, periappendiceal fat stranding, appendix wall enhancement, thickened appendix wall The radiologist graded the likelihood of appendicitis on a 5-point scale. 1: definitely absent, 2: probably absent, 3: indeterminate, 4: probably present, 5: definitely present. In the accuracy analyses, patients with grade 4 or 5 likelihood of appendicitis were considered CT-positive
Assessors of the CT-scan	3 radiologists with varying experience (20 years, 3 years, and 1 month after American Board of Radiology certification) examined both CT-scans from all patients. In the meta-analyses, average values of sensitivity and specificity across the 3 radiologists were used to generate 2×2 tables
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			

Jacobs 2001 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Jo 2010

Study characteristics	
Patient sampling	Consecutive patients presented to the Emergency Department with pain in the right lower quadrant of the abdomen. Patients younger than 15

Jo 2010 *(Continued)*

	years and patients referred from other hospitals with confirmed diagnoses of appendicitis were excluded
Patient characteristics and setting	Mean age: 37.3 years, 54% women. Patients younger than 15 years, pregnant women, patients with renal insufficiency, and patients with allergy to contrast medium were excluded Emergency Department in Seoul, Korea Disease spectrum: any suspicion of appendicitis
Index tests	CT of the entire abdomen with intravenous contrast enhancement (Brilliance, Philips Medical Systems). No further information about the CT-scanner and the CT-protocol
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery (telephone calls with structured interview 3 months after CT)
Flow and timing	278 patients were included; 91 were withdrawn (see notes). Of the remaining 187 patients, 120 had surgery and 67 had follow-up. 111 patients had appendicitis
Comparative	
Criteria for CT diagnosis of appendicitis	Not reported. Radiologists' confidence in the diagnosis was scored on a 5-point scale, with 1 indicating normal appendix and 5 indicating definite appendicitis
Assessors of the CT-scan	3 board certified body imaging radiologists
Notes	This study compares the accuracy of 3 index tests: CT and clinical assessments made by surgical and emergency medicine residents. The reason for withdrawal of 87 of 91 patients was lack of evaluation by the surgical resident

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Jo 2010 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Kan 2001

Study characteristics	
Patient sampling	Patients with an equivocal clinical diagnosis of appendicitis were referred for CT at the discretion of the Emergency Department staff. No exclusion criteria were reported. Recruitment period: September 2000 to March 2001
Patient characteristics and setting	Age range (mean): 18 to 57 years (34); 84% women Radiology Department in Chicago, Illinois, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	CT of the abdomen - extent unclear (Lightspeed, HiSpeed; General Electric Medical Systems) 74% of patients had enhancement with rectal and oral contrast medium 26% of patients had enhancement with rectal contrast only 23% of patients had additional enhancement with intravenous contrast medium

Kan 2001 *(Continued)*

	Slice thickness: not stated. Slice interval: not stated. Voltage: not stated. mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Clinical outcome was determined after chart review or telephone contact for included patients 1 to 4 months after diagnostic imaging. Unclear whether the reference standard in operated patients consisted of intraoperative findings or histological examination of the removed appendix
Flow and timing	35 patients were screened for inclusion and 31 were included. 4 patients had appendicitis. The number of patients who had surgery is unclear
Comparative	
Criteria for CT diagnosis of appendicitis	Non-filling appendix with diameter > 6 mm, periappendiceal fat stranding, appendicolith, caecal wall thickening, periappendiceal free fluid Integration of criteria not stated
Assessors of the CT-scan	Radiologists. Not otherwise specified
Notes	This study is included in subgroup meta-analysis of rectal and oral contrast enhancement because most patients (76%) had this type of enhancement

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**High**
DOMAIN 3: Reference Standard			

Kan 2001 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Karabulut 2014

Study characteristics	
Patient sampling	Patients with suspected appendicitis were included. No exclusion criteria were reported Recruitment period: December 2005 to December 2008
Patient characteristics and setting	Age range (mean): 6 to 77 years (27) - proportion younger than 15 years not reported (study authors contacted - no response). 52% women. Exclusion criteria not reported Disease spectrum and clinical setting: not stated. Single-centre study
Index tests	Two helical CT-scanners were used: Brillance 16 (Philips Medical Systems): 16-slice. Slice thickness: 3 mm. Slice interval: 1.5 mm. Voltage: 120 kV. mAs product: 50 mAs. Unenhanced. Lower abdomen MW8000 (Philips Medical Systems): 2-slice. Slice thickness: 3.2 mm. Slice interval: 1.6 mm. Voltage: 120 kV. mAs product: 50 mAs. Unenhanced. Lower abdomen
Target condition and reference standard(s)	Appendicitis. Histological examination in patients who had an appendectomy. Otherwise alternative intraoperative findings. Follow-up in patients who did not have surgery - review of medical charts and/or a telephone call after 21 to 31 days

Karabulut 2014 *(Continued)*

Flow and timing	104 patients were recruited. All had CT of the lower abdomen, 40 had surgery, and 39 had appendicitis confirmed histologically. 64 patients were followed up
Comparative	
Criteria for CT diagnosis of appendicitis	Enlarged outer appendix diameter (threshold not stated), thickened appendix wall, appendicolith, periappendicular fat stranding, pericaecal or periappendicular fluid or abscess. Enlarged appendix diameter was not accepted as a single criterion unless it was accompanied by intraluminal, mural, or periappendicular soft tissue changes
Assessors of the CT-scan	1 senior radiologist
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Unclear	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**

Karabulut 2014 *(Continued)*

DOMAIN 4: Flow and Timing	
Did all patients receive a reference standard?	Yes
Did all patients receive the same reference standard?	No
Did all patients with a positive CT-scan have surgery?	No
Did all patients with a negative CT-scan have clinical follow-up?	No
Was the choice of reference standard independent of the result of the index test?	No
Were all patients included in the analyses?	Yes
	High

Kepner 2012

Study characteristics	
Patient sampling	Patients 18 years of age or older with clinically suspected appendicitis were referred for CT by Emergency Department (ED) physicians Exclusion criteria included pregnancy, allergy to intravenous or oral contrast material, creatine level ≥ 1.5 (unit not reported), current incarceration, inability to give informed consent, appendicitis not primary concern of the ED physician Included patients were randomised to receive CT of the entire abdomen with either intravenous (IV) contrast material or IV and oral contrast material
Patient characteristics and setting	IV contrast: age quartiles (median): 22 to 40 years (32). 59% women IV and oral contrast: age quartiles (median): 25 to 43 years (32). 55% women Emergency Department in York, Pennsylvania, USA. Single-centre study Disease spectrum: unclear
Index tests	CT of the entire abdomen and pelvis via a 16-slice CT-scanner (Somatom Sensation, Siemens Medical Solutions). Slice thickness: 3 mm. Slice interval, voltage, and mAs product: not stated Included patients were randomised to receive enhancement by either intravenous (IV) contrast material or intravenous and oral contrast material
Target condition and reference standard(s)	Appendicitis. Intraoperative findings were used to confirm appendicitis in patients who were operated on. Patients who did not have surgery were followed up Follow-up consisted of telephone calls within 1 week to 1 month after discharge. Letters with questionnaires and stamped return envelopes were sent to patients who could not be reached by telephone
Flow and timing	Overall 244 patients were included. 17 patients were excluded, including 3 patients lost to follow-up. 114 patients were allocated to IV contrast, and 113 were allocated to IV and oral contrast IV contrast: 114 patients had CT. Appendicitis was found in 41 patients IV and oral contrast: 113 patients had CT. Appendicitis was found in 34 patients

Kepner 2012 *(Continued)*

	The number of patients who had surgery and the number who had follow-up are not reported for any of the groups
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 5 mm, localised abscess or fluid collection, appendicolith, periappendiceal fat stranding, hyperenhancement of the appendix mucosa, thickened appendix wall (> 2 mm) After evaluating the CT-scan, the radiologist had the option of 'yes', 'no', or 'possible' for the diagnosis of appendicitis. In the analyses, 'possible' assessments were counted as 'yes'
Assessors of the CT-scan	All CT-scans were initially assessed by an attending radiologist. These assessments were not analysed Subsequently, 2 board certified radiologists with 18 years' and 27 years' experience re-assessed all CT-scans independently. These reassessments were used in the analyses. The board certified radiologists were unaware of the initial assessment. It is unclear how the 2 radiologists' assessments were aggregated into the reported single estimates for sensitivity and specificity
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**

Kepner 2012 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	Yes		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Keyzer 2004

Study characteristics	
Patient sampling	Consecutive patients older than 15 years of age with right lower quadrant pain who had a CT-scan requested by the Emergency Department physician to evaluate for acute appendicitis. Exclusion criteria were prior appendectomy and pregnancy. Recruitment period: March 2002 to February 2003
Patient characteristics and setting	Age range (mean): 16 to 81 years (38); 63% women. Pregnant women and patients with prior appendectomy were excluded Emergency Department in Brussels, Belgium. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	• Standard dose, 4-slice CT of the entire abdomen without contrast enhancement (Somatom Plus Volume Zoom, Siemens). Slice thickness: 3 mm. Slice interval: 1.5 mm. Voltage: 120 kV. mAs product: 100 mAs • Low-dose CT with mAs product: 30. Otherwise as above

Keyzer 2004 *(Continued)*

	Paired design with direct comparison of low-dose and standard-dose CT
Target condition and reference standard(s)	Appendicitis. Histological examination in patients who had an appendectomy; follow-up for patients who did not have surgery (telephone calls 1 month after the CT-scan)
Flow and timing	94 patients were included. All had CT, 30 had appendicitis confirmed histologically, and the remainder had clinical follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Outer appendix diameter, appendicolith, periappendiceal fat stranding, caecal wall thickening, and abscess or phlegmon in the right iliac fossa. The presence of gas in the appendiceal lumen was considered to be a possible negative criterion for appendicitis After separate coding for each of these signs, readers were asked to propose an overall diagnosis of appendicitis (same for Keyzer 2004 and Keyzer 2005)
Assessors of the CT-scan	Keyzer 2004: a board certified radiologist and a 3-year radiologist resident with no specific coaching or training before the study Results from the former are used in the meta-analyses Keyzer 2005: radiologists responsible for emergency examinations (5 board certified radiologists with more than 10 years' experience, and 8 resident or general radiologists with 3 to 7 years' experience)
Notes	Keyzer 2004 and Keyzer 2005 are considered to report the same study because there is a 90% overlap in participants between the 2 reports (C. Keyzer, personal communication). Results from Keyzer 2005 are used in the overall meta-analyses because results from the initial clinical evaluation of the CT-scan are reported here. In Keyzer 2004, results stem from reevaluations of the CT-scans; these results are used in the subgroup analysis of low-dose CT

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Keyzer 2004 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Keyzer 2009

Study characteristics	
Patient sampling	Consecutive patients aged 18 years or older with right lower quadrant abdominal pain who were referred for a CT examination due to suspected acute appendicitis. Patients with prior appendectomy or possible pregnancy were excluded. Recruitment period: May 2005 to November 2005 Patients were randomly allocated to receive enhancement by oral contrast material (group 1) or no such enhancement (group 2). In both groups, CT was initially performed without intravenous contrast. Subsequently, intravenous contrast enhancement was administered and another CT was performed. The first and second CTs were acquired at standard dose. Both of these scans were then manipulated via a comput-

Keyzer 2009 *(Continued)*

	er-assisted method to produce simulated low-dose CTs. Hence, 4 different CT-scans were evaluated for each patient
Patient characteristics and setting	Group 1 (enhancement by oral contrast material): age range (mean): 18 to 87 years (36); 66% women Group 2 (unenhanced): age range (mean): 18 to 82 years (37); 56% women Emergency Department in Brussels, Belgium. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	4-slice CT of the entire abdomen (Somatom Plus Volume Zoom, Siemens). Slice thickness: 3 mm. Slice interval: 1.5 mm. Voltage: 120 kV. mAs product: 100 mAs. Noise was added to yield an mAs product of 30 for simulated low-dose CT
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy. Follow-up was provided for patients who did not have surgery (review of medical records and telephone calls 1 month after CT-scan)
Flow and timing	Group 1: of 65 allocated patients, 13 had histologically confirmed appendicitis Group 2: of 66 allocated patients, 20 had histologically confirmed appendicitis The number of patients having surgery or follow-up is unclear for both groups. No patients were lost to follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 8 mm, abscess or phlegmon in the right iliac fossa, appendicolith, periappendiceal fat stranding. Presence of gas or contrast material in the appendiceal lumen was considered evidence against appendicitis
Assessors of the CT-scan	2 board certified radiologists. Both radiologists evaluated CT-scans for all patients. No consensus evaluation is reported
Notes	Cell counts in the 2×2 tables are the mean numbers calculated by 2 radiologists

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Keyzer 2009 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Kim 2008

Study characteristics	
Patient sampling	All patients presenting to the Emergency Department with symptoms or signs suggestive of appendicitis were enrolled into the study. Exclusion criteria: younger than 15 years of age, pregnancy, previous CT contrast allergy, renal insufficiency, creatine level > 1.5 mg/dL. Patients transferred from another hospital were also excluded

Kim 2008 *(Continued)*

	Recruitment period: not stated
Patient characteristics and setting	Age range (mean): 15 to 84 (37.1). 60% women Clinical setting: 2 emergency departments in Seoul, Korea, and in Stony Brook, New York, USA Disease spectrum: any suspicion of appendicitis
Index tests	16-slice helical CT of the entire abdomen (Brilliance, Philips Medical Systems). Enhancement with intravenous contrast material. Slice thickness, slice interval, and mAs product: not stated. Voltage: 120 kV
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy. Follow-up was provided for patients who did not have surgery - review of hospital course and a telephone call within 3 months
Flow and timing	157 patients were recruited. All had CT of the lower abdomen, 91 had surgery, and 90 had appendicitis confirmed histologically. 66 patients were followed up
Comparative	
Criteria for CT diagnosis of appendicitis	Not stated
Assessors of the CT-scan	All scans were read by 2 board certified attending radiologists specialising in CT imaging
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		

Kim 2008 *(Continued)*

Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Kim 2012

Study characteristics	
Patient sampling	Patients aged 15 to 44 years were referred for CT examination by Emergency Department physicians due to clinically suspected appendicitis. Patients with prior appendectomy, pregnant women, patients with allergy to intravenous contrast material, patients with impaired renal function, and patients who had prior cross-sectional imaging tests to evaluate the presenting symptoms were excluded. Recruitment period: September 2009 to January 2011 Included patients were randomised to receive low-dose or standard-dose intravenous contrast-enhanced CT
Patient characteristics and setting	Low-dose CT: age quartiles (median): 22 to 36 years (29); 62% women Standard-dose CT: age quartiles (median): 22 to 37 years (30); 59% women Emergency Department and Department of Radiology in Seoul, Korea. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Intravenous contrast-enhanced abdominal CT using 16-, 64-, or 256-slice CT-scanners. Slice thickness: 2 to 5 mm. Slice interval, voltage, and mAs product: not stated. Unclear whether

