Cochrane Library

Cochrane Database of Systematic Reviews

Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis (Review)

D'Souza N, Hicks G, Beable R, Higginson A, Rud B

D'Souza N, Hicks G, Beable R, Higginson A, Rud B.
Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis.
Cochrane Database of Systematic Reviews 2021, Issue 12. Art. No.: CD012028.
DOI: 10.1002/14651858.CD012028.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 3
BACKGROUND 5
OBJECTIVES 6
METHODS 6
RESULTS 8
Figure 1. 9
Figure 2. 11
Figure 3. 13
Test 1. MRI 15
Figure 4. 16
Figure 5. 17
Figure 6. 18
Figure 7. 19
Figure 8. 20
Figure 9. 21
Figure 10. 22
Figure 11. 23
Figure 12. 24
DISCUSSION 25
AUTHORS' CONCLUSIONS 27
ACKNOWLEDGEMENTS 27
REFERENCES 28
CHARACTERISTICS OF STUDIES 38
DATA 149
ADDITIONAL TABLES 149
APPENDICES 150
WHAT'S NEW 160
HISTORY 160
CONTRIBUTIONS OF AUTHORS 160
DECLARATIONS OF INTEREST 160
SOURCES OF SUPPORT 160
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 161
INDEX TERMS 161

[Diagnostic Test Accuracy Review]

Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis

Nigel D'Souza[1], Georgina Hicks[2], Richard Beable[3], Antony Higginson[3], Bo Rud[4]

[1]University Hospital Southampton, Southhamptom, UK. [2]Hammersmith Hospital, London, UK. [3]Queen Alexandra Hospital, Portsmouth, UK. [4]Gastrounit, Copenhagen University Hospital Hvidovre , Hvidovre, Denmark

Contact: Nigel D'Souza, nigel151@gmail.com.

Editorial group: Cochrane Colorectal Group.
Publication status and date: New, published in Issue 12, 2021.

Citation: D'Souza N, Hicks G, Beable R, Higginson A, Rud B. Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis. *Cochrane Database of Systematic Reviews* 2021, Issue 12. Art. No.: CD012028. DOI: 10.1002/14651858.CD012028.pub2.

ABSTRACT

Background

Appendicitis remains a difficult disease to diagnose, and imaging adjuncts are commonly employed. Magnetic resonance imaging (MRI) is an imaging test that can be used to diagnose appendicitis. It is not commonly regarded as a first-line imaging test for appendicitis, but the reported diagnostic accuracy in some studies is equivalent to computed tomography (CT) scans. As it does not expose patients to radiation, it is an attractive imaging modality, particularly in women and children.

Objectives

The primary objective was to determine the diagnostic accuracy of MRI for detecting appendicitis in all patients.

Secondary objectives:

To investigate the accuracy of MRI in subgroups of pregnant women, children, and adults.

To investigate the potential influence of MRI scanning variables such as sequences, slice thickness, or field of view.

Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, and Embase until February 2021. We searched the references of included studies and other systematic reviews to identify further studies. We did not exclude studies that were unpublished, published in another language, or retrospective.

Selection criteria

We included studies that compared the outcome of an MRI scan for suspected appendicitis with a reference standard of histology, intraoperative findings, or clinical follow-up. Three study team members independently filtered search results for eligible studies.

Data collection and analysis

We independently extracted study data and assessed study quality using the Quality Assessment of Studies of Diagnostic Accuracy - Revised (QUADAS-2) tool. We used the bivariate model to calculate pooled estimates of sensitivity and specificity.

Main results

We identified 58 studies with sufficient data for meta-analysis including a total of 7462 participants (1980 with and 5482 without acute appendicitis). Estimates of sensitivity ranged from 0.18 to 1.0; estimates of specificity ranged from 0.4 to 1.0. Summary sensitivity was 0.95 (95% confidence interval (CI) 0.94 to 0.97); summary specificity was 0.96 (95% CI 0.95 to 0.97). Sensitivity and specificity remained high on subgroup analysis for pregnant women (sensitivity 0.96 (95% CI 0.88 to 0.99); specificity 0.97 (95% CI 0.95 to 0.98); 21 studies, 2282 women);

children (sensitivity 0.96 (95% CI 0.95 to 0.97); specificity 0.96 (95% CI 0.92 to 0.98); 17 studies, 2794 children); and adults (sensitivity 0.96 (95% CI 0.93 to 0.97); specificity 0.93 (95% CI 0.80 to 0.98); 9 studies, 1088 participants), as well as different scanning techniques. In a hypothetical cohort of 1000 patients, there would be 12 false-positive results and 30 false-negative results. Methodological quality of the included studies was poor, and the risk of bias was high or unclear in 53% to 83% of the QUADAS-2 domains.

Authors' conclusions

MRI appears to be highly accurate in confirming and excluding acute appendicitis in adults, children, and pregnant women regardless of protocol. The methodological quality of the included studies was generally low due to incomplete and low standards of follow-up, so summary estimates of sensitivity and specificity may be biased. We could not assess the impact and direction of potential bias given the very low number of high-quality studies. Studies comparing MRI protocols were few, and although we found no influence of MRI protocol variables on the summary estimates of accuracy, our results do not rule out that some MRI protocols are more accurate than others.

PLAIN LANGUAGE SUMMARY

Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis

Review question

To check the accuracy of magnetic resonance imaging (MRI), a medical imaging tool used for taking detailed pictures of the inside of the body, to test for appendicitis.

Why is diagnosing appendicitis important?

Appendicitis is a very common condition that is usually treated with emergency surgery, but it can be difficult to diagnose. Up to one in four patients may be incorrectly diagnosed with appendicitis. Tools such as MRI can help diagnose appendicitis quickly and early.

What was studied in this review?

We studied the accuracy of MRI for appendicitis in all patients.

What are the main results of the review?

We analysed the results of 58 studies with 7462 participants to calculate the accuracy of MRI. The results of these studies indicate that in theory, if MRI were to be used in 1000 patients with suspected appendicitis, where 250 patients actually had appendicitis, then:

• an estimated 250 patients will have an MRI result indicating appendicitis, 12 of whom will not actually have appendicitis; and

• of the 750 patients with a result indicating that appendicitis is not present, 30 will actually have appendicitis.

MRI remained very accurate when looking specifically at adults, pregnant women, and children.

How reliable are the results of the studies in this review?

There were problems with how most of the studies were conducted that may have resulted in MRI appearing more accurate than it actually is.

To whom do the results of this review apply?

The results apply to people with suspected appendicitis, including adults, pregnant women, and children. Most studies were conducted in Europe and North America in large university hospitals. Patients had often undergone an ultrasound scan without a clear result.

What are the key messages of this review?

Based on the studies included in this review, MRI seems to be a very accurate test for appendicitis. The chance of wrongly diagnosing someone with appendicitis or missing appendicitis was less than 5%. However, as most of the included studies had problems, we cannot trust their results completely. Although MRI is promising, until better studies have been performed, we cannot firmly recommend the use of MRI for the diagnosis of appendicitis.

How up-to-date is this review?

We searched for and used studies published up to February 2021.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table

Patients/population	Patients with suspected appendicitis
Settings	Mostly tertiary care settings in North America, Europe, Asia, and the Middle East
Index test	MRI
Reference standard	Surgery (if MRI positive) or follow-up (if MRI negative)
Target condition	Appendicitis
Number of studies	59 studies with a total of 7482 participants met the inclusion criteria. We excluded 1 study from meta-analysis as all patients had appendicitis, leaving 58 studies with 7462 participants that were meta-analysed.
Methodological concerns	Most studies were of poor methodological quality and at high risk of bias, although concerns about applicability were low. The nature of follow-up was frequently limited to case note review, as most studies were retrospective.

Results	Number of studies (participants)	Summary sensitivity (95% CI)	Summary specificity (95% CI)	Prevalence	Post-test probability following a positive MRI outcome	Post-test probability following a negative MRI outcome
Overall	58 (7462)	0.95 (0.94 to 0.97)	0.96 (0.95 to 0.97)	0.15 (lower quartile)	0.82 (0.76 to 0.87)	0.01 (0.01 to 0.01)
				0.25 (median)	0.90 (0.85 to 0.93)	0.02 (0.01 to 0.03)
				0.40 (upper quartile)	0.94 (0.92 to 0.96)	0.04 (0.03 to 0.06)
Adults	9 (1088)	0.96 (0.93 to 0.97)	0.93 (0.80 to 98)	0.25 (lower quartile)	0.82 (0.58 to 0.93)	0.02 (0.01 to 0.03)
				0.57 (median)	0.95 (0.85 to 0.98)	0.06 (0.04 to 0.10)
				0.67 (upper quartile)	0.96 (0.90 to 0.99)	0.09 (0.05 to 0.14)
Paediatric patients	17 (2794)	0.96 (0.95 to 0.97)	0.96 (0.92 to 0.98)	0.21 (lower quartile)	0.86 (0.76 to 0.92)	0.01 (0.01 to 0.01)
				0.35 (median)	0.92 (0.87 to 0.96)	0.02 (0.01 to 0.03)

				0.43 (upper quartile)	0.94 (0.90 to 0.97)	0.03 (0.02 to 0.04)
Pregnant women	21 (2282)	0.96 (0.88 to 0.99)	0.97 (0.95 to 0.98)	0.09 (lower quartile)	0.76 (0.62 to 0.85)	0.00 (0.00 to 0.01)
				0.13 (median)	0.82 (0.71 to 0.90)	0.01 (0.00 to 0.02)
				0.21 (upper quartile)	0.89 (0.82 to 0.94)	0.01 (0.00 to 0.04)
Conclusion	MRI has a very high diagnostic accuracy for appendicitis. The included studies were of poor methodological quality, as follow-up was frequently incomplete and of low standard. The accuracy of MRI remained high when studies of high or unclear risk of bias were excluded. Consequently, our results do not completely support (or refute) the use of MRI as a first-line imaging test.					