Kim 2012 *(Continued)*

	the CT-protocol included the entire abdomen and pelvis. Manufacturer of CT-scanners and model name are not reported Low-dose CT: intended radiation dose 2 mSv Standard-dose CT: intended radiation dose 8 mSv
Target condition and reference standard(s)	Appendicitis. The reference standard had 3 components: intraoperative findings in patients who had surgery but no appendectomy, histological examination of the resected appendix in patients who had an appendectomy, and follow-up in patients who did not have surgery. Follow-up was based on review of medical records and telephone interviews 3 months after presentation
Flow and timing	1035 patients were eligible for inclusion, 444 were randomised to have low-dose CT, and 447 were randomised to have standard-dose CT In the low-dose CT group, 189 patients had surgery, 172 had appendectomy (166 had appendicitis), 249 had follow-up, and 6 patients were lost to follow-up In the standard-dose CT group, 195 patients had surgery, 186 had appendectomy (180 had appendicitis), 246 had follow-up, and 6 patients were lost to follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, abscess or phlegmon in the right iliac fossa, appendicolith, periappendiceal fat stranding, abnormal appendix wall enhancement, thickened appendix wall The radiologist graded the likelihood of appendicitis on a 5-point scale. 1: definitely absent, 2: probably absent, 3: indeterminate, 4: probably present, 5: definitely present. In the accuracy analyses, patients with grade 3 to 5 likelihood of appendicitis were considered CT-positive
Assessors of the CT-scan	During daytime: 3 expert radiologists. During after-hours: on-call radiologists with various levels of expertise
Notes	The low-dose group and the standard-dose group enter the meta-analyses as 2 separate studies

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			

Kim 2012 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Lane 1999

Study characteristics

Lane 1999 *(Continued)*

Patient sampling	Consecutive patients with suspected appendicitis were referred for CT from Departments of Emergency Medicine and Surgery. Referral for CT was based on the clinical judgement of the referring physician. No exclusion criteria were reported Recruitment period: not stated
Patient characteristics and setting	Age range: 8 to 86 years. Mean/median age and proportion of patients younger than 15 years are not reported. 52% women Department of Radiology, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Unenhanced single-slice helical CT of the entire abdomen (HiSpeed Advantage, General Electric Medical Systems). Slice thickness and slice interval: 5 mm. Voltage: 120 kV. mAs product: 240 to 270 mAs
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. It is not stated how follow-up was performed
Flow and timing	300 patients were included. The number who had surgery is not reported. Appendicitis was confirmed histologically in 115 patients. All patients who did not have surgery were followed up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with periappendiceal inflammatory changes
Assessors of the CT-scan	Body imaging fellows or attending radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Lane 1999 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Lopez 2007

Study characteristics	
Patient sampling	Female patients between the ages of 18 and 45 years, presenting to the Emergency Department with possible appendicitis and an Alvarado score of 2 to 8, were included. Included patients were randomly allocated to clinical assessment with abdominal CT or clinical assessment alone Exclusion criteria: inability to receive intravenous contrast, pregnancy, HIV-positive, patients awaiting interval appendectomy, inflammatory bowel disease. No patients were excluded based on these criteria Recruitment period: November 1999 to February 2001; March 2003 to December 2004
Patient characteristics and setting	Age range (mean): 18 to 45 (27.9). 100% women Emergency Department in Miami, Florida, USA. Single-centre study

Lopez 2007 *(Continued)*

	Disease spectrum: intermediate suspicion of appendicitis
Index tests	Single-slice helical CT of the entire abdomen and pelvis with intravenous and oral contrast enhancement (HiSpeed Advantage, General Electric). Slice thickness: 7 mm. Slice interval, voltage, and mAs product: not reported
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Patients who did not have an appendectomy were followed up with a telephone interview after 1 week
Flow and timing	95 patients were screened for inclusion and 90 were included (2 were excluded due to Alvarado score < 2, and 3 refused to participate). 42 patients were allocated to clinical assessment and CT. Of these, 20 had surgery and 19 had appendicitis. Of the 22 patients scheduled for follow-up with telephone interviews, 12 could not be reached. These 12 patients were considered true-negatives in the analysis
Comparative	
Criteria for CT diagnosis of appendicitis	A dilated appendix (> 6 mm) with an enhancing rim or pericaecal soft tissue prominence
Assessors of the CT-scan	Senior radiology residents and attending radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**

Lopez 2007 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Malone 1993

Study characteristics	
Patient sampling	Patients with equivocal symptoms and signs of appendicitis were referred to a Radiology Department for an emergency barium enema. No exclusion criteria were stated Recruitment period: May 1991 to not stated
Patient characteristics and setting	Age range (mean): 4 to 91 years (ns). 59% women Radiology Department in Arlington Heights, Illinois, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Non-helical CT of the lower abdomen without contrast enhancement (GE 9800 or PACE, General Electric). Slice thickness and interval: 10 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Surgical reports and histopathological reports were provided for patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up clinically for up to 6 months - patients were contacted to determine if symptoms

Malone 1993 *(Continued)*

	had resolved, and if surgery had been performed elsewhere at a later date
Flow and timing	211 patients were included, 94 had surgery, and 75 had appendicitis. The 117 patients who did not have surgery were followed up, and none had appendicitis
Comparative	
Criteria for CT diagnosis of appendicitis	A thickened appendix > 6 mm with associated inflammatory changes in the periappendiceal fat and/or abnormal thickening in the right lateroconal fascia with or without an appendicolith
Assessors of the CT-scan	Not stated
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**

Malone 1993 *(Continued)*

DOMAIN 4: Flow and Timing		
Did all patients receive a reference standard?	Yes	
Did all patients receive the same reference standard?	No	
Did all patients with a positive CT-scan have surgery?	Yes	
Did all patients with a negative CT-scan have clinical follow-up?	No	
Was the choice of reference standard independent of the result of the index test?	No	
Were all patients included in the analyses?	Yes	
		High

Maluccio 2001

Study characteristics	
Patient sampling	Consecutive patients 18 years of age or older presented with symptoms and signs for which appendicitis was 1 of the first 3 considerations in the differential diagnosis. Patients who had a CT-scan at another institution before presentation were excluded. No other exclusion criteria were reported
Patient characteristics and setting	Mean age: 38 years. 66% women. Patients younger than 18 years were excluded Emergency Department in New York, New York, USA. Single-centre study. Recruitment period: July to December 1999 Disease spectrum: intermediate suspicion of appendicitis
Index tests	Helical CT of the entire abdomen with oral and intravenous contrast enhancement (HiSpeed Advantage, General Electric). Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery
Flow and timing	125 patients were included. 21 underwent appendectomy without preoperative CT. Of the 104 analysed patients, 35 had appendicitis. 15% of patients intended for follow-up could not be reached and were considered appendicitis negative in the analyses. The number of patients having surgery is unclear
Comparative	
Criteria for CT diagnosis of appendicitis	Not stated
Assessors of the CT-scan	Attending radiologists

Maluccio 2001 *(Continued)*

Notes

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		

Maluccio 2001 *(Continued)*

Were all patients included in the analyses?	Yes
	High

Megibow 2002

Study characteristics	
Patient sampling	Consecutive patients with suspected appendicitis referred for CT were included. It is unclear whether CT was performed in all patients with suspected appendicitis. No exclusion criteria were reported
Patient characteristics and setting	Age, gender distribution, and proportion of paediatric patients are not stated Hospital in New York, New York, USA. Single-centre study. Recruitment period: February to August 2000 Disease spectrum: unclear
Index tests	Single-slice CT of the entire abdomen with IV and oral contrast enhancement (CT/I Performix or HiSpeed RP, General Electric Medical Systems). Slice thickness: 7 mm. Slice interval: 6 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Intraoperative findings and clinical follow-up were reported for patients who did not have surgery. It is not stated how follow-up was undertaken
Flow and timing	58 patients were included, 32 had appendicitis, and 26 had alternative causes for the clinical presentation. The number who had surgery is not stated. 5 patients were excluded because they did not satisfy the requirements of the CT-protocol, hence 53 patients were included in the analyses
Comparative	
Criteria for CT diagnosis of appendicitis	Radiologists scored cases on a 0 to 4 confidence scale: 0: absolutely no appendicitis, 1: probably no appendicitis, 2: indeterminate, 3: probably appendicitis, 4: absolutely appendicitis In the analyses, cases with scores of 3 or 4 were counted as CT-positive for appendicitis
Assessors of the CT-scan	4 senior radiologists. The extracted 2×2 table was calculated from mean values of sensitivity and specificity for the 4 radiologists
Notes	The main purpose of the study was to assess if lossy wavelet compression can be applied to CT images without compromising diagnostic performance. The CT-scan for each patient was presented to the radiologist with 3 levels of compression and as uncompressed. The sequences of compressed and uncompressed images were randomised for each patient. The extracted 2×2 table is based on radiologists' assessment of uncompressed images

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Megibow 2002 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	Unclear		
Were all patients included in the analyses?	No		
		High	

Mittal 2004

Study characteristics	
Patient sampling	Patients for whom the clinical diagnosis of appendicitis was uncertain were included. Patients younger than 6 years of age, pregnant women, and patients with contraindications to contrast material were excluded. Recruitment period: December 2000 to December 2002 Based on the last digit of their medical record numbers, patients were allocated to triple-contrast CT of the entire abdomen (even number) or CT of the lower abdomen with rectal contrast only (odd number). Triple contrast consisted of enhancement with intravenous, oral, and rectal contrast material
Patient characteristics and setting	Triple-contrast group: mean age: 43; 52% women. Rectal contrast group: mean age: 33; 59% women Hospital in Southfield, Michigan, USA. Disease spectrum: intermediate suspicion of appendicitis
Index tests	Triple-contrast group: single-slice helical CT of the entire abdomen enhanced with intravenous, oral, and rectal contrast material. CT device not reported. Slice thickness: 5 mm. Slice interval: 5 mm. Voltage and mAs product: not stated Rectal contrast group: single-slice helical CT of the lower abdomen enhanced with rectal contrast material. CT device not reported. Slice thickness: 5 mm. Slice interval: 5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy. Otherwise alternative intraoperative findings. Follow-up was provided to rule out readmission for acute appendicitis for patients who did not have surgery
Flow and timing	Triple-contrast group: of 52 allocated patients, 48 had surgery and 44 had appendicitis. 4 patients had follow-up Rectal contrast group: all of the 39 allocated patients had surgery, and 36 had appendicitis. None had follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix outer diameter > 6 mm, adjacent inflammatory changes such as fat stranding, phlegmon, or fluid collection. Appendicitis was diagnosed in cases with non-visualisation of the appendix only in the presence of appendicolith, focal caecal apical thickening, an arrowhead sign, or a caecal bar sign. Contrast material in the appendiceal lumen was considered evidence against appendicitis
Assessors of the CT-scan	1 or more senior radiologists
Notes	The corresponding authors have been contacted about the proportion of patients younger than 15 years of age. No reply was received The rectal contrast group and the triple-contrast group are considered as 2 separate studies in the meta-analyses

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Mittal 2004 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Unclear		

Mittal 2004 *(Continued)*

Were all patients included in the analyses?	Yes

High

Moteki 2009

Study characteristics	
Patient sampling	Consecutive patients were referred for CT due to clinically suspected appendicitis. It is unclear whether all patients with suspected appendicitis were evaluated with CT. Patients younger than 16 years of age were excluded. Recruitment period: January 2004 to May 2006
Patient characteristics and setting	Age range (mean): 16 to 91 years (41); 42% women. Patients younger than 16 years were excluded Radiology Department - patients were referred by their attending doctor for CT studies. Fujioka, Japan. Single-centre study Disease spectrum: unclear
Index tests	4-slice helical CT of the entire abdomen with intravenous contrast enhancement (Lightspeed Plus, General Electric Medical Systems). Slice thickness: 2.5 to 3.75 mm. Voltage: 120 to 140. Slice interval and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination of the appendix specimen was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery
Flow and timing	285 patients were included. Results from 26 patients treated with antibiotics for appendicitis are excluded from the analyses in this review. 89 patients had surgery, and 86 had appendicitis confirmed histologically. 170 patients had follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm and periappendicular inflammatory changes Maximum depth of intraluminal appendiceal fluid > 2.6 mm
Assessors of the CT-scan	6 radiologists with 3 to 16 years' experience in radiology
Notes	Analysis using second criterion above is not included in meta-analyses because it is used only in this study

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Moteki 2009 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Nathan 2008

Study characteristics

Nathan 2008 *(Continued)*

Patient sampling	Consecutive adult patients presented to the Emergency Department with clinical suspicion of appendicitis. Patients younger than 18 years of age, pregnant women, and patients in whom intravenous contrast material was contraindicated were excluded Recruitment period: August 2006 to November 2006
Patient characteristics and setting	Age range (mean): 18 to 66 years (30); 75% women. Patients younger than 18 years, pregnant women, and patients in whom IV contrast material was contraindicated were excluded Emergency Department of a 300-bed community hospital in Seattle, Washington, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	16-slice CT of the lower abdomen with IV contrast enhancement (LightSpeed Pro 16, General Electric Healthcare). Slice thickness: 2.5 mm. Voltage: 120 kV. Slice interval and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Of the 100 analysed patients, 19 had surgery with histological assessment of removed appendices, and 81 received follow-up. Follow-up consisted of assuming the absence of appendicitis if no interval appendectomy was performed at the study hospital for at least 2 weeks after the Emergency Department encounter
Flow and timing	115 patients were included; 15 were excluded from the analyses due to insufficient data collection
Comparative	
Criteria for CT diagnosis of appendicitis	Radiologists were given no explicit criteria for diagnosis or exclusion of appendicitis. Equivocal CT results were coded as normal
Assessors of the CT-scan	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Nathan 2008 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Nemsadze 2009

Study characteristics	
Patient sampling	Patients with intermediate probability of appendicitis were recruited from the Emergency Department. Intermediate risk of appendicitis was defined as an Alvarado score of 4 to 6 Recruitment period: May 2007 to May 2008
Patient characteristics and setting	Age range: 18 to 43 years. 82% women Emergency Department in Georgia. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	64- or 16-slice abdominal CT (Lightspeed and Bright Speed, General Electric). 50% of study participants received oral contrast material

Nemsadze 2009 *(Continued)*

	Slice thickness, slice interval, voltage, and mAs product: not stated. 50% of study participants received oral contrast material
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological findings were reported for patients who had an appendectomy. It is unclear whether patients who did not have surgery were followed up
Flow and timing	60 patients were included; 55 had surgery and 41 had appendicitis. 5 patients who did not have surgery were excluded from the analysis
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, periappendiceal inflammation, appendicolith, absence of contrast material in the appendix lumen
Assessors of the CT-scan	Not stated
Notes	50% of patients had oral contrast material. Therefore, results from this study are not included in the subgroup analysis for CT with oral contrast enhancement Results are included only in the overall meta-analysis This study was reported in Russian. We are grateful to Dr. Anna Aaresøn for extracting data from this study. Study authors were not contacted for subgroup results according to type of contrast enhancement

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**High**
DOMAIN 3: Reference Standard			

Nemsadze 2009 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	Unclear		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	No		
		High	

Ozturk 2014

Study characteristics	
Patient sampling	Patients with suspected appendicitis were enrolled. Patients who had a definitive diagnosis and treatment without the use of CT were excluded. Otherwise no exclusion criteria were reported Recruitment period: July 2010 to November 2011
Patient characteristics and setting	Age range (median): 5 to 85 years (median age 33); 15% younger than 15 years. 42% women Hospital setting in Istanbul, Turkey. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Standard-dose 64-slice CT of the entire abdomen without contrast enhancement (Somatom Sensation, Siemens). Slice thickness, slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery. Follow-up consisted of outpatient visits, the timing of which is unclear
Flow and timing	125 patients with suspected appendicitis who had CT were included. Of these patients, 93 had surgery, 83 had appendicitis confirmed histologically, and 32 were intended for outpatient follow-up. It is unclear whether all 32 patients received follow-up. All 125 patients were included in the analyses

Ozturk 2014 *(Continued)*

Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 8 mm, appendicolith, periappendicular fluid collection, increased appendix wall thickness, appearance of inflammation in the mesoappendix
Assessors of the CT-scan	Radiologist on duty. Management of patients was planned according to results of the CT
Notes	The corresponding author provided information about the CT-scanner, the CT-protocol, and the numbers of true-positives, false-positives, false-negatives, and true-negatives

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		High	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			

Ozturk 2014 *(Continued)*

Did all patients receive a reference standard?	Unclear	
Did all patients receive the same reference standard?	No	
Did all patients with a positive CT-scan have surgery?	No	
Did all patients with a negative CT-scan have clinical follow-up?	No	
Was the choice of reference standard independent of the result of the index test?	No	
Were all patients included in the analyses?	Yes	
		High

Pakaneh 2008

Study characteristics	
Patient sampling	Patients with a clinical diagnosis of appendicitis were candidates for appendectomy. Patients with long-lasting abdominal pain, patients with suspected perforated appendix, and patients with unstable haemodynamics were excluded Recruitment period: May to July 2006
Patient characteristics and setting	Age range (median): 13 to 76 years (25). 26% women Department of Surgery, Imam Khomeini Hospital, Tehran, Iran. Single-centre study Disease spectrum: high suspicion of appendicitis
Index tests	Unenhanced helical CT of the lower abdomen. Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not stated. Manufacturer of CT-scanner, model name, and slice number: not stated
Target condition and reference standard(s)	Appendicitis. All patients had surgery with appendectomy and histological assessment of the removed appendix as the reference test
Flow and timing	50 patients were included; all had surgery, and 42 had appendicitis confirmed histologically
Comparative	
Criteria for CT diagnosis of appendicitis	Positive CT findings were defined as presence of at least 1 of the following: appendix diameter > 6 mm, periappendiceal fat stranding, appendicolith, periappendiceal free fluid, flegmone or abscess
Assessors of the CT-scan	
Notes	

Pakaneh 2008 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Low	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	Yes		

Pakaneh 2008 *(Continued)*

	Low

Park 2016

Study characteristics	
Patient sampling	Patients aged 15 to 44 years with suspected appendicitis were referred for CT. No exclusion criteria were stated. Recruitment period: June to December 2013
Patient characteristics and setting	Age range (mean): 15 to 44 (29.8). 60% women Radiology Department in Seoul, Korea. Single-centre study Disease spectrum: low or intermediate suspicion of appendicitis as assessed by the Alvarado score
Index tests	128-slice helical low-dose CT of the entire abdomen and pelvis with intravenous contrast enhancement (Brilliance iCT, Philips Healthcare). Slice thickness: 4 mm. Slice interval: 3 mm. Voltage: 120 kV. mAs product: 55 mAs. Effective dose: 2 mSv
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Patients who did not have an appendectomy were followed up with a telephone interview at least 3 months after the CT-scan
Flow and timing	168 patients were included and 61 were excluded due to standard-dose CT (n = 35); unavailable data (n = 24); no follow-up (n = 2). 107 patients had CT, 44 had an appendectomy, 42 had appendicitis confirmed histologically, and 63 received follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	No criteria were reported. Radiologists graded the likelihood of appendicitis on a 5-point Likert-like scale (1: definitely absent, 2: probably absent, 3: indeterminate, 4: probably present, 5: definitely present) A score of 3 or higher defined radiological evidence of appendicitis
Assessors of the CT-scan	Each of 7 radiologists examined CT-scans for 107 patients. Of these radiologists, 3 were fellowship-trained abdominal radiologists with 3 to 14 years' experience, 2 were board certified general radiologists with 2 to 3 years' experience, and 2 were radiology residents
Notes	Study data were collected prospectively during this pilot study for the LOCAT randomised trial. All trial procedures except randomisation were performed during the pilot study. The 7 radiologists assessed CT-scans in random order. These assessments were not used in clinical practice Furthermore, radiologists assessed images produced by 2 different reconstruction techniques: filtered back projection (FBP) and iterative reconstruction. As FBP is the conventional technology, we used the results for this technique in meta-analyses. Overall, 2×2 tables were constructed from appendicitis prevalence and mean values of sensitivity and specificity across the 7 radiologists

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Park 2016 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	Yes		

Park 2016 *(Continued)*

Were all patients included in the analyses?	Yes
	High

Pickuth 2001

Study characteristics	
Patient sampling	Patients with equivocal symptoms and signs of appendicitis were included over a 6-month period. Symptoms and signs were considered equivocal if they were insufficient for deciding whether to perform surgery or whether to discharge the patient. No exclusion criteria were reported Recruitment period: 6 months; otherwise not specified
Patient characteristics and setting	Age range (mean): 18 to 81 years (not reported). 53% women Radiology Department in Halle (Salle), Germany. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Unenhanced helical CT of the lower abdomen (Somatom Plus 4, Siemens; Tomoscan Aveo, Philips). Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not reported. If the appendix was not identified on the initial scan, then left lateral decubitus scanning with rectal contrast was performed. The number of patients who were examined with rectal contrast is not reported, hence the study was excluded from heterogeneity analyses of the effect of contrast enhancement
Target condition and reference standard(s)	Appendicitis. Surgical findings were reported for patients who had surgery without appendectomy. Histopathological reports were provided for patients who had an appendectomy. Patients who did not have surgery were followed up. Follow-up is not specified further
Flow and timing	120 patients were included; 86 patients with clinical unequivocal symptoms and signs of appendicitis were excluded. Appendicitis was diagnosed in 93 of the 120 patients. The number who had surgery and the number who had follow-up are not reported
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with periappendiceal inflammatory changes
Assessors of the CT-scan	Not reported
Notes	Results from this study are also reported in Pickuth 2000 The number of patients who were examined with rectal contrast is not reported. Study authors were contacted for subgroup results according to type of contrast enhancement. No reply was received, hence the study was excluded from heterogeneity analyses of the effect of contrast enhancement

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Pickuth 2001 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Unclear		
Were all patients included in the analyses?	Yes		
		High	

Platon 2009

Study characteristics	
Patient sampling	Consecutive adult (> 18 years) patients with suspected appendicitis presented to the Emergency Department during the daytime. Pregnant women were excluded
Patient characteristics and setting	Age range (median): 18 to 96 years (42.5); 52% women. Pregnant women were excluded Emergency Department in Geneva, Switzerland. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	4-slice CT of the entire abdomen (MX8000, Philips Medical Systems). Slice thickness: 5 mm. Slice interval: not stated. Voltage: 120 kV 2 CT-protocols were compared: Standard dose: tube current time product: 180 mAs. Enhancement by oral and intravenous contrast Low dose: tube current time product: 30 mAs. Enhancement by oral contrast For the low-dose protocol, effective doses were 1.2 ± 0.1 mSv for men; 1.7 ± 0.2 mSv for women
Target condition and reference standard(s)	Appendicitis was approached by intraoperative assessment. The proportions of patients with surgery and follow-up as reference standards are not stated. Follow-up consisted of recoding the definitive diagnosis in the discharge report
Flow and timing	86 patients were included in the study and in the analyses
Comparative	
Criteria for CT diagnosis of appendicitis	CT diagnosis of appendicitis was based on the following findings: appendix diameter > 6 mm, periappendiceal fat stranding, appendicolith, periappendiceal flegmone or abscess, periappendiceal free fluid, caecal wall thickening, arrowhead sign The relative importance and the logical combination of these findings are not stated Appendicitis was excluded when gas or contrast medium was depicted in the appendix lumen
Assessors of the CT-scan	Paired assessments by 2 experienced, board certified radiologists. Disagreements between radiologists were resolved by discussion
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Platon 2009 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Poortman 2003

Study characteristics	
Patient sampling	All patients presenting to the Emergency Department with symptoms and signs of appendicitis were included. Patients recruited between 10 pm and 8 am underwent CT the following morning Exclusion criteria: need for urgent surgery, pregnancy, claustrophobia Recruitment period: August 1998 to June 2000
Patient characteristics and setting	Age range (mean): 3 to 89 years (26). 6 (3%) patients were younger than 12 years. 55% women Emergency Department in a general community hospital in Tilburg, The Netherlands. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Unenhanced single-slice CT of the lower abdomen (Tomoscan AV, Philips). Slice thickness: 5 mm. Slice interval: 3 mm. Voltage: 120 kV. mAs product: 100 to 250 mAs, depending on patient age
Target condition and reference standard(s)	Appendicitis. Intraoperative visual assessment of the appendix was performed at laparoscopy - a normal looking appendix was considered uninflamed, and it was not resected. If the appendix was inflamed on visual assessment, it was resected. A normal looking appendix was resected for all patients who had open surgery with a muscle split laparotomy. The reference test was histological evaluation of the appendix in patients who had an appendectomy
Flow and timing	339 patients were screened for inclusion; 105 patients were excluded. Of the 234 included patients, 8 were excluded due to protocol violations and 27 were excluded because they did not have surgery, hence 199 patients were included from the analyses. All had surgery, and appendicitis was confirmed histologically in 132 patients
Comparative	
Criteria for CT diagnosis of appendicitis	Outer appendix diameter ≥ 6 mm. Ancillary signs of appendicitis including right lower quadrant inflammation, appendicoliths, and lymphadenopathy were recorded. CT findings were interpreted as negative if the appendix was visualised with intraluminal air If an appendix was not visualised and ancillary signs were or were not present, findings were interpreted as negative
Assessors of the CT-scan	2 body imaging radiologists and 10 other members of the radiology staff
Notes	Results from reassessment of CT-scans by 2 experienced body imaging radiologists reported in Poortman 2010 are not included in this review

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

Poortman 2003 *(Continued)*

Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	No		
		High	

Rao 1997

Study characteristics	
Patient sampling	Consecutive patients with suspected appendicitis were referred for CT examination of the appendix. Patients were referred from the Emergency Department or from private surgeons' offices. It is unclear whether all patients with clinically suspected appendicitis had CT. Pregnant women and patients younger than 6 years were excluded. Women with gynaecological abnormalities detected by pelvic ultrasonography were ineligible for the study Recruitment period: October 1995 to March 1996
Patient characteristics and setting	Age range: 6 to 84 years. The proportion of patients younger than 15 years is not reported. 54% women Department of Radiology, Boston, Massachusetts, USA. Single-centre study Disease spectrum: unclear
Index tests	Single-slice helical CT of the lower abdomen with oral and rectal contrast enhancement (HiSpeed Advantage, General Electric). Slice thickness and slice interval: 5 mm. Voltage and mAs product: not stated. 4 patients with a potential contraindication to rectal contrast received only oral contrast enhancement
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination in patients who had surgery with or without appendectomy. Follow-up in patients who did not have surgery. Follow-up consisted of at least 1 outpatient clinic visit and phone calls approximately 1 week, 1 month, and 3 months after the CT-scan
Flow and timing	103 patients were referred for CT examination. 2 patients declined to participate. Of the 101 included patients, 61 had surgery and 56 had appendicitis. Clinical follow-up was uneventful in 38 patients. 1 patient was lost to follow-up, and the final diagnosis was unclear in another. Both of these patients were excluded from the analyses, which included 99 patients
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with periappendiceal inflammatory changes such as fat stranding, fluid collection, phlegmon, or extraluminal gas
Assessors of the CT-scan	1 board certified radiologist
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Unclear	**High**
DOMAIN 2: Index Test All tests			

Rao 1997 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Rao 1998

Study characteristics	
Patient sampling	Patients with suspected appendicitis were referred for CT examination of the appendix. Patients were referred from the Emergency Department or from private surgeons' offices. 100 of 117 patients admitted with a principal diagnosis of appendicitis were referred for CT. It is unclear whether all patients with clinically suspected appendicitis had CT. Pregnant women, patients younger than 6 years, and patients

Rao 1998 *(Continued)*

	with a clinical contraindication to contrast material administered through the colon were excluded Recruitment period: July 1996 to November 1996
Patient characteristics and setting	Age range (mean): 6 to 75 (28) years (27% were paediatric patients). 57% women Department of Radiology, Boston, Massachusetts, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Single-slice helical CT of the lower abdomen with rectal contrast enhancement (HiSpeed Advantage, General Electric Medical Systems). Slice thickness and slice interval: 5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was documented for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. Follow-up included outpatient clinic visits and phone calls approximately 1 week and 2 months after CT-scan
Flow and timing	100 patients were included. No ineligible patients were referred for CT examination, and all referred patients agreed to participate. Surgery was performed in 59 patients, and 53 had appendicitis. Follow-up was performed in 41 patients - none were lost to follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with periappendiceal inflammatory changes such as fat stranding, fluid collection, phlegmon, or extraluminal gas. Appendicitis was diagnosed in cases with non-visualisation of the appendix only in the presence of specific CT signs of appendicitis, such as an appendicolith, focal caecal apical thickening, arrow head sign, or caecal bar sign. The appendix was considered normal if the appendiceal lumen filled completely with contrast material, air, or both, regardless of appendix diameter
Assessors of the CT-scan	3 board certified radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			

Rao 1998 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Rao 1999

Study characteristics	
Patient sampling	Consecutive female patients with 2 or more clinical signs or symptoms associated with appendicitis or acute gynaecological conditions who presented to the Emergency Department between October 1997 and March 1998 were included. No exclusion criteria were reported