Abbreviations: CI: confidence interval; MRI: magnetic resonance imaging

BACKGROUND

Target condition being diagnosed

Appendicitis is the most common abdominal emergency in general surgery. Over 42,000 and 270,000 appendicectomies are performed annually in the UK and USA, respectively (Hall 2002; Health and Social Care Information Centre 2012). Since a number of other medical conditions can mimic its symptoms and signs, appendicitis can be a challenging disease to diagnose (Bhangu 2020; Di Saverio 2020). Appendicitis is diagnosed clinically incorporating results from laboratory and imaging studies, but no single test or risk scoring system exists that can reliably identify it with 100% accuracy (Bhangu 2020; Di Saverio 2020).

Although spontaneous resolution of appendicitis has been previously reported (Liu 2011), the potential complications of septicaemia, peritonitis, or death from untreated appendicitis mean that treatment is mandated when appendicitis is provisionally diagnosed. A growing body of research has recently suggested that antibiotics may be considered as an alternative to surgery in uncomplicated appendicitis, but with a recurrence risk of up the 39% (CODA 2020; Di Saverio 2020; Harnoss 2017). Once a diagnosis of appendicitis is made, the traditional treatment is surgical excision of the appendix (appendicectomy) via open or laparoscopic approaches to the abdomen (Sauerland 2010).

An incorrect diagnosis of appendicitis may lead to unnecessary surgery if the underlying aetiology is self-limiting or requires medical treatment. Surgery will result in a negative appendicectomy, where the appendix is excised, but tissue analysis reveals no inflammation. Surgical complications from a negative appendicectomy occur in approximately 11% of patients (Bhangu 2013). The negative appendicectomy rate (NAR) in large-scale studies varies from 6.4% (Switzerland, Guller 2011), 11.8% (USA, Seetahal 2011), 18.2% (Hong Kong, Ma 2010) to 20.6% (UK, Bhangu 2013). More recent studies from the Netherlands, Van Rossem 2015, and the USA, Tseng 2019, have found a decreased NAR of 3.3% and 2.5%, respectively, with mandatory imaging.

Several Cochrane Reviews have investigated interventions for appendicitis (Andersen 2005; Cheng 2015; Rehman 2011; Sauerland 2010; Wilms 2011). Ultrasonography (US) and computed tomography (CT) are the other commonly used imaging modalities for appendicitis. The accuracy of CT was investigated by a Cochrane Review (Rud 2019).

Index test(s)

Magnetic resonance imaging (MRI) is an imaging modality that is increasingly used for the diagnosis of gastrointestinal (GI) disease (Stoker 2010). MRI uses magnetic fields to create images of the body, and is described as a safe imaging technology, with no exposure to radiation (Stoker 2010). Safety guidelines specify subgroups of patients that may be harmed during an MRI scan, for example patients with metallic implants or foreign bodies (Dill 2008). People with claustrophobia and most young children or babies may also not tolerate the noise and closed space within an MRI scanner (Aspelund 2014; Dill 2008; Thieme 2014).

MRI is frequently used to investigate gastrointestinal pathology (Martin 2005; Tkacz 2009), particularly Crohn's disease (Florie 2006; Sempere 2005). It can diagnose other groups of conditions that mimic appendicitis, such as gynaecological (Birchard 2005; Sohaib 2007; Zanardi 2003), or urinary tract pathology (Leyendecker 2008).

Historically, MRI has not been used as an imaging test for emergency abdominal conditions, where computed tomography (CT) or ultrasound (US) are the default modalities to image the appendix (Harringa 2019; Leeuwenburgh 2012; Rankey 2008). Previous generations of MRI scanners would take up to 30 minutes to scan the abdomen (Hormann 1998; Pedrosa 2009), whilst a CT took less than 5 minutes. Furthermore, MRI scans of the abdomen require a subspecialist interest in GI radiology or further training to interpret accurately (Leeuwenburgh 2012; Thieme 2014).

MRI scanning technology was developed in the 1970s, and subsequent advances in MRI hardware (coil technology), software (protocols and sequences), and radiology expertise have led to an increase in its diagnostic capabilities and quicker scan times (Johnson 2012; Stoker 2010; Zhu 2012). As MRI accuracy has increased and scanning time has reduced, a growing number of primary research studies support the use of MRI to diagnose appendicitis in adults as well as women and children, where avoidance of radiation from CT scanning is highly desirable (Blumenfeld 2011; Moore 2016; Repplinger 2016). A previous systematic review of eight studies on the diagnostic accuracy of MRI for appendicitis calculated the summary sensitivity and specificity at 0.97 and 0.95, respectively (Barger 2010), which has been consistent in subsequent meta-analyses (Blumenfeld 2011; Duke 2016; Kave 2019; Moore 2016; Repplinger 2016). This is comparable to the sensitivity and specificity of CT, at 0.95 and 0.94, respectively (Rud 2019). If MRI is confirmed to be an accurate, radiation-free imaging test, then it could be a valid alternative or even first-line imaging modality for appendicitis, particularly in children and pregnant women, in whom avoidance of radiation is especially desirable.

Clinical pathway

People admitted with a potential diagnosis of appendicitis should routinely undergo clinical assessment by history and examination from an emergency general surgical team (The Royal College of Surgeons of England 2014); on that basis alone, a diagnosis may be formed, and the decision to operate, discharge, or perform further investigations may be made. Urinalysis and blood tests are commonly performed investigations, followed by imaging studies (The Royal College of Surgeons of England 2014). Since the symptoms and signs of appendicitis are variable, and investigations may be falsely positive or negative, the diagnosis of appendicitis is based on clinical judgement, weighing relevant information from the patient's history and examination and investigation results (Di Saverio 2020).

US and CT are the two commonly used preoperative imaging tests (Bhangu 2020; Di Saverio 2020). If US or CT is positive for appendicitis, the patient will proceed to surgery. If US is inconclusive, the patient will either be admitted for observation, proceed to CT as a second-line test, or proceed to diagnostic laparoscopy (Bakker 2010). If CT is inconclusive, the person will be admitted for observation, or proceed to diagnostic laparoscopy (Rud 2019).

In most countries, MRI is not commonly used in individuals with suspected appendicitis, but MRI could replace US or CT as a first-line imaging test, or could be used as a second-line imaging test

following a negative or inconclusive US (Di Saverio 2020; Tseng 2019).

Alternative test(s)

Blood tests for appendicitis are used to check whether inflammatory markers (white blood cell count (WBC) or C-reactive protein (CRP)) are elevated, with a clinical suspicion (based on history and examination) of appendicitis (Bhangu 2020; Di Saverio 2020). In this clinical context, normal WBC and CRP values mean that appendicitis is unlikely (Gronroos 1999; Sengupta 2009). Other markers have also emerged such as bilirubin, D'Souza 2013; Giordano 2013, and procalcitonin (Yu 2013), although their exact role in the diagnosis of appendicitis is not established.

US is a commonly used investigation in the UK (Bhangu 2020; Jaunoo 2012), particularly in young women to exclude gynaecological abnormalities. It is cheaper than CT with no radiation burden to the patient, but as its diagnostic accuracy depends directly on the expertise of the operator, its sensitivity and specificity is frequently inferior to CT (D'Souza 2015; Terasawa 2004).

CT has excellent sensitivity and specificity of 95% and 94%, respectively, on meta-analysis (Rud 2019), and is widely available and quick to perform. It is still not commonly used in the UK and other countries due to its expense and radiation dose. An abdominal CT exposes the recipient to as much radiation as 2.7 years of background radiation (U.S. Department of Health and Human Services 2015). It is estimated that 0.4% of all cancers diagnosed in the USA will be due to radiation exposure from CT scans (Brenner 2007), and national data from Korea suggest an increased risk of haematological malignancies in patients undergoing CT to diagnose appendicitis (Lee 2020). However, new, low-dose CT protocols (2 to 3.3 mSv versus 16 mSv for standard CT abdomen-pelvis protocols) are also effective to diagnose appendicitis (Kim 2012; Kim 2017; Sippola 2020).