Rao 1999 *(Continued)*

Patient characteristics and setting	Age range (mean): 11 to 63 (28) years (21% of patients were younger than 18 years). 100% women Emergency Department, Boston, Massachusetts, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Single-slice helical CT of the lower abdomen with rectal contrast enhancement (model name and manufacturer of CT-scanner not stated). Slice thickness and slice interval: 5 mm. Voltage and mAs product: not stated. In 14 patients, the scan included the entire abdomen; 2 patients also had intravenous contrast
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. Follow-up included outpatient clinic visits and phone calls at least 2 months after CT-scan
Flow and timing	100 patients were included, and all eligible patients agreed to participate. Surgery was performed in 41 patients; 32 had appendicitis. Follow-up was provided for 59 patients - none were lost to follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm with periappendiceal inflammatory changes. Appendicitis was diagnosed in cases with non-visualisation of the appendix only in the presence of specific CT signs of appendicitis, such as appendicolith, focal caecal apical thickening, arrow head sign, or caecal bar sign. The appendix was considered normal if the appendiceal lumen filled completely with contrast material, air, or both, regardless of appendix diameter
Assessors of the CT-scan	Residents or staff members of the Emergency Radiology Division
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		Unclear	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Rao 1999 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Repplinger 2015

Study characteristics	
Patient sampling	A convenience sample of patients older than 12 years of age who had CT ordered to evaluate for appendicitis. In this study, the accuracy of CT and MRI was compared via a paired design Exclusion criteria: contraindication to gadolinium-based contrast administration or MR imaging (metallic implants), inability to provide informed consent. The number of patients excluded due to MRI-related exclusion criteria is not stated Recruitment period: February 2012 to August 2014
Patient characteristics and setting	Age range (mean): 12 to 81 (31.5); 8 (4%) patients were younger than 15 years. 58% were women

Repplinger 2015 *(Continued)*

	Emergency Department in Madison, Wisconsin, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	64-slice helical CT of the entire abdomen and pelvis with intravenous and oral contrast enhancement (General Electric Healthcare, model name not reported) Slice thickness: 5 mm. Slice interval: 3 mm. Voltage: 100 to 140 kV. mAs product: 30 to 600 mAs
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Patients who did not have an appendectomy were followed up by a telephone interview 1 month after the visit to the Emergency Department
Flow and timing	210 patients were included; all had CT. 6 patients were excluded due to an incomplete MRI scan. Appendicitis was confirmed histologically in 64 patients. 6 patients with no follow-up clinic notes who could not be reached for telephone interview were counted as not appendicitis, hence 204 patients were included in the analysis
Comparative	
Criteria for CT diagnosis of appendicitis	Maximum short-axis width of the appendix, appendiceal wall thickening, fluid within the appendix lumen, presence of appendicolith, degree of periappendiceal inflammation Based on these criteria, the likelihood of appendicitis was rated on a 5-point scale (1: definitely not, 2: probably not, 3: possible, 4: probably, 5: definitely). A positive test result was a priori defined as a score ≥ 3
Assessors of the CT-scan	3 fellowship-trained abdominal radiologists independently interpreted all CT-scans. Based on the majority, 3 radiologist assessments were combined into an overall assessment, which was defined as the primary outcome
Notes	The study is reported in 2 conference abstracts. A submitted, unpublished study report was kindly provided by Michael D. Repplinger

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Repplinger 2015 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Sammalkorpi 2017

Study characteristics	
Patient sampling	Patients 16 years of age or older with suspected acute appendicitis who had a CT-scan were included. No exclusion criteria were reported
Patient characteristics and setting	Age range (mean): > 15 years; otherwise not specified. The proportion of women is not stated Emergency Department in Helsinki, Finland. Single-centre study

Sammalkorpi 2017 *(Continued)*

	Disease spectrum: any suspicion of appendicitis
Index tests	128-slice helical CT of the entire abdomen and pelvis with intravenous contrast enhancement (Somatom Definition AS+, Siemens Medical Systems). Slice thickness: 3 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: 110 mAs Effective doses of low-dose CT: 3.2 mSv in women; 2.6 mSv in men
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Follow-up by review of medical records was provided after a minimum of 1 month for patients who did not have an appendectomy
Flow and timing	1545 patients presented with suspected appendicitis. Of these, 489 had CT and 257 had appendicitis confirmed histologically. The number who had surgery, an appendectomy, or follow-up is not stated
Comparative	
Criteria for CT diagnosis of appendicitis	Appendiceal diameter > 6 mm with or without an appendicolith, appendiceal wall thickening, increased appendix wall enhancement, periappendiceal fat infiltration
Assessors of the CT-scan	Staff radiologists during working hours, thereafter radiological residents
Notes	An unspecified proportion of study participants had an ultrasound examination of the abdomen before CT

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		

Sammalkorpi 2017 *(Continued)*

		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Scott 2015

Study characteristics	
Patient sampling	Patients admitted with suspected appendicitis were referred for CT. The referral was provided at the discretion of the clinical team. Recruitment period: August 2012 to July 2013. Observational study - no exclusion criteria were applied
Patient characteristics and setting	Age range (median): 13 to 93 (46). 1 patient was younger than 15 years of age. 58% women. The proportion of patients younger than 15 years is not stated Department of General Surgery in West Middlesex, England. Single-centre study Disease spectrum: unclear
Index tests	16-slice or 128-slice CT of the entire abdomen with oral contrast material (Toshiba, Aquillion 16 or Aquillion 128). Slice thickness: 1 mm. Slice interval: 1 mm. Voltage: 120 kV. mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed in patients who had an appendectomy. Patients were classified as not having appendicitis if they did not require surgery, if a macroscopically normal appendix was found intraoperatively, or if the removed appendix was without a transmural neutrophilic infiltrate on histological examination. Follow-up for patients who did not have surgery consisted of checking for readmission for a minimum of 30 days after discharge

Scott 2015 *(Continued)*

Flow and timing	476 patients with suspected appendicitis were included. Among the 86 patients referred for CT, 39 had surgery and appendicitis was confirmed histologically in 34. Analyses include all 86 patients who had CT
Comparative	
Criteria for CT diagnosis of appendicitis	Radiologists were unaware of the study and followed standard practice when assessing the CT-scan
Assessors of the CT-scan	Radiologist on call
Notes	The primary aim of this study was to evaluate the accuracy of the Appendicitis Inflammatory Response Score in patients with suspected appendicitis. Only a subset of patients were referred for CT as part of the evaluation Additional information about this study was kindly provided by Alasdair Scott

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		High	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Scott 2015 *(Continued)*

		High	Low
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Sim 2013

Study characteristics	
Patient sampling	Consecutive patients underwent CT examination for suspected appendicitis because of acute right lower abdominal pain. It is unclear whether all patients with clinically suspected appendicitis had CT. Observational study - no exclusion criteria were applied Recruitment period: April 2011 to October 2011
Patient characteristics and setting	Age range (mean): 4 to 90 years (33) - 18% of patients were younger than 15 years of age. 55% women Radiology Department in Sungnam Si, Korea. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	16-slice and 64-slice CT of the entire abdomen with intravenous contrast enhancement (Brilliance 16, Philips Healthcare and Somatom Sensation, Siemens Healthcare). Transverse and coronal reformations were used. Slice thickness: 3 to 4 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery. Follow-up consisted of telephone calls in 161 and review of medical records in 322
Flow and timing	1012 patients were enrolled; 143 were subsequently withdrawn, hereof 59 due to alternative diagnosis and 49 due to loss to follow-up. Of the remaining 869 patients, 386 had surgery and 374 had appendicitis Of the 869 patients, 738 were 15 years of age or older; of these 320 had appendicitis
Comparative	

Sim 2013 *(Continued)*

Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm, thickened appendix wall, periappendiceal fat stranding, appendix wall hyperenhancement, extraluminal air adjacent to the appendix, caecal wall thickening, periappendiceal free fluid Definitive appendicitis when 3 or more criteria were present Probable appendicitis when 2 criteria were present Equivocal findings when 1 criteria was present
Assessors of the CT-scan	2 residents performed the initial CT assessments, which were subsequently checked by an expert abdominal radiologist
Notes	Study authors have provided 2×2 tables for the 738 patients who were 15 years of age or older. These results were used in the meta-analyses

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Sim 2013 *(Continued)*

		High	Low
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Stacher 1999

Study characteristics	
Patient sampling	Adult patients (> 18 years) presented with suspected appendicitis between December 1997 and December 1998. It is unclear whether all patients who had an appendectomy during the study period also had a CT-scan. Pregnant and breastfeeding women were excluded
Patient characteristics and setting	Mean age: 42 years. 41% women. Pregnant or nursing women and patients younger than 18 years were excluded Clinical setting unclear. Austria. Single-centre study Disease spectrum: unclear
Index tests	Unenhanced helical CT of the lower abdomen (Somatom Plus 4, Siemens). Slice thickness: 5 mm. Slice interval: 4 mm. Voltage: 140 kV. mAs product: 92.5 mAs. Multi-planar reconstructions were used if axial images were insufficient for definitive diagnosis
Target condition and reference standard(s)	Appendicitis. Of the 56 included patients, 32 had surgery with histological assessment of the removed appendix, and 24 had follow-up. Follow-up was performed after 2 months, but otherwise the concept of follow-up is not stated
Flow and timing	56 patients were included and analysed
Comparative	
Criteria for CT diagnosis of appendicitis	Positive CT findings were defined as a combination of the following features: appendix diameter > 6 mm; and periappendiceal fat stranding

Stacher 1999 *(Continued)*

Assessors of the CT-scan	3 experienced radiologists. Non-paired assessments were performed, but consensus assessments were used if 1 radiologist was in doubt about the diagnosis
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		

Stacher 1999 *(Continued)*

Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Tan 2015

Study characteristics			
Patient sampling	Patients with suspected appendicitis were referred for CT. The decision on inclusion and CT was made by the attending surgeon during the initial assessment No exclusion criteria were stated Recruitment period: August 2013 to March 2014		
Patient characteristics and setting	Age range (median): 15 to 82 years (33). 62% women General Surgery and Radiology Department in Singapore. Single-centre study Disease spectrum: mainly patients with intermediate suspicion of appendicitis		
Index tests	256-slice CT of the entire abdomen with intravenous contrast enhancement (iCT 256, Philips Healthcare). Slice thickness: 0,625 mm. Slice interval: not stated. Voltage: 120 kV. mAs product: up to 1000 mAs		
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery. Follow-up consisted of checking for readmission within 2 weeks after discharge		
Flow and timing	450 patients with suspected appendicitis were eligible for inclusion. Of these, 350 were included and all had CT. 168 patients had surgery, 155 had appendicitis confirmed histologically, and 182 had follow-up		
Comparative			
Criteria for CT diagnosis of appendicitis	Appendix diameter, any involvement of base, any gas pockets to suggest perforation, any fat stranding or periappendicular abscess		
Assessors of the CT-scan	Radiologist on duty		
Notes	Information about the CT-scanner and radiological criteria for appendicitis were provided by the corresponding author		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Tan 2015 *(Continued)*

Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Unclear	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Togawa 2005

Study characteristics

Togawa 2005 *(Continued)*

Patient sampling	Consecutive patients presented with rebound tenderness and muscular rigidity or guarding in the right lower quadrant No exclusion criteria were stated. Recruitment period: September 1999 to March 2001
Patient characteristics and setting	Age range (median): 15 to 86 years (45). 55% women. Kamitsuga General Hospital, Japan. Single-centre study Disease spectrum: high suspicion of appendicitis
Index tests	Unenhanced CT; unclear whether helical or not. Manufacturer of the CT-scanner and the model name were not reported Slice thickness, slice interval, voltage, mAs product: not reported
Target condition and reference standard(s)	It is unclear if the reference standard test is histological evaluation of the removed appendix or intraoperative findings of an inflamed appendix. It is unclear how patients who did not have surgery were followed up, but it is stated that none of these patients underwent delayed appendectomy
Flow and timing	Of 100 included patients, 86 had surgery and 58 had appendicitis. None of the 14 patients who were followed up had a delayed appendectomy. A diagnosis alternative to appendicitis was established in 28 patients
Comparative	
Criteria for CT diagnosis of appendicitis	Thickened appendix wall, periappendiceal high-dense fat tissue, or calcifications in the appendix
Assessors of the CT-scan	Unclear
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Togawa 2005 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Torbati 2003

Study characteristics	
Patient sampling	All patients presenting to the Emergency Department with equivocal symptoms and signs of appendicitis were studied with helical CT. Patients with clinically evident appendicitis were referred directly for surgical intervention. Pregnant women were excluded. Recruitment period: September1998 to March 2000
Patient characteristics and setting	Age range (mean): 7 to 75 (31) years (12% of patients were younger than 18 years). 56% women Emergency Department, USA. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Single-slice helical CT of the entire abdomen (model name and manufacturer of the CT-scanner were not stated). Slice thickness and slice interval: 5 mm. Voltage and mAs product: not stated. CT-scan was without contrast enhancement in 167 patients (71%), intravenous contrast was used in 33 patients (14%), and various combinations of intravenous, rectal, and oral contrast were used in the remaining patients (15%)

Torbati 2003 *(Continued)*

Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. Follow-up included telephone calls after 2 weeks to patients discharged from the Emergency Department and review of hospital courses for admitted patients
Flow and timing	310 patients were included. CT was performed in 250 patients; 60 were admitted to the surgical service without CT. Of the 250 patients who had CT, 51 had appendicitis - the total number who had surgery is not reported. CT-scans were equivocal in 17 patients, and 15 patients were lost to follow-up. These patients were excluded from analyses, which included 218 patients
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm was associated with an appendicolith or with periappendiceal inflammatory changes such as fat stranding, fluid collection, phlegmon, or extraluminal gas
Assessors of the CT-scan	Board certified radiologists
Notes	Patient characteristics (age and gender) stated above pertain to all 310 included patients. Results from this study are allocated to the unenhanced CT category, although 29% of patients received some kind of contrast enhancement

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Low	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**

Torbati 2003 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Tsai 2001

Study characteristics	
Patient sampling	A convenience sample of patients with suspected appendicitis and atypical clinical features were included. Patients with BMI ≥ 30 had CT; patients with BMI < 30 had ultrasonography. Patients younger than 16 years of age and pregnant women were excluded. Recruitment period: February 1999 to September 1999
Patient characteristics and setting	Adult patients (> 15 years of age) with suspected appendicitis and atypical clinical features were eligible Age range (mean): 16 to 84 years (41). 80% women. 30 participants were included Emergency Department and physician offices in Springfield, Missouri, USA. Single-centre study Diasease spectrum: intermediate suspicion of appendicitis
Index tests	Helical CT of the entire abdomen was performed with oral contrast enhancement. Rectal contrast was administered at the discretion of the radiologist to an undisclosed number of patients. Slice thickness: 5 mm. CT