Diagnostic laparoscopy is an invasive, intraoperative diagnostic modality to confirm appendicitis by direct visualisation of the appendix or finding other intra-abdominal pathologies during keyhole surgery. The diagnostic capability of laparoscopy in cases of uncertainty has probably lowered the threshold for surgery. However, as intraoperative laparoscopic diagnosis of appendicitis can be difficult, diagnostic laparoscopy can paradoxically increase the NAR. In some studies, over 30% of appendices that look normal at laparoscopy are inflamed on histological analysis (Phillips 2009; Roberts 2008; Slotboom 2014). If no other significant pathology is seen inside the abdomen, some intraoperative protocols will mandate the appendix is removed, even if it looks normal, to ensure that microscopic appendicitis is not missed. The NAR has therefore gone up in some centres since the advent of laparoscopy (Akbar 2010; Jones 2012). Some centres or guidelines advocate leaving a normal appendix in situ, consequently decreasing the NAR rate (Teh 2000; Van Rossem 2015), but still subjecting the patient to surgery to achieve a diagnosis.

Rationale

Many conditions mimic the symptoms and signs of appendicitis. Up to one-third of all women of childbearing age with right iliac fossa pain are incorrectly diagnosed with appendicitis due to similar symptoms caused by a wide range of common gynaecological conditions (Bhangu 2013; Rothrock 1995). Women have a higher NAR in most studies compared to men (28.6% versus 12.8%) (Bhangu 2013).

All groups of patients, including children, women, and the elderly, also have alternate diagnoses that may mimic appendicitis (des Plantes 2016; Dillman 2016). Some of these conditions may be self-limiting (e.g. mesenteric adenitis or gastroenteritis) and will resolve without any treatment, or may require medical treatment only (Byott 2016). Other unexpected conditions found at surgery may result in patients not being appropriately informed of potential complications, or the procedure being performed by a non-specialist surgeon (Boyd-Carson 2019), with potentially worse outcomes.

When appendicitis is incorrectly diagnosed, the decision to operate may subject a patient to an avoidable operation with the risk of complications (Bhangu 2013). It additionally incurs costs to the hospital (costs of inpatient stay, surgery, treatment of complications); to the wider healthcare setting (costs of community follow-up by a general practitioner or family doctor); and to the economy (costs of time off work for the patient and their caregiver) (D'Souza 2018).

A lack of access to imaging resources can contribute to a higher NAR. CT has excellent diagnostic accuracy for appendicitis, and evidence exists from previous studies showing that routine CT scanning can decrease the NAR by excluding appendicitis or finding alternate diagnoses (Drake 2012; Tseng 2019). Due to its cost, CT may not be used routinely, but studies from the USA have confirmed that the cost of surgery and inpatient stay in hospitals with a high NAR can outweigh the cost of routine CT scanning in all patients (Pena 1999; Rao 1998). However, concerns still exist over the radiation exposure from CT, which may increase the scanned patient's lifetime risk of cancer (Lee 2020; Sippola 2020).

MRI is not commonly used to diagnose appendicitis (Di Saverio 2020; Tseng 2019). However, there is a growing body of evidence that MRI may be used as a radiation-free modality to diagnose appendicitis in all patient groups.

OBJECTIVES

The primary objective was to determine the diagnostic accuracy of MRI for detecting appendicitis in all patients.

Secondary objectives

- To investigate the accuracy of MRI in subgroups of pregnant women, children, and adults.
- To investigate the potential influence of MRI scanning variables such as sequences, slice thickness, or field of view.

METHODS

Criteria for considering studies for this review

Types of studies

We included studies that compared the outcome of an MRI scan for suspected appendicitis with a reference standard.

Observational studies (cohort or cross-sectional studies) and randomised test accuracy studies were eligible for inclusion. We used data from randomised test accuracy studies (if available) to extract measures of diagnostic test accuracy for MRI, not to

compare diagnostic accuracy of MRI with alternative tests. We excluded studies with fewer than 10 participants because such studies were considered to be case reports with insufficient information. We also excluded studies with a case-control design, as diagnostic accuracy studies with this design are prone to bias (Whiting 2013). We included studies irrespective of their publication status and language.

Participants

People with suspected appendicitis or with abdominal pain in the right lower quadrant. We excluded studies in people with abdominal pain in general.

Index tests

An abdominal MRI scan performed to assess for the presence of appendicitis.

Target conditions

The target condition is acute appendicitis. We considered disease status as dichotomous: appendicitis or not appendicitis.

Reference standards

The reference test to diagnose the presence or absence of appendicitis was histological analysis of the appendix specimen following surgery.

In a person who did not undergo appendicectomy, appendicitis was considered as not present if one of several conditions were satisfied:

- if there is a normal appearance to the appendix at surgery, with or without alternative intraoperative findings that explain right iliac fossa pain, and clinical follow-up that excludes a missed diagnosis of appendicitis;
- if patients are discharged without treatment for appendicitis and have an uneventful follow-up.

Search methods for identification of studies

Electronic searches

We searched the following bibliographic databases:

- The Cochrane Central Register of Controlled Trials (CENTRAL, in the Cochrane Library; Issue 1, 2021) 01 February 2021 (Appendix 1);
- Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) (1946 to 01 February 2021) (Appendix 2);
- Ovid Embase (1974 to 2021 Week 5) (Appendix 3).

We included studies in all languages. We developed our search strategy in conjunction with the Cochrane Colorectal Cancer Group editorial office (Managing Editor and Information Specialist) and the Cochrane Diagnostic Test Accuracy editors.

Searching other resources

Reference lists

We checked the bibliographies of all included or relevant studies, such as existing reviews, for further eligible studies. We also performed forward tracking of publications that cited the included studies. We planned to revise the search terms if over half of the finally included references originated from sources other than the electronic searches, but this was not necessary.

Grey literature

We checked published, citeable reports and international conference proceedings from the last 10 years for eligible data or references.

Correspondence

If we could not retrieve the full text of a study or extract data from potentially eligible studies, we contacted study authors to obtain a copy of the full text or data.

Data collection and analysis

Selection of studies

Three study team members (ND'S, AT, GH) independently screened the titles and abstracts for potentially relevant studies. We retrieved full-text articles of all potentially relevant studies and assessed them for eligibility.

Three study team members (ND'S, AT, GH) independently performed selection and data extraction processes. Any discrepancies were resolved by discussion or by referral to a third review author (BR) for arbitration.

Data extraction and management

We collected data using a standard data extraction form and analysed the collected data using Review Manager 5 software (RevMan 2014). We extracted data in duplicate for quality assurance. Any discrepancies were resolved by discussion or by referral to a third review author (BR) for arbitration.
The data collection form included the following variables.

- Patient demographics
- Selection criteria
- Recruitment procedure
- Clinical setting
- MRI scanner generation
- Body region scanned
- MRI sequence
- MRI scan time
- Contrast administration
- Radiologist number, experience or specialisation
- MRI tolerability
- MRI criteria for appendicitis
- Method of diagnosis of appendicitis
- Prevalence of appendicitis
- Type of appendicitis present (simple or complicated (gangrenous, perforated, abscess)).

Assessment of methodological quality

Three study team members (ND'S, AT, GH) used the QUADAS-2 tool to assess methodological quality (Whiting 2011). A rating guideline was developed (see Appendix 4). This tool was revised during the study to best capture all elements of bias present in

the included studies. We presented outcomes of methodological quality assessment in table format.

Statistical analysis and data synthesis

Primary study estimates of sensitivity and specificity were plotted in forest plots and in receiver-operating characteristic plots to visually explore variation between studies. We considered summary estimates of sensitivity and specificity most relevant because the outcome of MRI evaluations for appendicitis is essentially binary. Moreover, we anticipated little variation between studies in MRI criteria for appendicitis (i.e. criteria for positive MRI outcome). We therefore used the bivariate random-effects model to summarise sensitivity and specificity (Reitsma 2005). We included results from all studies in an overall meta-analysis and performed subgroup analyses to explore variation in test performance between adults, children, and pregnant women. We explored the effect of different aspects of the MRI protocol in meta-regression analyses (see below). In these analyses we added covariates to the bivariate model one at a time and assumed equal variance between groups for the random effects of logit sensitivity and logit specificity because the number of studies in the groups was generally low. We used likelihood ratio tests to compare the fit between models. We used the xtmelogit-command in Stata version 13 to perform the analyses (Stata 2015; Takwoingi 2013), following the guidelines in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). We computed summary positive and negative likelihood ratios from summary estimates of sensitivity and specificity. We calculated post-test probabilities of appendicitis following positive and negative MRI outcomes for the minimum, 25%, median, 75%, and maximum percentiles of pre-test probabilities in the included studies. If a study reported estimates of accuracy for several MRI criteria, we focused on the criterion that conferred the highest degree of clinical homogeneity with the other studies. If sensitivity and specificity were reported for several observers in studies with a paired design, we calculated mean counts for true positives, false positives, false negatives, and true negatives and rounded them to integers when overall results across observers were not available.

Investigations of heterogeneity

We performed meta-regression analyses to explore the effect on test performance of the following MRI protocol variables:

- field of view (e.g. whole abdomen versus limited area);
- slice thickness (≦ 4 mm versus > 4 mm);
- sequence (e.g. T2 weighted images only versus T2 and T1 weighted images);
- contrast (intravenous or oral contrast versus no contrast);
- total scan time (≦ 10 minutes versus > 10 minutes).

We made receiver operating characteristic (ROC) plots for all analyses of heterogeneity. When meta-analyses were unfeasible due to low numbers of studies in one of the groups, there is no summary point with confidence and prediction regions in the plot, just the primary study results.

Sensitivity analyses

We performed sensitivity analyses to assess if results from low-quality studies influenced summary estimates of sensitivity and specificity. We excluded studies with low risk of bias for domain one, two, and three, or three of four domains in these analyses. We also performed sensitivity analyses to assess if studies with outlying estimates of sensitivity or specificity in the ROC plot influenced summary estimates of accuracy. We excluded results from studies with outlying results in these analyses. These analyses were not planned in the protocol (see Differences between protocol and review).

We performed a further unplanned sensitivity analysis to investigate whether diagnostic accuracy differed when MRI was performed after a negative or inconclusive ultrasound for appendicitis.

Assessment of reporting bias

We did not assess reporting bias.

RESULTS

Results of the search

Our study retrieval process is documented in a PRISMA flow diagram (see Figure 1). Our search terms identified 2632 references. These included 655 MEDLINE references, 1894 Embase references, and 83 references from the Cochrane Library. We identified a further six additional references through other sources such as reference lists. We excluded 492 duplicates and 2005 irrelevant references through title and abstract screening. We retrieved 141 full texts for further assessment of eligibility. We excluded 76 of these studies, the reasons for which are provided in Figure 1, and assessed 14 studies as awaiting classification. A total of 59 studies met the inclusion criteria. One study was included in the systematic review but excluded from the meta-analysis, as all included participants had appendicitis, therefore specificity was indeterminable (Hormann 1998). The 58 studies included in the meta-analysis comprised 7462 participants, 1980 with and 5482 without appendicitis. The median prevalence of appendicitis was 0.25 (interquartile range 0.15 to 0.40).

Figure 1. PRISMA flow diagram

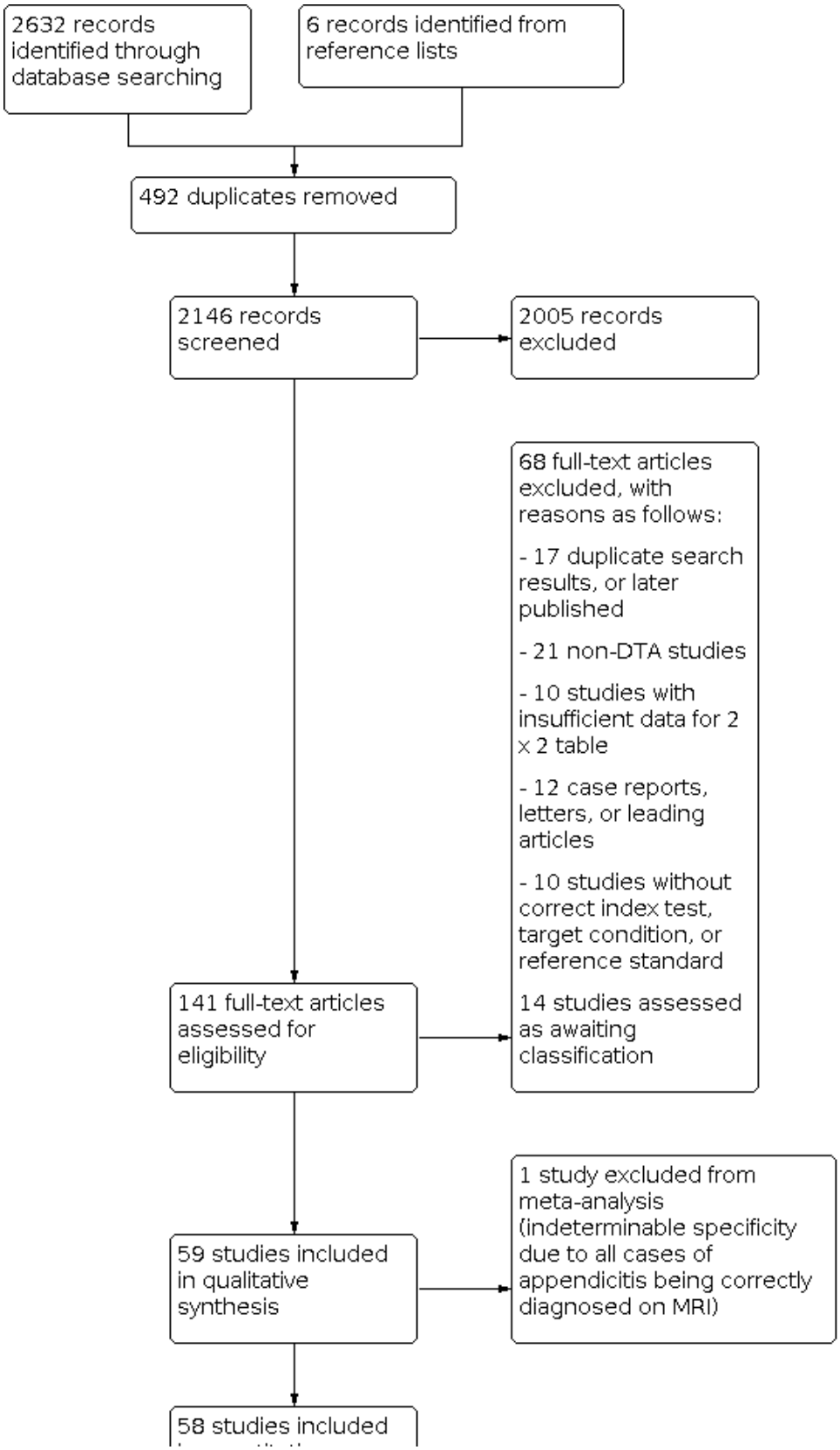
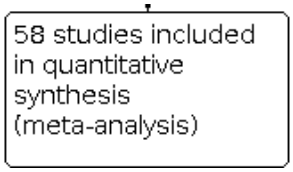


Figure 1. (Continued)



The study populations were classified as adults in 9 studies (1088 participants), children in 17 studies (2794 participants), and pregnant women in 21 studies (2282 participant). The remaining 13 studies recruited patients of different ages not classified within these groups.

The numbers of participants varied from 12 to 709 across the 58 studies. There were seven studies with fewer than 30 participants, 13 studies with fewer than 40 participants, and 23 studies with fewer than 50 participants.

The majority of included studies were conducted in the USA (35/59), followed by the Netherlands (5/59) and Turkey (5/59). By continent, 38 studies were conducted in North America (the USA and Canada), 16 in Europe (Turkey, the Netherlands, Italy, Israel, Germany, Denmark, and Austria), and 4 in Asia (China, Japan, Korea). The majority of study designs were retrospective (39/59). The remainder of studies (19/59) were prospective, and one study was a randomised trial. Nine studies did not specify the start date of recruitment, and one study began recruitment before 2000. The remaining studies (49/59) began recruitment after 2000. Twenty studies began recruitment after 2010. Most studies were performed on 1.5-Tesla MRI scanners (37/59); three studies used 3-Tesla MRI scanners exclusively; and two studies used 0.5-Tesla scanners exclusively. The remaining studies used a variety of 0.5- to 3-Tesla MRI scanners, whilst five studies did not describe the MRI scanner.

The MRI criteria for appendicitis was reported in 42 of the 59 studies (Appendix 5). The six most common features were appendix diameter (6 to 7 mm, 29 studies) and periappendicular inflammation (29 studies), wall thickening (16 studies), intraluminal fluid (15 studies), periappendiceal fluid (13 studies), and appendicolith (8 studies). Further data on the characteristics of included studies can be found below and in the Characteristics of included studies section.

Methodological quality of included studies

The methodological quality of the included studies is summarised in Figure 2 and Figure 3. Poor reporting in the primary studies limited methodological quality assessment.

Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

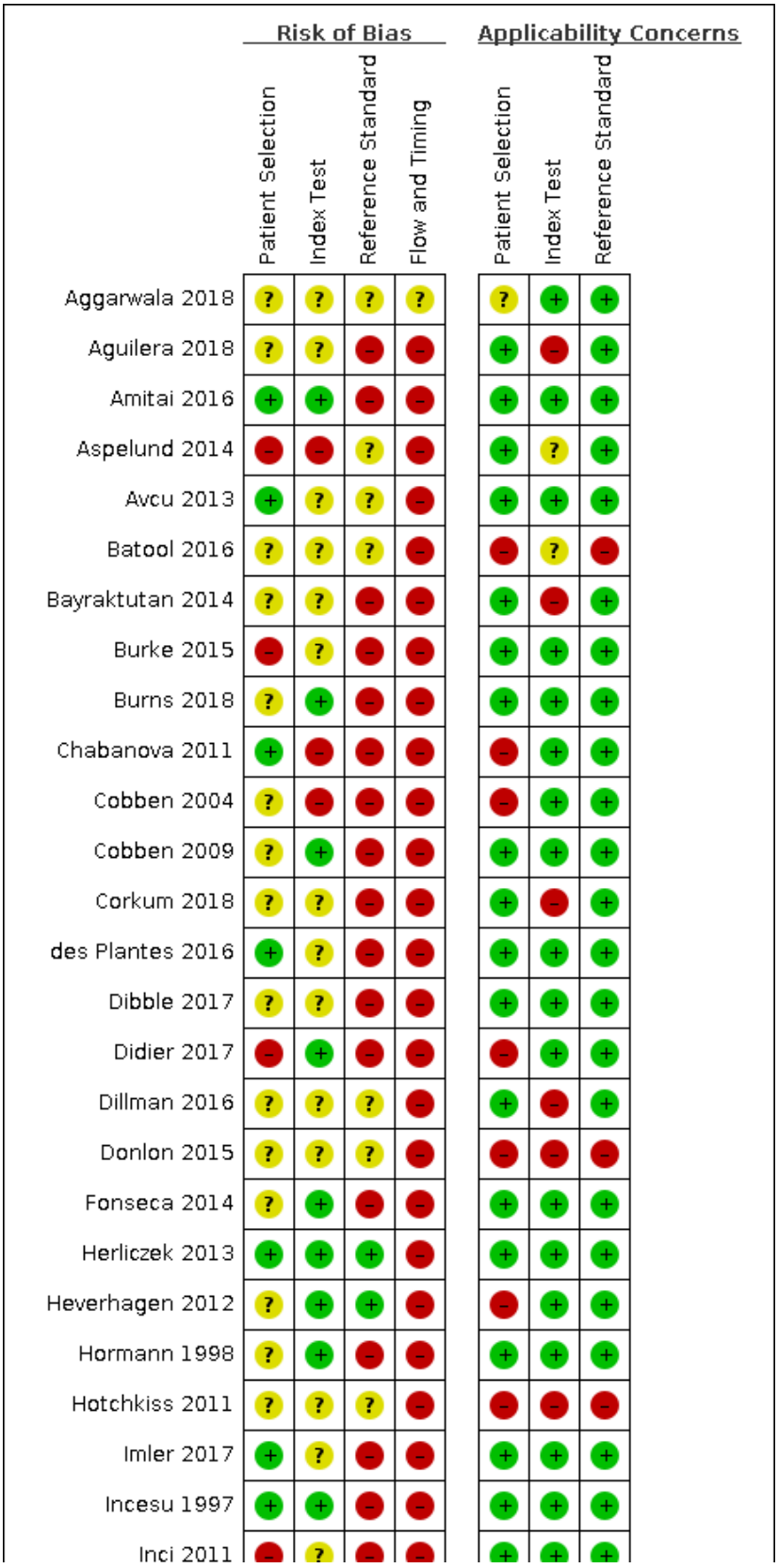


Figure 2. (Continued)

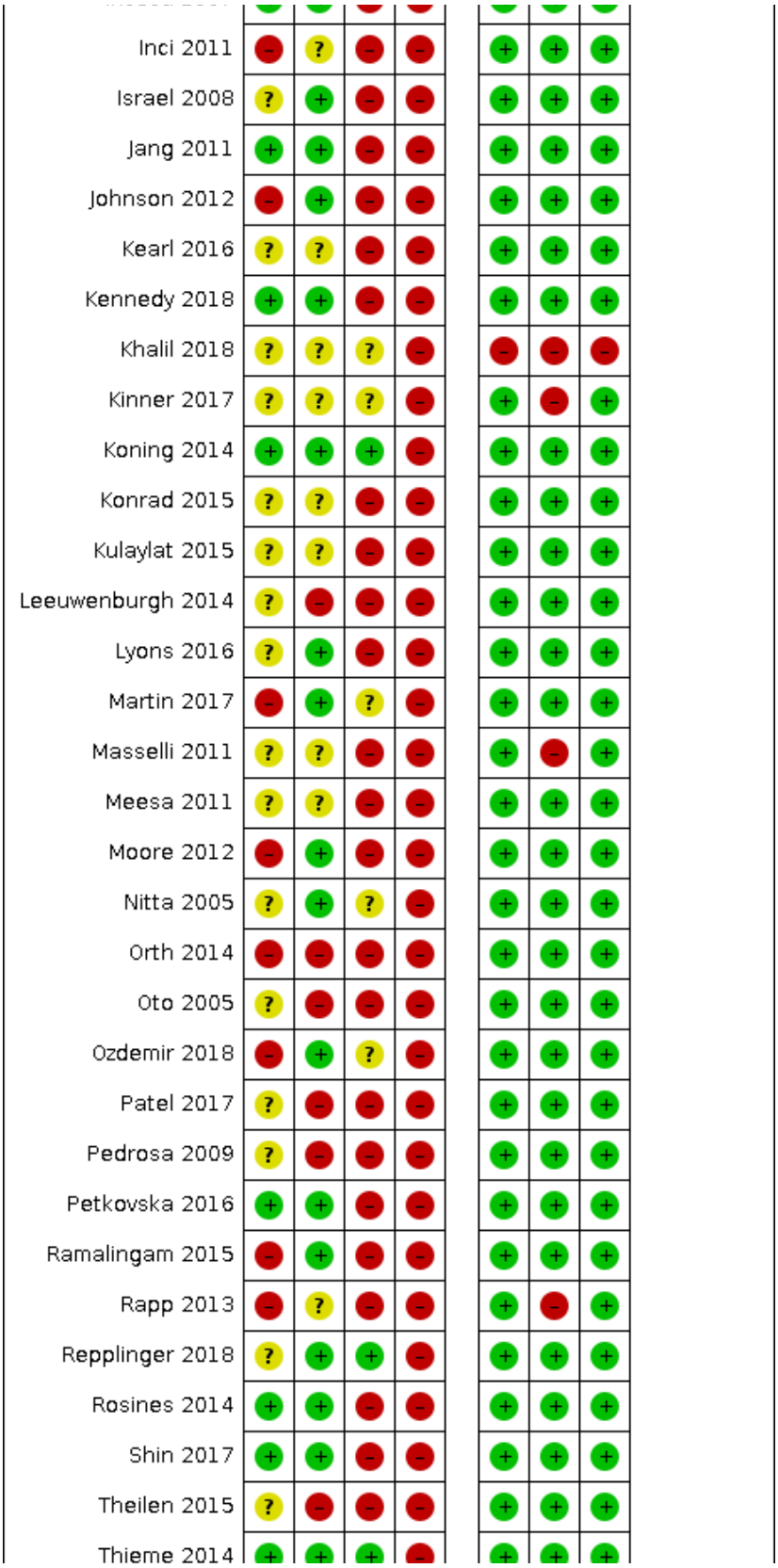


Figure 2. (Continued)

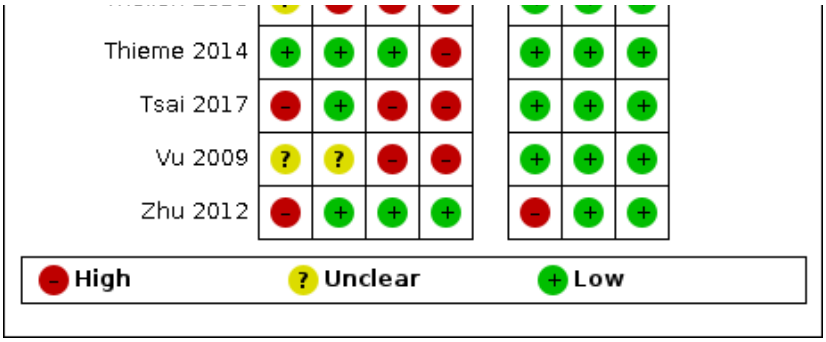
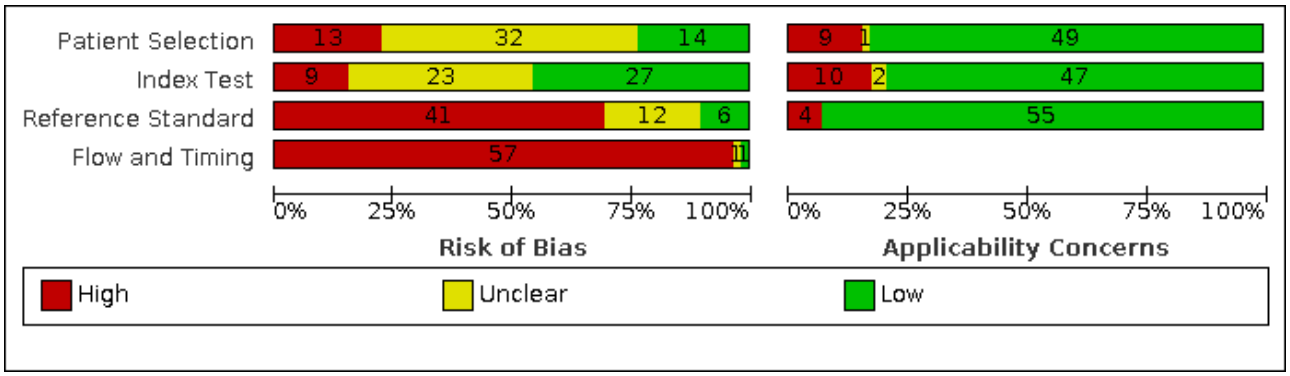


Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.



Domain 1: Patient selection

The risk of bias in patient selection was low in 14/59 (24%) of studies. The sole inclusion criterion was frequently that patients had suspected appendicitis. Studies failed to state whether patient recruitment was consecutive or random in 22/59 (37%) of studies, and exclusion criteria were often not described (33/59 (56%) of studies).