Tsai 2001 *(Continued)*

	manufacturer and model name: not stated. Slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Of the 26 analysed patients, 4 had surgery with histological assessment of the removed appendix and 22 had follow-up. Follow-up consisted of telephone interviews with the primary care physician at least 3 months after initial presentation
Flow and timing	30 patients were included; 4 had surgery and all 4 had appendicitis. 26 patients were scheduled for follow-up; 4 were lost to follow-up and were excluded from analyses
Comparative	
Criteria for CT diagnosis of appendicitis	Positive CT findings were based on the following features: appendix diameter > 4 mm; periappendiceal fat stranding; appendicolith; periappendiceal flegmone or abscess; extraluminal air adjacent to the appendix; caecal wall thickening
Assessors of the CT-scan	Not stated
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	No		
		High	**High**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Unclear	**High**
DOMAIN 3: Reference Standard			

Tsai 2001 *(Continued)*

Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	No		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	Yes		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	No		
		High	

Uzunosmanoglu 2017

Study characteristics	
Patient sampling	Patients between 18 and 65 years of age with suspected appendicitis were enrolled. Pregnant women, patients who could not give consent for the study, and patients presenting before 8 am or after 5 pm were excluded Recruitment period: October 2012 to April 2013
Patient characteristics and setting	Age range (median): 19 to 61 years (mean age 30.3). 45% women Emergency Department in Ankara, Turkey. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	16-slice helical CT of the entire abdomen with intravenous contrast enhancement (Activion 16 Multislice CT, Toshiba). Slice thickness, slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Histological examination of the removed appendix was performed. All patients included in the analysis had an appendectomy
Flow and timing	92 patients were included; 32 of these were excluded due to missing data and lack of consent. 60 patients had CT and surgery with histological examination of the removed appendix
Comparative	
Criteria for CT diagnosis of appendicitis	None stated

Uzunosmanoglu 2017 *(Continued)*

Assessors of the CT-scan	Radiologist; not otherwise specified
Notes	This study compares the accuracy of CT, ultrasonography, and Doppler ultrasonography using a paired design. Patients included in the analyses had data for all included tests Study authors provided supplementary information about study design and the CT-scanner

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Did all patients with a positive CT-scan have surgery?	Yes		

Uzunosmanoglu 2017 *(Continued)*

Did all patients with a negative CT-scan have clinical follow-up?	No	
Was the choice of reference standard independent of the result of the index test?	Yes	
Were all patients included in the analyses?	Unclear	
		Unclear

Walker 2000

Study characteristics	
Patient sampling	Patients receiving general surgery consultation for appendicitis in an Emergency Department were eligible. If suspicion for appendicitis warranted either inpatient observation or operation, informed consent was obtained and the patient was randomised to receive CT or standard management. Exclusion criteria: age younger than 18 years, pregnancy, contraindication to instillation of rectal contrast material, appendiceal ultrasound performed before general surgical evaluation. No account of excluded patients was given. Recruitment period: July 1998 to June 1999
Patient characteristics and setting	Age range (mean): 18 to 77 (36) years. 66% women Emergency Department in Denver, Colorado, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Single-slice CT of the lower abdomen was enhanced with rectal contrast. Model name and manufacturer of the CT-scanner were not stated. Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. Follow-up consisted of a telephone call. The time period between CT-scan and the telephone call is not stated
Flow and timing	128 patients were included; 65 were allocated to CT. Of these, 39 had surgery and 35 had appendicitis confirmed histologically. CT was equivocal in 8 patients who were excluded from the analysis of accuracy. It is not reported whether all of the 26 patients who did not have surgery were followed up, but 57 patients were included in the analysis of accuracy
Comparative	
Criteria for CT diagnosis of appendicitis	CT-scans were categorised as positive, negative, or equivocal for appendicitis. CT-scans with appendiceal diameter > 6 mm without periappendiceal inflammatory changes were considered equivocal. Otherwise no specific criteria for diagnosis or exclusion of appendicitis were given to the radiology staff
Assessors of the CT-scan	Staff radiologists
Notes	
Methodological quality	

Walker 2000 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		

Walker 2000 *(Continued)*

Were all patients included in the analyses?	No
	High

Wang 2012

Study characteristics	
Patient sampling	Adult patients presenting to the Emergency Department with right lower quadrant pain, lower abdominal tenderness, and an Alvarado score of 4 to 7 were included. Patients younger than 18 years of age, pregnant women, and patients with known contrast allergy or reduced renal function were excluded. No account of excluded patients was given. Recruitment period: July and October 2010
Patient characteristics and setting	Age range: 18 years or older. 54% women Emergency Department at a tertiary hospital in Taoyuan, Taiwan. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	64-slice helical CT of the entire abdomen and pelvis with intravenous contrast enhancement (General Electric Healthcare; model name not available). Slice thickness: 5 mm. Slice interval, voltage, and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Histology was performed in patients who had had surgery with appendectomy; follow-up was provided after 2 weeks for patients who did not. No further description of follow-up was provided
Flow and timing	Of 60 included patients, 26 had appendicitis confirmed histologically. The number who had surgery is not stated. One patient was lost to follow-up. It is unclear whether this patient was included in the analysis
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm and pericaecal fat stranding
Assessors of the CT-scan	Not stated
Notes	The corresponding author provided information about the CT scanner, the CT-protocol, criteria for the CT diagnosis of appendicitis, and numbers of true-positives, false-positives, false-negatives, and true-negatives

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		

Wang 2012 *(Continued)*

Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Weltman 2000

Study characteristics	
Patient sampling	Consecutive patients aged 3 years or older suspected of having appendicitis underwent CT of the abdomen and pelvis. It is unclear whether all patients with suspected appen-

Weltman 2000 *(Continued)*

	dicitis were evaluated with CT. Exclusion criteria and recruitment period were not stated. No account of exclusions was given
Patient characteristics and setting	Age range: 3 to 73 years (mean age 34). The percentage of patients younger than 15 years of age is not stated. 46% women Radiology Department in East Meadows, New York, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	Helical single-slice CT of the abdomen and pelvis (XPress/SX, Toshiba). Contrast enhancement: 92% of participants had rectal contrast, 60% had IV contrast, and 2% had oral contrast. Slice thickness: 5 mm and 10 mm. Slice interval: 5 mm. Voltage and mAs product: not stated. Study authors compared the accuracy of CT with 5-mm and 10-mm slice thickness. Results for 5-mm slice thickness are included in the meta-analyses
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not. Follow-up for the 51 patients who did not have surgery consisted of outpatient visits for 30 patients 1 to 2 months after CT and telephone calls for 21 patients
Flow and timing	103 patients were enrolled and all had had CT. 3 patients were subsequently withdrawn because symptoms resolved after antibiotic treatment. Surgery was performed in 49 patients, 48 had appendicitis, and 51 patients received follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Confidence in the radiological diagnosis of appendicitis was graded from 1 to 3: 1: > 85% certainty - abnormal appendix or appendicolith associated with periappendiceal inflammatory changes 2: 40% to 85% certainty - right lower quadrant inflammatory changes, abscess, caecal wall thickening 3: < 40% certainty - cannot rule out appendicitis due to equivocal, but potentially abnormal, findings Patients with grade 1 to 3 were considered CT-positive in the accuracy analyses
Assessors of the CT-scan	All CT-scans were reevaluated for the study by 2 fellowship-trained body imaging attending physicians with several years' experience in interpretation of CT-scans for appendicitis
Notes	Study authors were contacted for additional data. No response was received

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		

Weltman 2000 *(Continued)*

Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Wijetunga 2001

Study characteristics	
Patient sampling	Consecutive patients with equivocal symptoms and clinical signs of appendicitis were included. Patients with characteristic symptoms and signs of appendicitis were excluded (central or right iliac fossa pain, vomiting, fever, increased white blood cell count, signs of peritonitis in the right lower quadrant). No other exclusion criteria are reported. No account of exclusions is given
Patient characteristics and setting	Age range (mean): 14 to 81 years (31). 56% women Emergency Department in Sydney, Australia. Single-centre study Disease spectrum: intermediate suspicion of appendicitis
Index tests	Helical single-slice CT of the lower abdomen with oral contrast enhancement (Twin Flash V. 3.3, Elscint). Slice thickness and slice interval: 5 mm. Voltage and mAs product: not stated
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported for patients who had surgery with or without appendectomy. Follow-up was provided for patients who did not have surgery. Follow-up included review of hospital notes and telephone calls to patients within 1 to 8 months after CT
Flow and timing	105 patients were included; all had CT, 34 had surgery, and 30 had appendicitis. Follow-up was provided for 66 patients, and 5 patients were lost to follow-up. Patients lost to follow-up were excluded from analyses
Comparative	
Criteria for CT diagnosis of appendicitis	CT-scans were interpreted as positive for appendicitis if 3 or more of the following criteria were present: • Maximum appendix diameter > 6 mm • No contrast material in the appendiceal lumen • Periappendicular inflammatory changes such as fat stranding, fluid collection, phlegmon, or extraluminal gas • Appendicolith • Thickening of the caecal wall (focal thickening, arrowhead sign, caecal bar sign)
Assessors of the CT-scan	Specialist radiologists
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		

Wijetunga 2001 *(Continued)*

Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Unclear		
Were all patients included in the analyses?	Yes		
		High	

Wilson 2001

Study characteristics

Wilson 2001 *(Continued)*

Patient sampling	All patients receiving a surgery consultation for acute appendicitis were eligible. Patients were excluded if they refused consent, had diffuse peritonitis, were physiologically compromised requiring immediate operation, were pregnant, or were nursing. Recruitment period: not stated
Patient characteristics and setting	Age range (mean): 4 to 81 (27) years. The proportion of patients younger than 15 years is not reported. 49% women Emergency Department, USA. Single-centre study Disease spectrum: any suspicion of appendicitis
Index tests	CT of the lower abdomen with rectal contrast material. Not otherwise specified
Target condition and reference standard(s)	Appendicitis. Intraoperative findings or histological examination was reported in patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up with telephone calls after 1 day and 7 days
Flow and timing	104 patients were evaluated for study enrolment; 99 were included. 50 patients had appendicitis. The numbers that had surgery and follow-up are not stated
Comparative	
Criteria for CT diagnosis of appendicitis	Not reported
Assessors of the CT-scan	Experienced resident or staff radiologists
Notes	CT interpretation was equivocal in 28 patients, of whom 15 had appendicitis. In the extracted 2×2 table, we counted patients with equivocal CT as CT-positive because this is how patients with equivocal CT evaluations were counted in the analyses in most included studies

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Wilson 2001 *(Continued)*

Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Wise 2001

Study characteristics	
Patient sampling	Patients presented to the Emergency Department or outpatient clinic with appendicitis in the first 3 of the differential diagnoses. Exclusion criteria: age younger than 19 years, pregnancy, history of allergy to intravenous contrast material, immediate surgery needed Recruitment period: March 1998 to June 1999
Patient characteristics and setting	Age range (mean): 18 to 86 years (38). 74% women Radiology Department in Hershey, Pennsylvania, USA. Single-centre study Disease spectrum: unclear
Index tests	Single-slice helical CT (PQ 5000, Picker International). Slice thickness: 4 mm. Slice interval: not stated. Voltage: 100 to 200 kV. mAs product: 200 to 300 mAs. Patients were randomised to 2 predetermined sequences of contrast enhancement: Group A

Wise 2001 *(Continued)*

	First: CT of the lower abdomen with oral contrast material Second: CT of the entire abdomen and pelvis with oral and intravenous contrast material Third: CT of the lower abdomen with oral, intravenous, and rectal contrast material Group B First: CT of the lower abdomen with oral contrast material Second: CT of the lower abdomen with oral and rectal contrast material Third: CT of the entire abdomen and pelvis with oral, rectal, and intravenous contrast material All patients also had graded compression ultrasonography of the right lower quadrant performed by dedicated sonographers or sonologists The radiologist on duty initially interpreted all CT and sonographic studies as a unit, with the overall interpretation used for clinical treatment. At a later date, 4 observers independently interpreted each of the CT-scans in random order
Target condition and reference standard(s)	Appendicitis. Surgical reports and histopathological reports were provided for patients who had surgery with or without appendectomy. Patients who did not have surgery were followed up with telephone calls 1 week, 1 month, and 3 months after presentation
Flow and timing	149 patients were eligible; 49 were excluded. Of the 100 included patients, 24 had appendicitis. The number that had surgery is not stated, and there is no account of the completeness of follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Each assessor graded the confidence in the radiological diagnosis of appendicitis on a 10-point scale. No morphological criteria for the radiological diagnosis nor for a threshold for positivity on the 10-point scale are reported
Assessors of the CT-scan	3 fellowship-trained radiologists and 1 third year radiology resident
Notes	It was feasible to extract a 2×2 table only for CT of the lower abdomen with oral contrast material. This 2×2 table represents an average for the 4 observers. Results of the initial evaluation by the radiologist on duty are ignored because they may incorporate the outcome of graded compression ultrasonography of the right lower quadrant

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**

Wise 2001 *(Continued)*

DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it prespecified?	No		
Is the index test described in sufficient detail to permit its replication?	Yes		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	No		
		High	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Unclear		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	Unclear		
Did all patients with a negative CT-scan have clinical follow-up?	Unclear		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

Wong 2002

Study characteristics	
Patient sampling	Patients suspected of having appendicitis and scheduled for surgery were recruited. Patients who were pregnant, who were younger than 16 years, or who could not have contrast medium administered via the rectum were excluded. Recruitment period: not stated
Patient characteristics and setting	Age range: 16 years or older. 42% women. 50 participants were included Setting: hospital in Singapore - otherwise unclear. Single-centre study Disease spectrum: high suspicion of appendicitis
Index tests	1-slice helical CT of lower abdomen and pelvis with rectally administered colonic contrast material (CT-X Vision, Toshiba) Slice thickness: 5 mm. Slice interval: 5 mm. Voltage and mAs product: not stated. Additional reconstruction of the axial images to 1-mm slice interval was done to identify the appendix if there were difficulties locating it from the initial CT images
Target condition and reference standard(s)	Appendicitis. Histological examination of the removed appendix was performed - all patients had surgery with appendectomy
Flow and timing	50 patients were included; all had CT. Surgery was performed in all patients; 37 had appendicitis
Comparative	
Criteria for CT diagnosis of appendicitis	If the appendix was visualised: external appendix diameter > 6 mm and/or periappendiceal inflammatory changes (fat stranding, fluid collection, or enlarged mesenteric nodes) If the appendix was not visualised: appendicolith, caecal apical wall thickening, arrowhead sign, or caecal bar sign The appendix was considered normal if the lumen was completely filled with air, contrast material, or both
Assessors of the CT-scan	Not stated
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Yes		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**

Wong 2002 *(Continued)*

DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Unclear		
		Low	**High**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**High**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Did all patients with a positive CT-scan have surgery?	Yes		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	Yes		
Were all patients included in the analyses?	Yes		
		Low	

Yuksekkaya 2004

Study characteristics	
Patient sampling	Patients with suspected appendicitis were included. Patients younger than 14 years of age and pregnant women were excluded
Patient characteristics and setting	Age range: 14 to 62 years. 52% women Emergency Department in Turkey

Yuksekkaya 2004 *(Continued)*

	Disease spectrum: any suspicion of appendicitis
Index tests	Unenhanced single-slice helical CT of the lower abdomen (General Electric, ProSpeed S) Slice thickness and slice interval: 5 mm. Voltage: 120 kV. mAs product: not reported
Target condition and reference standard(s)	Appendicitis. Histological examination was performed in patients who had an appendectomy; follow-up was provided for patients who did not have surgery. Follow-up consisted of monitoring readmission with appendectomy within 3 months
Flow and timing	65 patients were included; all had CT. 37 patients had surgery; 27 had appendicitis confirmed by histology and 28 patients received follow-up
Comparative	
Criteria for CT diagnosis of appendicitis	Appendix diameter > 6 mm and periappendiceal stranding
Assessors of the CT-scan	2 radiologists
Notes	This study is reported in Turkish. We are grateful to Dr. Fatma Kara for extracting data from this study

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Does the study population represent an unselected sample of adults with suspected appendicitis?	Unclear		
		Unclear	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Is the index test described in sufficient detail to permit its replication?	No		
Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?	Yes		
		Low	**Unclear**

Yuksekkaya 2004 *(Continued)*

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Did all patients with a positive CT-scan have surgery?	No		
Did all patients with a negative CT-scan have clinical follow-up?	No		
Was the choice of reference standard independent of the result of the index test?	No		
Were all patients included in the analyses?	Yes		
		High	

BMI: body mass index.
CT: computed tomography.
ED: Emergency Department.
IV: intravenous.
MRI: magnetic resonance imaging.
SD: standard deviation.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abo 2011	Study in paediatric population
Al-Faouri 2016	Extraction of 2×2 table not possible. Study authors contacted by email; no reply received
Ali 2017	Target disorder was perforated appendicitis
Anderson 2009	Study in patients with suspected appendicitis, diverticulitis, or small bowel obstruction
Antevil 2004	Retrospective data collection
Bachar 2013	Retrospective data collection
Balthazar 1998	Retrospective data collection
Bendeck 2002	Retrospective data collection
Bixby 2006	Target disorder was perforated appendicitis

Study	Reason for exclusion
Brandt 2003	Retrospective data collection
Brontvein 2002	Retrospective data collection
Caglayan 2010	Retrospective data collection
Callahan 2015	Study in paediatric population
Castro 2001	Study in paediatric population
Ceydeli 2006	Retrospective data collection
Chang 2013	Retrospective data collection
Chang 2016	Retrospective data collection
Chen 2010	Retrospective data collection
Chen 2016	Retrospective data collection
Chiu 2012	Retrospective data collection
Cho 1999	Retrospective data collection
Choi 1998	Retrospective data collection
Cuschieri 2008	Retrospective data collection
Davis 2017	Study in paediatric population
Dearing 2008	Retrospective data collection
Debnath 2015	Inconclusive ultrasonography performed before CT for all participants
Deneuville 1995	Retrospective data collection
Dibble 2016	Study in paediatric population
Didier 2015	Study in paediatric population
Dillman 2016	Study in paediatric population
Elikashvili 2014	Study in paediatric population. Inconclusive ultrasonography performed before CT for all participants
Fefferman 2001	Study in paediatric population
Fefferman 2005	Study in paediatric population
Foley 2005	Target disorder was perforated appendicitis
Fraser 2010	Study in paediatric population
Fuchs 2002	Retrospective data collection

Study	Reason for exclusion
Gaitini 2008	Inconclusive ultrasonography performed before CT for all participants
Garcia Pena 1999	Study in paediatric population
Gaskill 2016	Target disorder was perforated appendicitis
Giuliano 2004	Extraction of 2×2 table not possible. Study authors contacted by email; no reply was received
Gwynn 2001	Retrospective data collection
Hill 2010	Study in paediatric population
Hoecker 2005	Study in paediatric population
Hookman 2000	Study in paediatric population
Horrow 2003	Target disorder was perforated appendicitis
Huynh 2007	Retrospective data collection
Ives 2008	Retrospective data collection
Iwahashi 2005	Retrospective data collection
Iwama 2002	Retrospective data collection
Jabra 1997	Study in paediatric population
Johansson 2007	Retrospective data collection
Kahn 2013	Case-control design
Kailidou 2006	Retrospective data collection
Kaiser 2002	Study in paediatric population
Kaiser 2004	Study in paediatric population
Kamel 2000	Retrospective data collection
Karakas 2000	Study in paediatric population
Kene 2016	Study in paediatric population
Kharbanda 2007	Study in paediatric population
Kilincer 2017	Excluded for other reasons This study compares appendiceal outer diameter, wall thickness, and enteric contrast filling of the appendix in 2 separate groups In group A, participants were prospectively recruited and underwent CT with external compression to the right lower quadrant Group B was an age- and sex-matched historical control group of persons who had CT for a presumptive clinical diagnosis of appendicitis This study was excluded because the CT-protocol in group A was experimental and out of scope with respect to the review question

Study	Reason for exclusion
	In group B, data collection was retrospective
Kim 2008a	Retrospective data collection
Kim 2011	Retrospective data collection
Kimura 2016	Retrospective data collection
Latifi 2011	Retrospective data collection
Lazarus 2007	Study in population of pregnant women
Lee 2001	Retrospective data collection
Lee 2006	Retrospective data collection
Lee 2016	Study in paediatric population
Leeuwenburgh 2013	Inconclusive ultrasonography performed before CT for all participants
Lin 2016	Study in paediatric population
Liu 2015	Retrospective data collection
Lowe 2000	Study in paediatric population
Lowe 2001	Study in paediatric population
Lowe 2001a	Study in paediatric population
Lu 2007	Retrospective data collection
McDonough 2002	Retrospective data collection
Miki 2005	Case-control design
Mizuo 1999	Retrospective data collection
Morris 2002	Retrospective data collection
Moteki 2007	Case-control design
Moteki 2011	Retrospective data collection
Mullins 2001	Study in population of pregnant women
Mullins 2001a	Study in paediatric population
Mun 2006	Retrospective data collection
Naeger 2011	Case-control design
Naffaa 2005	Retrospective data collection
Naoum 2002	Retrospective data collection

Study	Reason for exclusion
Ng 2007	Case-control design
Oliak 1999	Target disorder was perforated appendicitis
Ozkan 2014	Retrospective data collection
Park 2013	Retrospective data collection
Partrick 2003	Study in paediatric population
Paulson 2005	Retrospective data collection
Peck 2000	Retrospective data collection
Pena 1999	Study in paediatric population
Pena 2002	Study in paediatric population. Inconclusive ultrasonography performed before CT for all participants
Perez 2003	Retrospective data collection
Pickhardt 2011	Retrospective data collection
Poh 2004	Retrospective data collection
Poletti 2011	Inconclusive ultrasonography performed before CT for all participants
Pooler 2012	Retrospective data collection
Poortman 2009	Inconclusive ultrasonography performed before CT for all participants
Ramalingam 2016	Retrospective data collection
Raman 2002	Retrospective data collection
Ramarajan 2009	Study in paediatric population
Rao 1997b	Other reason. This publication reports the sensitivity and specificity of individual CT signs for appendicitis (e.g. fat stranding, enlarged unopacified appendix, adenopathy). No estimates are presented for the overall assessment that integrates all signs. Moreover, hospital and recruitment period overlap with other studies reported by the same trial author
Reeve 2010	Retrospective data collection
Reich 2011	Retrospective data collection
Rhea 2005	Retrospective data collection
Rosengren 2004	Retrospective data collection
Sakai 2007	Retrospective data collection
Schuler 1998	Retrospective data collection
Seo 2009	Retrospective data collection

Study	Reason for exclusion
Siddiqui 2007	Target disorder was perforated appendicitis
Sivit 2000	Study in paediatric population
Sovtsov 2017	Retrospective data collection
Srinivasan 2010	Study in paediatric population
Stabile 2010	Case-control design
Stephen 2003	Study in paediatric population
Stromberg 2007	Participants recruited with abdominal pain at any location
Styrud 2000	Retrospective data collection
Suh 2011	Target disorder was perforated appendicitis
Suthikeeree 2010	Target disorder was perforated appendicitis
Tamburrini 2007	Retrospective data collection
Tan 2013	Retrospective data collection
Tatar 2016	Retrospective data collection
Teo 2000	Study in paediatric population. Inconclusive ultrasonography performed before CT for all participants
Toorenvliet 2010	Inconclusive ultrasonography performed before CT for all participants
Ujiki 2002	Retrospective data collection
Unlu 2009	Inconclusive ultrasonography performed before CT for all participants
Uyeda 2015	Participants recruited with abdominal pain at any location
Vajtai 2013	Study in paediatric population
Van Randen 2011	Study focuses on the accuracy of CT for diagnosing causes of abdominal pain in general. Study participants were not included due to a particular suspicion of appendicitis
Wadhwani 2016	Retrospective data collection
Wallace 2008	Study in population of pregnant women
Westerland 2016	Retrospective data collection
Weyant 2000	Retrospective data collection
Weyant 2001	Study in paediatric population
Yeoh 2016	Retrospective data collection
Yi 2017	Study in paediatric population

Study	Reason for exclusion
Yoo 2009	Study in paediatric population
Yun 2016	Retrospective data collection
Zourob 2016	Retrospective data collection

CT: computed tomography.

Characteristics of studies awaiting classification *[ordered by study ID]*

D'Ippolito 1998

Study characteristics	
Patient sampling	52 patients with clinical signs of acute appendicitis who underwent surgery from September 1993 to March 1995 were included
Patient characteristics and setting	Women 52%. Mean age 29 years (range 6 to 64 years). Setting not stated
Index tests	Unenhanced CT of the lower abdomen
Target condition and reference standard(s)	Acute appendicitis. Surgical findings and histopathology were used as reference tests
Flow and timing	52 participants were included; all had surgery and all were included in the analyses. 44 (85%) participants had appendicitis
Comparative	No
Notes	Unclear if data collection was prospective or retrospective

Ege 2002

Study characteristics	
Patient sampling	Adult patients who consulted general surgeons and were suspected to have acute appendicitis between January 1998 and December 2000 were included
Patient characteristics and setting	Women 37%. Mean age 25 years (range 16 to 69 years). Setting not stated
Index tests	Unenhanced CT of the lower abdomen
Target condition and reference standard(s)	Acute appendicitis. Histopathological analysis of resected appendices served as the reference test for the diagnosis of appendicitis. If surgery was not performed, clinical follow-up was obtained. If no surgery was undertaken and the patient's symptoms had resolved, this was recorded as a true-negative finding
Flow and timing	296 participants were included in the study and in the analyses. 123 (42%) participants had surgery; 108 (36%) had appendicitis
Comparative	No

Ege 2002 *(Continued)*

Notes	Unclear if data collection was prospective or retrospective

Elghany 2011

Study characteristics	
Patient sampling	Patients presenting with pain on the right side of the abdomen between 2009 and 2010
Patient characteristics and setting	Women 56%. Mean age 38 years (range 16 to 81 years). Emergency Department in Cairo, Egypt
Index tests	Abdominopelvic CT with IV and oral contrast material
Target condition and reference standard(s)	Acute appendicitis. Histopathological analysis of the removed appendix in participants who had appendicectomy. Follow-up by review of medical charts for participants who did not have appendicectomy
Flow and timing	63 participants were included in the study and in the analyses. 37 (59%) participants had appendicitis
Comparative	Ultrasonography in all participants
Notes	Unclear if data collection was prospective or retrospective

Lane 1997

Study characteristics	
Patient sampling	Adult patients (> 17 years) with suspected acute appendicitis referred for CT between September 1994 and December 1995
Patient characteristics and setting	Women 47%. Mean age and age range not stated. Setting not stated
Index tests	Unenhanced abdominopelvic CT
Target condition and reference standard(s)	Acute appendicitis. Histopathological analysis of the removed appendix in participants who had appendicectomy. Clinical follow-up for participants who did not have appendicectomy
Flow and timing	109 participants were included in the study and in the analyses. 37 (34%) participants had appendicitis
Comparative	No
Notes	Unclear if study participants are also included in Lane 1999. Study author contacted by email. No reply received

Lietzen 2018

Study characteristics	

Lietzen 2018 *(Continued)*

Patient sampling	Patients with suspected acute appendicitis recruited for the APPAC trial comparing surgery vs antibiotic therapy for treatment of uncomplicated acute appendicitis. APPAC participants allocated to antibiotic treatment were excluded. Patients were enrolled from November 2009 to June 2012
Patient characteristics and setting	Women 44%. Mean age 38 years (range 17 to 65 years). Emergency Departments at 6 hospitals in Finland
Index tests	Abdominopelvic CT with IV contrast
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix in participants who had surgery. Follow-up with review of medical records for participants who did not have surgery
Flow and timing	1065 participants were included in the study and in the analyses. 714 (67%) participants had acute appendicitis
Comparative	No
Notes	This study was identified by a search update during the editorial process of the review

LOCAT Group 2018

Study characteristics	
Patient sampling	Patients aged 15 to 44 years who were referred for IV contrast-enhanced CT because of suspected acute appendicitis. Participants were randomised to receive low-dose or standard-dose CT
Patient characteristics and setting	Low-dose group: women 55%, median age 28 years (interquartile range 21 to 35 years) Standard-dose group: women 54%, median age 28 years (interquartile range 21 to 35 years) Emergency and Radiology Departments at 20 centres in Korea
Index tests	Low-dose (2 mSv) and standard-dose (3 to 8 mSv) abdominopelvic CT with IV contrast
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix in participants who had surgery. Follow-up based on medical records and telephone interviews after 3 months for participants who did not have surgery
Flow and timing	1535 participants were allocated to low-dose CT; 1539 were allocated to standard-dose CT Low-dose CT: 1459 participants were included in the analyses; 76 were excluded due to an incomplete reference standard. Appendicitis was confirmed in 524 (36%) participants Standard-dose CT: 1429 participants were included in the analyses; 110 were excluded due to an incomplete reference standard. Appendicitis was confirmed in 564 (39%) participants
Comparative	Accuracy of low-dose and standard-dose CT was compared in a non-inferiority, multi-centre randomised trial
Notes	This study was identified by a search update during the editorial process of the review

Park 2018

Study characteristics	
Patient sampling	Young adults (18 to 44 years of age) who underwent CT for suspected appendicitis from August to October 2015
Patient characteristics and setting	Women 57%. Mean age 27 years (age range not stated). Emergency Department at a tertiary hospital in Seoul, Korea
Index tests	Low-dose abdominopelvic CT with IV contrast. Assessments by 6 radiologists
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix for participants who had surgery. Follow-up based on medical records and standardised telephone interviews after 3 months for participants who did not have surgery
Flow and timing	57 patients were eligible; 30 were included in the study and in the analyses. 9 (30%) participants had appendicitis
Comparative	Step-wise comparisons of 1.5-, 1.0-, and 0.5-mSv low-dose CT vs 2.0-mSv low-dose CT using a paired non-inferiority design
Notes	This study was identified by a search update during the editorial process of the review