Domain 2: Index test

The risk of bias in the index test domain was low in 28/59 (47%) of studies. In 19/59 (32%) studies it was unclear whether the MRI was read without results of the reference standard. Prespecified MRI criteria for appendicitis were present in 35/59 (59%) of studies. The MRI protocol was described in some detail in most studies, but full information about the sequences included in the protocol, the field of view, the use of sedation, contrast enhancement and scanning time was frequently omitted.

Domain 3: Reference standard

The reference standard was the most important methodological limitation, resulting in low risk of bias in 10/59 (17%) of studies. When MRI was positive, the reference standard was operative or histological findings. Histology and surgical findings were the sole reference standard in three prospective studies where only patients scheduled for surgery were included (Chabanova 2011; Hormann 1998; Zhu 2012). When MRI scans did not report appendicitis, and the clinical suspicion for appendicitis was low, patients were managed non-operatively. In this case, patient follow-up was used as the reference standard to exclude appendicitis. The majority of studies were retrospective (39/59); patients were not contacted, and follow-up was frequently limited to case note review to exclude readmission, or not described at all. Only three retrospective studies described a methodology (telephone follow-up) that was not reliant on case note review. There was a reliable reference standard in only 20/59 (34%) of studies (e.g. adequate follow-up of sufficient duration). No studies described treatment of appendicitis with antibiotics during patient admission or as treatment for MRI-diagnosed appendicitis.

Domain 4: Flow and timing

Only one study achieved low risk of bias in the flow and timing domain (1/59 (2%) of studies). Per the QUADAS-2 tool, to achieve a low risk of bias in this domain a study would have to ensure all patients were included and received the same reference standard. The study that achieved low risk of bias in flow and timing was a series of 41 patients with suspected appendicitis, all of whom underwent surgery following their MRI scan (Zhu 2012). Only 4 studies (2/19 prospective, 2/39 retrospective) reported loss to follow-up or explicitly described no loss to follow-up. Studies that utilised case note review reported no loss to follow-up. As stated above, all but three studies incorporated differential verification, which was highly dependent on the MRI result. When we ignored differential verification in our assessment, 43, 6, and 9 studies had low, high, and unclear risk of bias for the flow and timing domain, respectively. However, as stated above, follow-up in most studies

was based on review of case notes to exclude readmission, and we considered follow-up complete in these studies.

Findings

The full results are shown in Data table 1 and are summarised in Summary of findings 1. The meta-analysis of 58 studies (7462 participants) reported an overall summary sensitivity and specificity of MRI for appendicitis of 0.95 (95% confidence interval (CI) 0.94 to 0.97; 58 studies, 7462 participants) and 0.96 (95% CI 0.95 to 0.97), respectively. The summary positive likelihood ratio was 25.8 (95% CI 17.6 to 37.7), and the summary negative likelihood ratio was 0.05 (95% CI 0.03 to 0.07). A flowchart of test performance in a theoretical cohort of 1000 patients with suspected appendicitis is shown in Figure 4. The forest plot is shown in Figure 5, and the ROC plot in Figure 6. Outliers are discussed below.

Test 1. MRI

MRI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Aggarwala 2018	9	1	1	41	0.90 [0.55, 1.00]	0.98 [0.87, 1.00]
Aguilera 2018	2	0	9	41	0.18 [0.02, 0.52]	1.00 [0.91, 1.00]
Amitai 2016	4	1	0	44	1.00 [0.40, 1.00]	0.98 [0.88, 1.00]
Aspelund 2014	61	1	0	80	1.00 [0.94, 1.00]	0.99 [0.93, 1.00]
Avcu 2013	39	0	1	15	0.97 [0.87, 1.00]	1.00 [0.78, 1.00]
Batool 2016	39	1	1	59	0.97 [0.87, 1.00]	0.98 [0.91, 1.00]
Bayraktutan 2014	33	0	3	9	0.92 [0.78, 0.98]	1.00 [0.66, 1.00]
Burke 2015	61	5	2	641	0.97 [0.89, 1.00]	0.99 [0.98, 1.00]
Burns 2018	10	3	1	48	0.91 [0.59, 1.00]	0.94 [0.84, 0.99]
Chabanova 2011	26	7	4	11	0.87 [0.69, 0.96]	0.61 [0.36, 0.83]
Cobben 2004	3	0	0	9	1.00 [0.29, 1.00]	1.00 [0.66, 1.00]
Cobben 2009	62	1	0	75	1.00 [0.94, 1.00]	0.99 [0.93, 1.00]
Corkum 2018	16	7	1	101	0.94 [0.71, 1.00]	0.94 [0.87, 0.97]
des Plantes 2016	25	0	4	83	0.86 [0.68, 0.96]	1.00 [0.96, 1.00]
Dibble 2017	14	1	2	2	0.88 [0.62, 0.98]	0.67 [0.09, 0.99]
Didier 2017	31	3	2	2	0.94 [0.80, 0.99]	0.40 [0.05, 0.85]
Dillman 2016	17	0	1	85	0.94 [0.73, 1.00]	1.00 [0.96, 1.00]
Donlon 2015	2	0	1	26	0.67 [0.09, 0.99]	1.00 [0.87, 1.00]
Fonseca 2014	11	0	0	20	1.00 [0.72, 1.00]	1.00 [0.83, 1.00]
Herliczek 2013	10	2	0	48	1.00 [0.69, 1.00]	0.96 [0.86, 1.00]
Heverhagen 2012	11	1	2	38	0.85 [0.55, 0.98]	0.97 [0.87, 1.00]
Hormann 1998	20	0	0	0	1.00 [0.83, 1.00]	Not estimable
Hotchkiss 2011	5	1	0	38	1.00 [0.48, 1.00]	0.97 [0.87, 1.00]
Imler 2017	8	1	0	28	1.00 [0.63, 1.00]	0.97 [0.82, 1.00]
Incesu 1997	33	2	1	24	0.97 [0.85, 1.00]	0.92 [0.75, 0.99]
Inci 2011	55	3	2	25	0.96 [0.88, 1.00]	0.89 [0.72, 0.98]
Israel 2008	4	0	1	28	0.80 [0.28, 0.99]	1.00 [0.88, 1.00]
Jang 2011	5	0	0	13	1.00 [0.48, 1.00]	1.00 [0.75, 1.00]
Johnson 2012	12	1	0	29	1.00 [0.74, 1.00]	0.97 [0.83, 1.00]
Kearl 2016	46	25	2	119	0.96 [0.86, 0.99]	0.83 [0.75, 0.88]
Kennedy 2018	124	36	6	446	0.95 [0.90, 0.98]	0.93 [0.90, 0.95]
Khalil 2018	182	6	6	374	0.97 [0.93, 0.99]	0.98 [0.97, 0.99]
Kinner 2017	12	1	2	33	0.86 [0.57, 0.98]	0.97 [0.85, 1.00]
Koning 2014	127	10	5	222	0.96 [0.91, 0.99]	0.96 [0.92, 0.98]
Konrad 2015	16	2	0	96	1.00 [0.79, 1.00]	0.98 [0.93, 1.00]
Kulaylat 2015	122	10	4	374	0.97 [0.92, 0.99]	0.97 [0.95, 0.99]
Leeuwenburgh 2014	113	7	4	99	0.97 [0.91, 0.99]	0.93 [0.87, 0.97]
Lyons 2016	23	8	0	58	1.00 [0.85, 1.00]	0.88 [0.78, 0.95]
Martin 2017	9	1	1	19	0.90 [0.55, 1.00]	0.95 [0.75, 1.00]
Masselli 2011	5	0	0	35	1.00 [0.48, 1.00]	1.00 [0.90, 1.00]
Meesa 2011	8	2	0	15	1.00 [0.63, 1.00]	0.88 [0.64, 0.99]
Moore 2012	40	5	1	162	0.98 [0.87, 1.00]	0.97 [0.93, 0.99]
Nitta 2005	29	1	1	6	0.97 [0.83, 1.00]	0.86 [0.42, 1.00]
Orth 2014	28	1	2	50	0.93 [0.78, 0.99]	0.98 [0.90, 1.00]
Oto 2005	3	1	1	18	0.75 [0.19, 0.99]	0.95 [0.74, 1.00]
Ozdemir 2018	67	3	4	20	0.94 [0.86, 0.98]	0.87 [0.66, 0.97]
Patel 2017	3	3	2	34	0.60 [0.15, 0.95]	0.92 [0.78, 0.98]
Pedrosa 2009	14	9	0	125	1.00 [0.77, 1.00]	0.93 [0.88, 0.97]
Petkovska 2016	65	2	2	334	0.97 [0.90, 1.00]	0.99 [0.98, 1.00]
Ramalingam 2015	8	6	0	88	1.00 [0.63, 1.00]	0.94 [0.87, 0.98]
Rapp 2013	18	10	2	187	0.90 [0.68, 0.99]	0.95 [0.91, 0.98]
Repplinger 2018	62	14	2	120	0.97 [0.89, 1.00]	0.90 [0.83, 0.94]
Rosines 2014	16	0	1	32	0.94 [0.71, 1.00]	1.00 [0.89, 1.00]
Shin 2017	21	47	1	56	0.95 [0.77, 1.00]	0.54 [0.44, 0.64]
Theilen 2015	12	6	1	152	0.92 [0.64, 1.00]	0.96 [0.92, 0.99]
Thieme 2014	58	5	0	41	1.00 [0.94, 1.00]	0.89 [0.76, 0.96]
Tsai 2017	13	21	1	198	0.93 [0.66, 1.00]	0.90 [0.86, 0.94]
Vu 2009	1	0	1	17	0.50 [0.01, 0.99]	1.00 [0.80, 1.00]
Zhu 2012	33	0	3	5	0.92 [0.78, 0.98]	1.00 [0.48, 1.00]