Rao 1996

Study characteristics	
Patient sampling	Consecutive patients with clinically suspected acute appendicitis referred for CT
Patient characteristics and setting	Gender and age distributions: not stated. Departments of Emergency and Radiology in Boston, Massachusetts, USA
Index tests	CT of the lower abdomen with oral and rectal contrast medium
Target condition and reference standard(s)	Acute appendicitis. Histopathological analysis of the removed appendix in participants who had appendicectomy. Clinical follow-up in participants who did not have appendicectomy
Flow and timing	35 participants were included in the study and in the analyses. 17 (49%) participants had appendicitis
Comparative	No
Notes	Unclear if the 35 persons included in this study were also included in later studies by the same trial author

Singh 2018

Study characteristics	
Patient sampling	Patients older than 18 years of age presenting with acute right iliac fossa pain and suspected of acute appendicitis were enrolled in the study from October 2014 to September 2016

Singh 2018 *(Continued)*

Patient characteristics and setting	Women 47%. Mean age 37 years (range 18 to 85 years). Departments of Radiodiagnosis and Surgery in Imphal, Manipur, India
Index tests	Low-dose unenhanced abdominopelvic CT
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix in participants who had surgery. Follow-up based on medical records and telephone interviews after 1 month for participants who did not have surgery
Flow and timing	83 participants were included in the study and in the analyses. 56 (67%) participants had appendicitis
Comparative	No
Notes	This study was identified by a search update during the editorial process of the review

Sippola 2018

Study characteristics	
Patient sampling	Patients between 18 and 60 years of age admitted with clinical suspicion of acute appendicitis
Patient characteristics and setting	Women 46%. Mean age 33 years (age range not stated). Emergency Department in Turku, Finland
Index tests	Low-dose and standard-dose abdominopelvic CT with IV contrast
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix in participants who had surgery. Follow-up after 6 months for participants who did not have surgery
Flow and timing	60 participants were recruited and 3 were excluded. Low-dose and standard-dose CT was available for 57 and 55 participants, respectively. 55 participants were included in the comparative analyses. 49 (86%) had appendectomy with histological confirmation of the diagnosis
Comparative	Accuracy of low-dose and standard-dose treatment was compared in a paired design. The order of low-dose and standard-dose CT was randomised
Notes	This study was identified by a search update during the editorial process of the review

Stroman 1999

Study characteristics	
Patient sampling	Over a 12-month period (1 December 1997 to 1 December 1998), 107 patients with suspected acute appendicitis, but with equivocal symptoms, underwent CT
Patient characteristics and setting	Women 59%. Median age 33 years (range 13 to 89 years). Setting not stated

Stroman 1999 *(Continued)*

Index tests	Abdominopelvic CT with intravenous and oral contrast material
Target condition and reference standard(s)	Acute appendicitis. Histopathological analysis of the removed appendix in participants who had appendicectomy. Follow-up by review of medical charts for participants who did not have appendicectomy
Flow and timing	107 participants were included; all were included in the analyses. 36 (34%) underwent appendicectomy with histological confirmation of the diagnosis
Comparative	Ultrasonography was performed in 43 participants
Notes	Unclear if data collection was prospective or retrospective. Study author contacted by email. No reply received

Yang 2016

Study characteristics	
Patient sampling	Patients suspected of acute appendicitis
Patient characteristics and setting	Not stated
Index tests	Low-dose CT
Target condition and reference standard(s)	Acute appendicitis. Histological analysis of the removed appendix in all participants
Flow and timing	Not stated
Comparative	No
Notes	Unclear if data collection was prospective or retrospective. Study authors contacted by email. No reply received

Yetkin 2002

Study characteristics	
Patient sampling	Patients with equivocal clinical findings suggesting acute appendicitis who had CT were included
Patient characteristics and setting	Women 40%. Median age 26 years (range 17 to 64 years). Setting not stated
Index tests	Unenhanced CT of the lower abdomen
Target condition and reference standard(s)	Acute appendicitis. Operative findings and histopathological analysis of the removed appendix in participants who had surgery. Participants who did not have surgery were followed up clinically
Flow and timing	65 participants were included in the study and in the analyses. 48 (74%) participants had surgery; 45 (69%) had appendicitis
Comparative	No
Notes	Unclear if data collection was prospective or retrospective

Yildirim 2008

Study characteristics	
Patient sampling	Patients with abdominal pain who presented to the Emergency Department between June 2003 and February 2006
Patient characteristics and setting	Women 45%. Mean age 34 years (range 18 to 76 years). Emergency Department in Ankara, Turkey
Index tests	Abdominopelvic CT with intravenous and oral contrast material. 48 participants did not tolerate oral contrast material
Target condition and reference standard(s)	Acute appendicitis. Histological analysis of the removed appendix in all participants
Flow and timing	143 participants were included in the study and in the analyses. All had appendicectomy; 130 (91%) had appendicitis
Comparative	No
Notes	Unclear if data collection was prospective or retrospective. Study authors contacted by email. No reply received

CT: computed tomography.
IV: intravenous.

DATA

Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 CT (unenhanced)	19	2140
2 CT (IV contrast)	17	4265
3 CT (oral contrast)	7	673
4 CT (rectal contrast)	9	1098
5 CT (IV+oral contrast)	15	2074
6 CT (oral+rectal contrast)	3	230
7 CT (IV+oral+rectal contrast)	2	152
8 Low-dose CT	7	1445
9 CT (overall)	64	10380

Test	No. of studies	No. of participants
10 Standard-dose CT	61	9292

Test 1. CT (unenhanced).

Test 2. CT (IV contrast).

Test 3. CT (oral contrast).

Test 4. CT (rectal contrast).

Test 5. CT (IV+oral contrast).

Test 6. CT (oral+rectal contrast).

Test 7. CT (IV+oral+rectal contrast).

Test 8. Low-dose CT.

Test 9. CT (overall).

Test 10. Standard-dose CT.

ADDITIONAL TABLES

Table 1. Components of CT-protocols in the 64 included studies

CT-protocol components	Number of studies

Table 1. Components of CT-protocols in the 64 included studies *(Continued)*

Slice thickness (mm)	6
0.6-2.9	9
3.0-4.9	36
5.0-6.9	4
7.0-10.0	9
not stated	
Slice interval (mm)	6
0.6-2.9	5
3.0-4.9	16
5.0	1
10.0	36
not stated	
Voltage (kV)	21
120	4
140	1
200	38
not stated	
mAs product (mAs)	4
30-100	5
100-199	5
200-299	4
≥ 300	46
not stated	

CT: computed tomography.
Atema 2015 was a multi-centre study including six centres.
The most commonly used CT-protocol specified the following: slice thickness 3 mm; voltage 120 kV; and mAs product 165 mAs.
These values are used in the table.

Table 2. Results from studies comparing different types of contrast enhancement using a randomised or a paired design *(Continued)*

Study	Design	Sensitivity/specificity according to type of contrast enhancement					
		None	IV	Oral	Rectal	IV+oral	IV+oral +rectal
Hekimoglu 2011	Randomised	-	0.77/0.93	-	-	0.97/0.99	-
Hershko 2007	Randomised	0.90/0.86	-	-	0.95/0.92	1.00/0.88	-
Kepner 2012	Randomised	-	1.00/0.99	-	-	1.00/0.95	-
Mittal 2004	Randomised	-	-	-	1.00/1.00	-	0.98/0.50
Keyzer 2009	Randomised & paired	0.75/0.93	0.85/0.98	0.85/0.96	-	1.00/0.98	-
Jacobs 2001	Paired	-	-	0.76/0.94	-	0.92/0.95	-

Results for the standard-dose CT-protocols.

Table 3. Results from studies comparing low-dose and standard-dose CT-protocols using a randomised or a paired design *(Continued)*

Study	Design	Contrast enhancement	Sensitivity/specificity	
			Low-dose protocol	Standard-dose protocol
Kim 2012	RCT	Intravenous	0.95/0.93	0.95/0.94
Keyzer 2004[a]	Paired	Unenhanced	0.97-1.00/0.80-0.94	0.97-1.00/0.82-0.94
Keyzer 2009[b]	Paired	Unenhanced	0.80-0.85/0.91-0.93	0.75-0.75/0.93-0.93
		Intravenous	0.70-0.80/1.0-1.0	0.85-0.85/0.98-0.98
		Oral	0.85-1.0/0.88-0.96	0.85-0.92/0.96-0.96
		Oral and intravenous	0.85-1.0/0.96-0.98	1.0-1.0/0.96-1.0
Platon 2009	Paired	Oral (low dose)	0.95/0.96	1.0/0.96
		Oral and intravenous (standard dose)		

CT: computed tomography.
RCT: randomised controlled trial.
[a]Results are given as the range of sensitivity and specificity for the four participating radiologists.
[b]Results are given as the range of sensitivity and specificity for the two participating radiologists.

Table 4. Subgroup analyses according to type of contrast enhancement and radiation dose *(Continued)*

Subgroups by enhancement and dose	Number of analyses (studies)[a]	Summary estimates with 95% CI				Absolute differences in summary estimates with 95% CI	
		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Sensitivity	Specificity
Unenhanced	19 (19)	0.91 (0.87-0.93)	0.94 (0.90–0.96)	15 (9–24)	0.10 (0.07–0.14)	–	–
IV contrast	18 (17)	0.96 (0.92–0.98)	0.93 (0.90–0.95)	14 (9–20)	0.04 (0.02–0.09)	0.04[b] (0.00–0.09)	-0.01[b] (-0.04–0.03)
IV and oral contrast	15 (15)	0.96 (0.93–0.98	0.94 (0.92–0.96)	17 (12–26)	0.04 (0.02–0.07)	0.05[b] (0.01–0.09)	0.01[b] (-0.03–0.04)
Rectal contrast	9 (9)	0.97 (0.93–0.99)	0.95 (0.90–0.98)	21 (9–51)	0.04 (0.02–0.08)	0.05[b] (0.01–0.09)	0.01[b] (-0.03–0.06)
Oral contrast	7 (7)	0.89 (0.81–0.94)	0.94 (0.90–0.97)	16 (9–29)	0.11 (0.06–0.21)	-0.01[b] (-0.08–0.6)	0.01[b] (-0.03–0.05)
Standard dose	67 (64)	0.95 (0.93–0.96)	0.94 (0.92–0.95)	15.6 (12.3–19.7)	0.05 (0.04–0.07)	–	–
Low dose	8 (7)	0.94 (0.90–0.97)	0.94 (0.91–0.96)	16 (10–24)	0.06 (0.03–0.11)	0.00[c] (-0.04–0.05)	0.00[c] (-0.04–0.03)
Overall	71	0.95 (0.93–0.96)	0.94 (0.92–0.95)	15 (12–19)	0.05 (0.04–0.07)	-	-

CI: confidence interval.
IV: intravenous.
[a]Randomised and paired studies provided two or more analyses.

bAbsolute difference compared to unenhanced CT.
cAbsolute difference compared to standard-dose CT.

Table 5. Sensitivity analysis - effects of selecting results for other CT-protocols in paired studies *(Continued)*

Subgroup - by enhancement and dose	Number of analyses (studies)	Summary estimates with 95% CI			
		Original analysis		Sensitivity analysis	
		Sensitivity	Specificity	Sensitivity	Specificity
Unenhanced	19 (19)	0.91 (0.87-0.93)	0.94 (0.90-0.96)	0.91 (0.88-0.94)	0.94 (0.90-0.96)
Intravenous contrast	18 (17)	0.96 (0.92-0.98)	0.93 (0.90-0.95)	0.96 (0.91-0.98)	0.93 (0.90-0.95)
Intravenous and oral contrast	15 (15)	0.96 (0.93-0.98	0.94 (0.92-0.96)	0.96 (0.93-0.98	0.94 (0.92-0.96)
Oral contrast	7 (7)	0.89 (0.81-0.94)	0.94 (0.90-0.97)	0.90 (0.82-0.95)	0.94 (0.90-0.96)
Low dose	8 (7)	0.94 (0.90-0.97)	0.94 (0.91-0.96)	0.95 (0.91-0.97)	0.94 (0.91-0.96)
Overall	71 (64)	0.95 (0.93-0.96)	0.94 (0.92-0.95)	0.95 (0.93-0.96)	0.94 (0.92-0.95)

CI: confidence interval.
CT: computed tomography.

Table 6. Results from previously published meta-analyses

Author and publication year	Number of included studies	Focus of review	Summary sensitivity (95% CI)	Summary specificity (95% CI)
Terasawa 2004	12	Adults, any CT modality, prospective studies	0.94 (0.91-0.95)	O.95 (0.93-0.96)
Anderson 2005	23	Adults, comparison of enhancement with: oral contrast vs any enhancement excluding oral contrast	0.92 0.95	0.94 0.97
Weston 2005	12	Adults, any CT modality	0.97 (0.94–0.98)	0.94 (0.92–96)
Doria 2006	21[a]	Any CT modality, separate results for adults and children	0.94 (0.92-0.95)[a]	0.94 (0.94-0.96)[a]
Al-Khayal 2007	25	Adults and children, any CT modality, prospective studies	0.93 (0.92-0.95)	0.93 (0.92-0.95)
van Randen 2008	6	Mainly adults or adolescents, any CT modality, prospective studies with direct comparisons of CT and US	0.91 (0.84-0. 95)	0.90 (0.85-0.94)

Table 6. Results from previously published meta-analyses *(Continued)*

Hlibczuk 2010	7	Unenhanced, helical CT	0.93 (0.90-0.95)	0.96 (0.94-0.98)
Dahabreh 2015	72[a]	Any CT modality Separate results for adults, children, women of reproductive age, and pregnant women	0.96 (0.95-0.97)[a]	0.96 (0.93-0.97)[a]
Xiong 2015	7	Unenhanced CT, prospective studies	0.90 (0.86–0.92)	0.94 (0.92–0.97)
Aly 2016	5	Comparison of: low-dose CT vs standard-dose CT	0.93 (0.89-0.96) 0.94 (0.91-0.96)	0.93 (0.90-0.96) 0.94 (0.92-0.96)
Yun 2017	9	Comparison of: low-dose CT and standard-dose CT in adults and children	0.96 (0.92-0.98) 0.96 (0.94-0.98)	0.93 (0.89-0.96) 0.92 (0.88-0.95)

CI: confidence interval.
CT: computed tomography.
[a]Studies and results in adults.