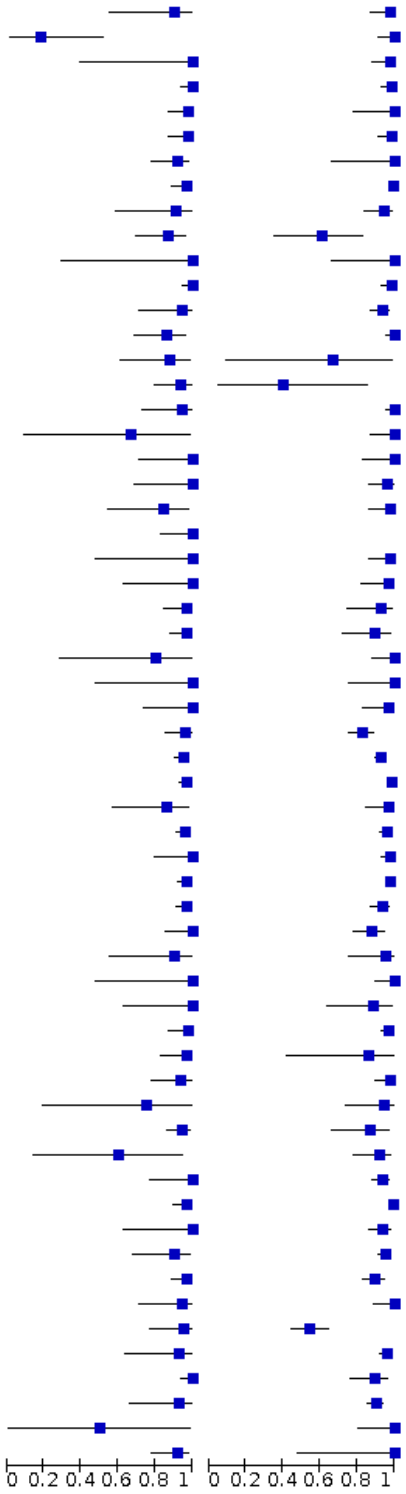
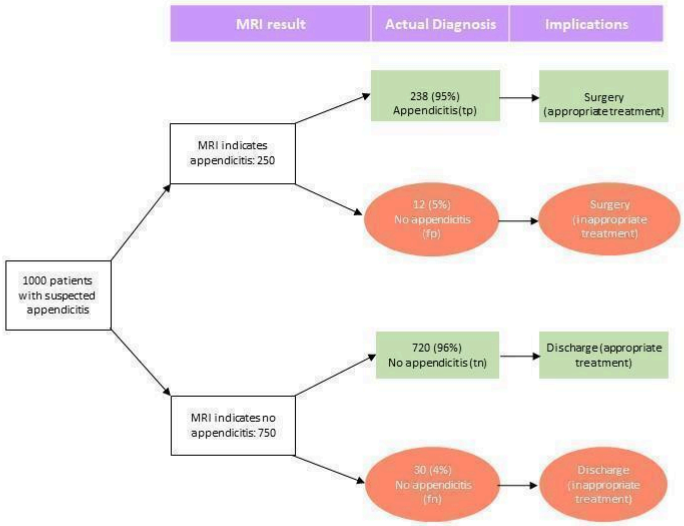
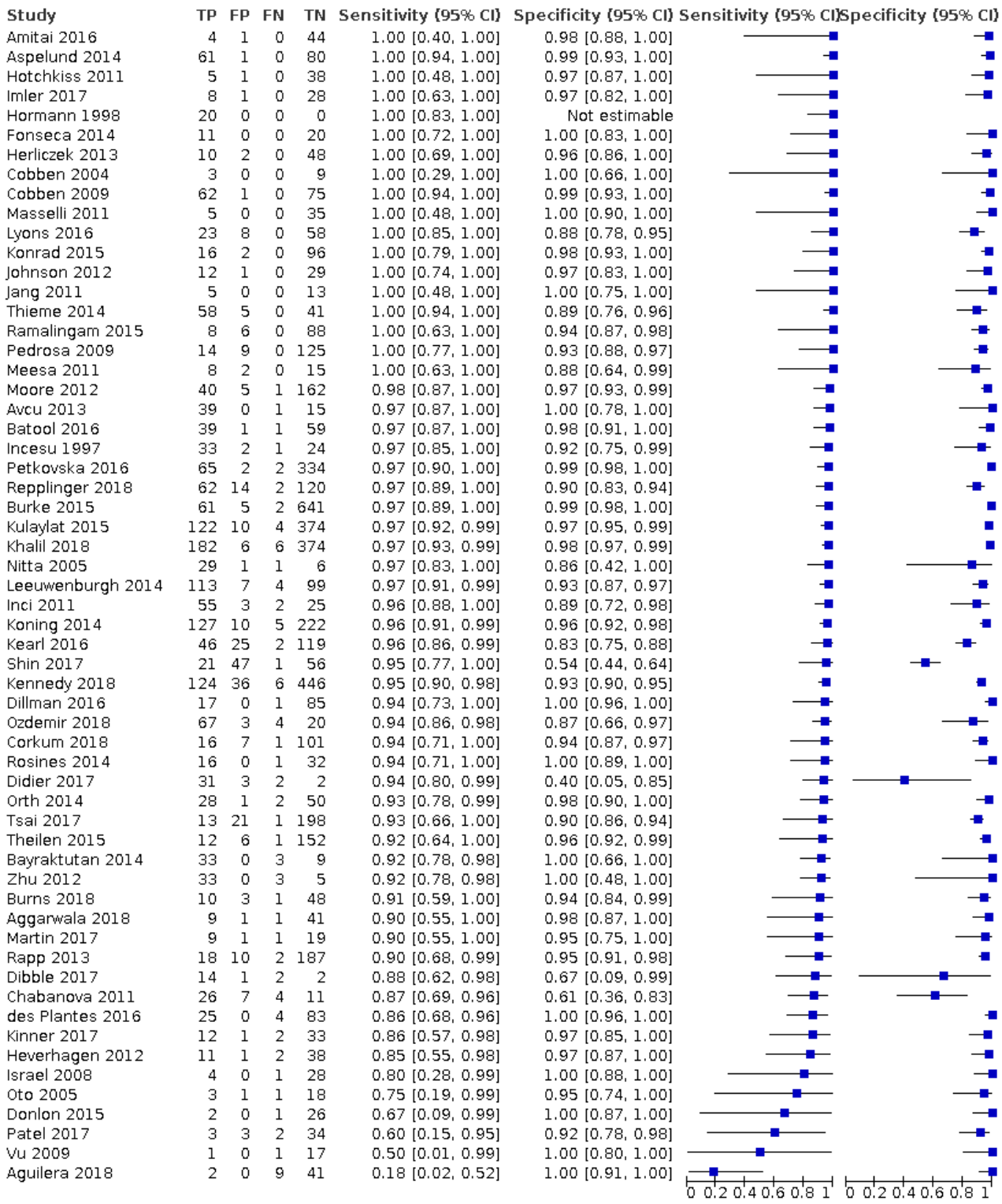


Figure 4. Flowchart of test performance in a theoretical cohort of 1000 patients with suspected appendicitis