APPENDICES

Appendix 1. MEDLINE search strategy

MEDLINE Ovid (Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present), 16 June 2017

1. Appendicitis/

2. Appendicitis.tw,kf.

3. (right adj2 (iliac fossa* or quadrant pain)).tw,kf.

4. Appendix/su

5. Appendectomy/

6. (appendec* or appendicec* or appendicit*).tw,kf.

7. Or/1-6

8. Tomography, x-ray computed/

9. *tomography, spiral computed/

10. (compute* tomography or computer assisted tomography).tw,kf.

11. ((ct or cat) adj (x-ray or scan*)).tw,kf.

12. ((compute* tomography or computer assisted tomography or ct or cat) adj1 (spiral or helical)).tw,kf.

13. Or/8-12

14. 7 and 13

15. Exp animals/ not humans.sh.

16. 14 not 15

Appendix 2. Embase search strategy

Embase Ovid (1974 to 2017 Week 24), 16 June 2017

1. appendicitis/ or acute appendicitis/ or appendix perforation/

2. ((right adj2 (iliac fossa* or quadrant pain)).tw,kw.

3. Appendix/su

4. Appendectomy/

5. (appendec* or appendicec* or appendicit*).tw,kw.

6. ((operat* or resect* or remov* or suger* or surgical or laparoscop* or acute) adj5 appendi*).tw,kw.

7. Or/1-6

8. Computer assisted tomography/

9. *spiral computer assisted tomography/

10. (ct or cat) adj (x-ray or scan*).tw,kw.

11. ((spiral or helical) adj1 (compute* tomography or computer assisted tomography or cat or ct).tw,kw.

12. Or/8-11

13. 7 and 12

14. (exp animal/ or exp invertebrate/ or animal.hw. or nonhuman/) not (exp human/ or human cell/ or (human or humans or man or men or wom?n).ti.)

15. 13 not 14

Appendix 3. QUADAS-2 rating guideline

Domain 1: patient selection

Signalling questions and answering guidelines

1) *Was a consecutive or a random sample of* person*enrolled?*
Answer 'yes' if one of the following conditions is met.
a. It is explicitly stated in the study report that enrolment was consecutive (or random).
b. It is reported that all eligible, screened, or potential study participants were included, and that enrolment took place at all hours on any day during the enrolment period.
Answer 'no' if neither of the conditions is met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

2) *Was a case-control design avoided?*
This question is irrelevant because studies with case-control design are excluded from the review.

Guidelines for assessing risk of bias

Risk of bias from patient selection will be assessed as 'low' when signalling question 1 is answered 'yes'.
Risk will be assessed as 'high' when signalling question 1 is answered 'no'.
Risk will be assessed as 'unclear' when insufficient information is reported to answer signalling question 1.

Guidelines for assessing concern regarding applicability

Concern regarding applicability in relation to patient selection will be assessed as 'low' when the study population represents an unselected sample of adults with suspected appendicitis. We will consider the sample selected in case of inappropriate exclusions, which we define as exclusions that are unrelated to execution of the index test (i.e. fear of radiation exposure, allergy to the contrast agent, inability to be positioned). Hence, exclusion of women or persons with diabetes will be considered inappropriate because the study question concerns the accuracy of CT for appendicitis in adults in general. By contrast, we do not consider it inappropriate if persons with extreme a priori probabilities of appendicitis are excluded. As stated in the background section, it is probably in persons with intermediate a priori probability that CT has the greatest role in guiding decisions on management. We are planning a sensitivity analysis of studies

including persons with intermediate a priori risk of appendicitis. Finally, exclusion of severely, acutely ill (i.e. septiceamic) persons and persons with mental incapacities is not considered inappropriate. If inappropriate exclusions account for 5% or less of the number of included persons, the potential impact of inappropriate exclusions will be considered negligible.

Concern will be assessed as 'high' when the study population does not represent an unselected sample of adults with suspected appendicitis.

Concern will be assessed as 'unclear' when insufficient information is available.

Domain 2: index test

Signalling questions and answering guidelines

1) *Were the index test results interpreted without knowledge of the results of the reference standard?*
For practical reasons, the CT-scan must take place before it is decided if the patient should have surgery with possible appendectomy or clinical follow-up. However, the CT evaluations used in the analyses do not necessarily have to take place in relation to acquisition of the scan. Thus, analyses may be based on CT evaluations that are performed subsequent to surgery. Such analyses could be biased if the radiologist is aware of the intraoperative findings.
Answer 'yes' if one of the following conditions is met.
a. The CT evaluations used in the analyses were performed before the patient had surgery.
b. The CT evaluations used in the analyses were postponed evaluations or reevaluations, and the radiologists were kept unaware of whether persons had surgery or not, as well as of intraoperative findings.
Answer 'no' if neither of the conditions is met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

2) *If a threshold was used, was it pre-specified?*
Answer 'yes' if the following two conditions are met.
a. The components (i.e. appendix diameter, presence of appendicolith, periappendiceal inflammation/edema) included in the evaluation of the CT-scan are explicitly reported in the study report.
b. The hierarchy and logical combination of components are explicitly reported in the study report.
Answer 'no' if one or more of the conditions above are not met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

Risk of bias from index test execution will be assessed as 'low' when signalling questions 1 and 2 are answered 'yes'.
Risk will be assessed as 'high' when signalling question 1 or 2 is answered 'no'.
Risk will be assessed as 'unclear' when insufficient information is reported to answer signalling questions 1 and 2.

Guidelines for assessing concern regarding applicability

Two issues will influence our assessment concerning applicability in relation to execution of the index test.
1)*Is the index test described in sufficient detail to permit its replication?*
Aswer 'yes' when the following details are reported.
a. Number of slices of the CT device.
b. Use of multi-planar reformations (assumed not used if the number of slices of the CT device is less than 16, unless stated otherwise).
c. Use of peroral, intravenous, or rectal contrast.
d. Region included in the scan (entire abdomen vs lower abdomen).
e. Slice thickness, slice interval, and mAs product.
Answer 'no' if one or more of the details listed above (a to e) are not described.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

2)*Was the analysis based on the initial evaluation of the CT-scan by the radiologist on call?*
Answer 'yes' if the analysis is based on the initial assessment of the CT-scan by the radiologist on call.
Answer 'no' if the analysis is based on a reassessment of the CT-scan by a senior radiologist or a consensus panel.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

Concern regarding applicability in relation to index test execution will be assessed as 'low' when questions 1 and 2 are answered 'yes'.
Concern will be assessed as 'high' when question 1 or 2 is answered 'no'.
Concern will be assessed as 'unclear' when insufficient information is reported to answer questions 1 and 2.

Domain 3: reference standard

Signalling questions and answering guidelines

1) *Is the reference standard likely to correctly classify the target condition?*
Answer 'yes' if the following conditions are met.
a. The diagnosis of appendicitis is based on the judgement of the surgeon during laparoscopy or laparotomy. Also classify as 'yes' if the diagnosis of appendicitis is based on histological examination of the removed appendix.
b. The diagnosis of appendicitis in patient who did not have surgery is based on clinical follow-up. A clinical examination, a letter with a questionnaire or a phone call from a doctor or a nurse with standardised questions to confirm recovery within 7 to 31 days from discharge will qualify as adequate clinical follow-up.
Answer 'no' if the diagnosis of appendicitis (or its absence) is not based on the conditions stated above.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

2) *Were the reference standard results interpreted without knowledge of results of the index test?*
Answer 'yes' if the following three conditions are met.
a. The surgeons performing the laparoscopies or the laparotomies are kept unaware of the results of the CT-scan (this condition is irrelevant if the diagnosis of appendicitis is based on histological assessment and if the appendix is removed in all persons who have surgery).
b. The pathologists examining the removed appendices are kept unaware of the result of the CT-scan (this condition is irrelevant if the diagnosis of appendicitis is based on the macroscopic appearance of the appendix during surgery).
c. The members of the study staff in charge of clinical follow-up are kept unaware of the results of the CT-scan.
Answer 'no' if one of the relevant conditions stated above is not met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

Risk of bias related to the reference standard will be assessed as 'low' when signalling questions 1 and 2 are answered 'yes'.
Risk will be assessed as 'high' when signalling question 1 or 2 is answered 'no'.
Risk will be assessed as 'unclear' when insufficient information is reported to answer signalling questions 1 and 2.

Guidelines for assessing concern regarding applicability

The use of intraoperative assessment of the appendix, as opposed to histological assessment, could potentially influence applicability of study results to settings where the appendix is always removed for histological assessment during surgery for suspected appendicitis, and vice versa. However, as the validity of intraoperative assessment is unsettled, it is not feasible to specify a concern regarding intraoperative versus histological assessment of the appendix as inflamed or normal. Both assessments are considered appropriate reference standards. For descriptive purposes, we will extract data concerning this issue.

Domain 4: flow and timing

Signalling questions and answering guidelines

1) *Did all* persons*receive a reference standard?*
Answer 'yes' if at least 95% of included persons had surgery with macroscopic assessment of the appendix, histological assessment of the removed appendix, or clinical follow-up.
Answer 'no' if less than 95% of included persons had surgery with macroscopic assessment of the appendix, histological assessment of the removed appendix, or clinical follow-up.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

2) *Did all* persons*receive the same reference standard?*
Answer 'yes' if one of the following conditions is met.
a. 90% of included persons had surgery with macroscopic assessment of the appendix or histological assessment of the removed appendix.
b. 90% of included persons were managed by clinical follow-up.
Answer 'no' if neither of the conditions is met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

It could be argued that surgery in persons with low a priori risk of appendicitis is unethical. However, in our view, this does not change the potential for differential verification bias when more than one reference standard is used.

3) *Did all persons with a positive CT-scan undergo surgery?*
Answer 'yes' if all persons with a positive CT-scan underwent surgery.
Answer 'no' if some persons with a positive CT-scan had clinical follow-up.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

4) Did all persons*with a negative CT-scan have clinical follow-up?*
Answer 'yes' if all persons with a negative CT-scan had clinical follow-up.
Answer 'no' if some persons with a negative CT-scan underwent surgery.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

5) *Was the choice of reference standard independent of the result of the index test?*
Answer 'yes' if the surgeons deciding on surgery or clinical follow-up were kept unaware of the outcome of the CT-scan.
Answer 'no' when the surgeons deciding on surgery or clinical follow-up were aware of the outcome of the CT-scan.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

6) *Were all persons included in the analysis?*
Answer 'yes' if the analyses encompassed all included persons. Also, answer 'yes' if 5% or less were excluded from the analysis because no reference standard assessment was available (to accommodate signalling question 1).
Answer 'no' if the requirement stated above is not met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

7) *Was there an appropriate interval between index test and reference standard?*
The appropriate time interval between the CT-scan and surgery is unclear. After careful consideration, we have reached the conclusion that we are unable to specify this interval in a meaningful way.
If the patient does not have appendicitis at the time of the CT-scan, it is unlikely that appendicitis will occur within the next weeks. Hence, the intraoperative appearance of the appendix and the histological appearance of the removed appendix are unlikely to change if surgery is undertaken within weeks after the index CT-scan. One caveat, however, relates to serosal inflammation of the appendix (periappendicitis) caused by disease processes in neighbouring organs that may hamper the intraoperative assessment of the appendix.

On the other hand, if the patient has appendicitis at the time of the CT-scan, the intraoperative appearance of the appendix and the histological appearance of the removed appendix will depend on the stage of the disease at the time of the CT-scan and the progression of the inflammatory process until surgery. In the light of the spectrum of disease courses ranging from spontaneous recovery to perforation and abscess formation, it is difficult to specify what makes up an appropriate interval between the CT-scan and surgery.

With respect to clinical follow-up, we believe that it should take place within 7 to 31 days from discharge. We admit this is arbitrary; however, if the interval is too short, cases with appendicitis may be overlooked, whereas 'new' cases of appendicitis may be mistaken for the index case if the interval is too long. Nevertheless, we consider the time interval an integral part of clinical follow-up, which we assess in signalling question 1, domain 2. For descriptive purposes, we will extract data on intervals between CT-scans and reference standards.

Guidelines for assessing risk of bias

Risk of bias related to patient flow and timing will be assessed as 'low' when signalling questions 1, 2, and 6 are answered 'yes'.
Risk will be assessed as 'high' when signalling question 1, 2, or 6 is answered 'no'.
Risk will be assessed as 'unclear' when insufficient information is reported to answer signalling questions 1, 2, and 6.

CONTRIBUTIONS OF AUTHORS

Drafting the protocol	TS Vejborg, ED Rappeport, JB Reitsma, Peer Wille-Jørgensen, B Rud
Searching the literature	TS Vejborg, B Rud
Extracting study data	ED Rappeport, B Rud
Entering data into RevMan	B Rud
Performing and interpreting analyses	B Rud
Drafting the review	B Rud
Revising the review draft	TS Vejborg, ED Rappeport, JB Reitsma, Peer Wille-Jørgensen,B Rud

DECLARATIONS OF INTEREST

The review authors have no conflicts of interests

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Selection criteria

In the protocol, we planned that we would contact study authors and request subgroup results when more than 10% of participants were younger than 15 years of age. This turned out to be not feasible because 28 studies with 30 study populations included more than 10% or an unclear proportion of participants younger than 15 years of age. Therefore, we decided to include these 28 studies, and we planned to perform a sensitivity analyses to explore whether summary sensitivity and specificity differed in these studies compared to the remaining studies. Studies that explicitly focused on a paediatric population were still excluded.

Subgroup analyses

Subgroup analyses according to type of contrast enhancement and a subgroup analysis for low-dose CT were added. In the protocol, the former analyses were planned as sensitivity analyses.

Sensitivity analyses

In the protocol, we planned the following sensitivity analyses.

- CT without contrast enhancement.
- CT with intravenous contrast enhancement.
- CT with 16 or higher slice technology.
- Participants with intermediate clinical suspicion of acute appendicitis (as defined by degree of clinical suspicion, prior testing, or prevalence of appendicitis).
- Subgroup analyses according to CT criteria used for the diagnosis of appendicitis.

In the review, the first two bullets are included as part of the subgroup analyses according to type of contrast enhancement and radiation dose, the third and fourth bullets are included in the analysis of heterogeneity, and the fifth bullet is cancelled due to the consistency in criteria for the CT diagnosis of appendicitis.

Sensitivity analyses have been added to explore the effects of including different analyses from paired studies evaluating the accuracy of two or more CT-protocols in the same participants.

Sensitivity analyses were added to explore the effects of methodological quality on summary estimates of sensitivity and specificity. In the protocol, it was planned to investigate the impact of each of the four domains in meta-regression analyses.

We also added a sensitivity analysis to explore whether summary estimates of sensitivity and specificity were affected by the inclusion of studies that used laparoscopic assessment of the appendix as a reference standard.

The title of the protocol was "Diagnostic accuracy of computed tomography for appendicitis in adults". We have revised the title of the review to make it consistent with the guidelines regarding Cochrane Reviews of diagnostic test accuracy.