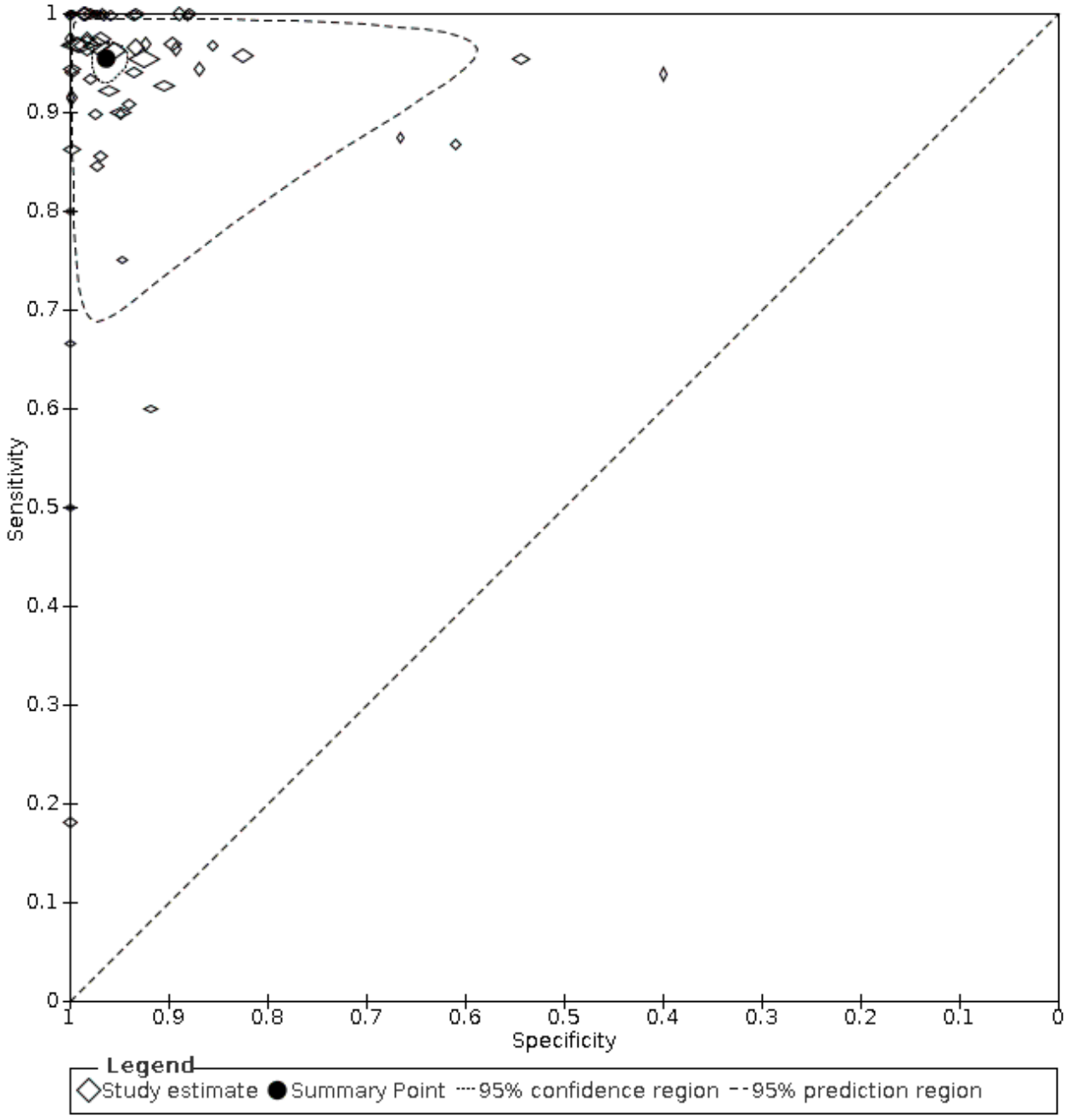
tp: true positive – test is positive (indicates appendicitis) and patient has appendicitis
fp: false positive – test is positive (indicates appendicitis) but patient does not have appendicitis
tn: true negative – test is negative (indicates no appendicitis) and patient does not have appendicitis
fn: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

Figure 5. Forest plot of MRI for appendicitis.



Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amitai 2016	4	1	0	44	1.00 [0.40, 1.00]	0.98 [0.88, 1.00]
Aspelund 2014	61	1	0	80	1.00 [0.94, 1.00]	0.99 [0.93, 1.00]
Hotchkiss 2011	5	1	0	38	1.00 [0.48, 1.00]	0.97 [0.87, 1.00]
Imler 2017	8	1	0	28	1.00 [0.63, 1.00]	0.97 [0.82, 1.00]
Hormann 1998	20	0	0	0	1.00 [0.83, 1.00]	Not estimable
Fonseca 2014	11	0	0	20	1.00 [0.72, 1.00]	1.00 [0.83, 1.00]
Herliczek 2013	10	2	0	48	1.00 [0.69, 1.00]	0.96 [0.86, 1.00]
Cobben 2004	3	0	0	9	1.00 [0.29, 1.00]	1.00 [0.66, 1.00]
Cobben 2009	62	1	0	75	1.00 [0.94, 1.00]	0.99 [0.93, 1.00]
Masselli 2011	5	0	0	35	1.00 [0.48, 1.00]	1.00 [0.90, 1.00]
Lyons 2016	23	8	0	58	1.00 [0.85, 1.00]	0.88 [0.78, 0.95]
Konrad 2015	16	2	0	96	1.00 [0.79, 1.00]	0.98 [0.93, 1.00]
Johnson 2012	12	1	0	29	1.00 [0.74, 1.00]	0.97 [0.83, 1.00]
Jang 2011	5	0	0	13	1.00 [0.48, 1.00]	1.00 [0.75, 1.00]
Thieme 2014	58	5	0	41	1.00 [0.94, 1.00]	0.89 [0.76, 0.96]
Ramalingam 2015	8	6	0	88	1.00 [0.63, 1.00]	0.94 [0.87, 0.98]
Pedrosa 2009	14	9	0	125	1.00 [0.77, 1.00]	0.93 [0.88, 0.97]
Meesa 2011	8	2	0	15	1.00 [0.63, 1.00]	0.88 [0.64, 0.99]
Moore 2012	40	5	1	162	0.98 [0.87, 1.00]	0.97 [0.93, 0.99]
Avcu 2013	39	0	1	15	0.97 [0.87, 1.00]	1.00 [0.78, 1.00]
Batool 2016	39	1	1	59	0.97 [0.87, 1.00]	0.98 [0.91, 1.00]
Incesu 1997	33	2	1	24	0.97 [0.85, 1.00]	0.92 [0.75, 0.99]
Petkovska 2016	65	2	2	334	0.97 [0.90, 1.00]	0.99 [0.98, 1.00]
Repplinger 2018	62	14	2	120	0.97 [0.89, 1.00]	0.90 [0.83, 0.94]
Burke 2015	61	5	2	641	0.97 [0.89, 1.00]	0.99 [0.98, 1.00]
Kulaylat 2015	122	10	4	374	0.97 [0.92, 0.99]	0.97 [0.95, 0.99]
Khalil 2018	182	6	6	374	0.97 [0.93, 0.99]	0.98 [0.97, 0.99]
Nitta 2005	29	1	1	6	0.97 [0.83, 1.00]	0.86 [0.42, 1.00]
Leeuwenburgh 2014	113	7	4	99	0.97 [0.91, 0.99]	0.93 [0.87, 0.97]
Inci 2011	55	3	2	25	0.96 [0.88, 1.00]	0.89 [0.72, 0.98]
Koning 2014	127	10	5	222	0.96 [0.91, 0.99]	0.96 [0.92, 0.98]
Kearl 2016	46	25	2	119	0.96 [0.86, 0.99]	0.83 [0.75, 0.88]
Shin 2017	21	47	1	56	0.95 [0.77, 1.00]	0.54 [0.44, 0.64]
Kennedy 2018	124	36	6	446	0.95 [0.90, 0.98]	0.93 [0.90, 0.95]
Dillman 2016	17	0	1	85	0.94 [0.73, 1.00]	1.00 [0.96, 1.00]
Ozdemir 2018	67	3	4	20	0.94 [0.86, 0.98]	0.87 [0.66, 0.97]
Corkum 2018	16	7	1	101	0.94 [0.71, 1.00]	0.94 [0.87, 0.97]
Rosines 2014	16	0	1	32	0.94 [0.71, 1.00]	1.00 [0.89, 1.00]
Didier 2017	31	3	2	2	0.94 [0.80, 0.99]	0.40 [0.05, 0.85]
Orth 2014	28	1	2	50	0.93 [0.78, 0.99]	0.98 [0.90, 1.00]
Tsai 2017	13	21	1	198	0.93 [0.66, 1.00]	0.90 [0.86, 0.94]
Theilen 2015	12	6	1	152	0.92 [0.64, 1.00]	0.96 [0.92, 0.99]
Bayraktutan 2014	33	0	3	9	0.92 [0.78, 0.98]	1.00 [0.66, 1.00]
Zhu 2012	33	0	3	5	0.92 [0.78, 0.98]	1.00 [0.48, 1.00]
Burns 2018	10	3	1	48	0.91 [0.59, 1.00]	0.94 [0.84, 0.99]
Aggarwala 2018	9	1	1	41	0.90 [0.55, 1.00]	0.98 [0.87, 1.00]
Martin 2017	9	1	1	19	0.90 [0.55, 1.00]	0.95 [0.75, 1.00]
Rapp 2013	18	10	2	187	0.90 [0.68, 0.99]	0.95 [0.91, 0.98]
Dibble 2017	14	1	2	2	0.88 [0.62, 0.98]	0.67 [0.09, 0.99]
Chabanova 2011	26	7	4	11	0.87 [0.69, 0.96]	0.61 [0.36, 0.83]
des Plantes 2016	25	0	4	83	0.86 [0.68, 0.96]	1.00 [0.96, 1.00]
Kinner 2017	12	1	2	33	0.86 [0.57, 0.98]	0.97 [0.85, 1.00]
Heverhagen 2012	11	1	2	38	0.85 [0.55, 0.98]	0.97 [0.87, 1.00]
Israel 2008	4	0	1	28	0.80 [0.28, 0.99]	1.00 [0.88, 1.00]
Oto 2005	3	1	1	18	0.75 [0.19, 0.99]	0.95 [0.74, 1.00]
Donlon 2015	2	0	1	26	0.67 [0.09, 0.99]	1.00 [0.87, 1.00]
Patel 2017	3	3	2	34	0.60 [0.15, 0.95]	0.92 [0.78, 0.98]
Vu 2009	1	0	1	17	0.50 [0.01, 0.99]	1.00 [0.80, 1.00]
Aguilera 2018	2	0	9	41	0.18 [0.02, 0.52]	1.00 [0.91, 1.00]

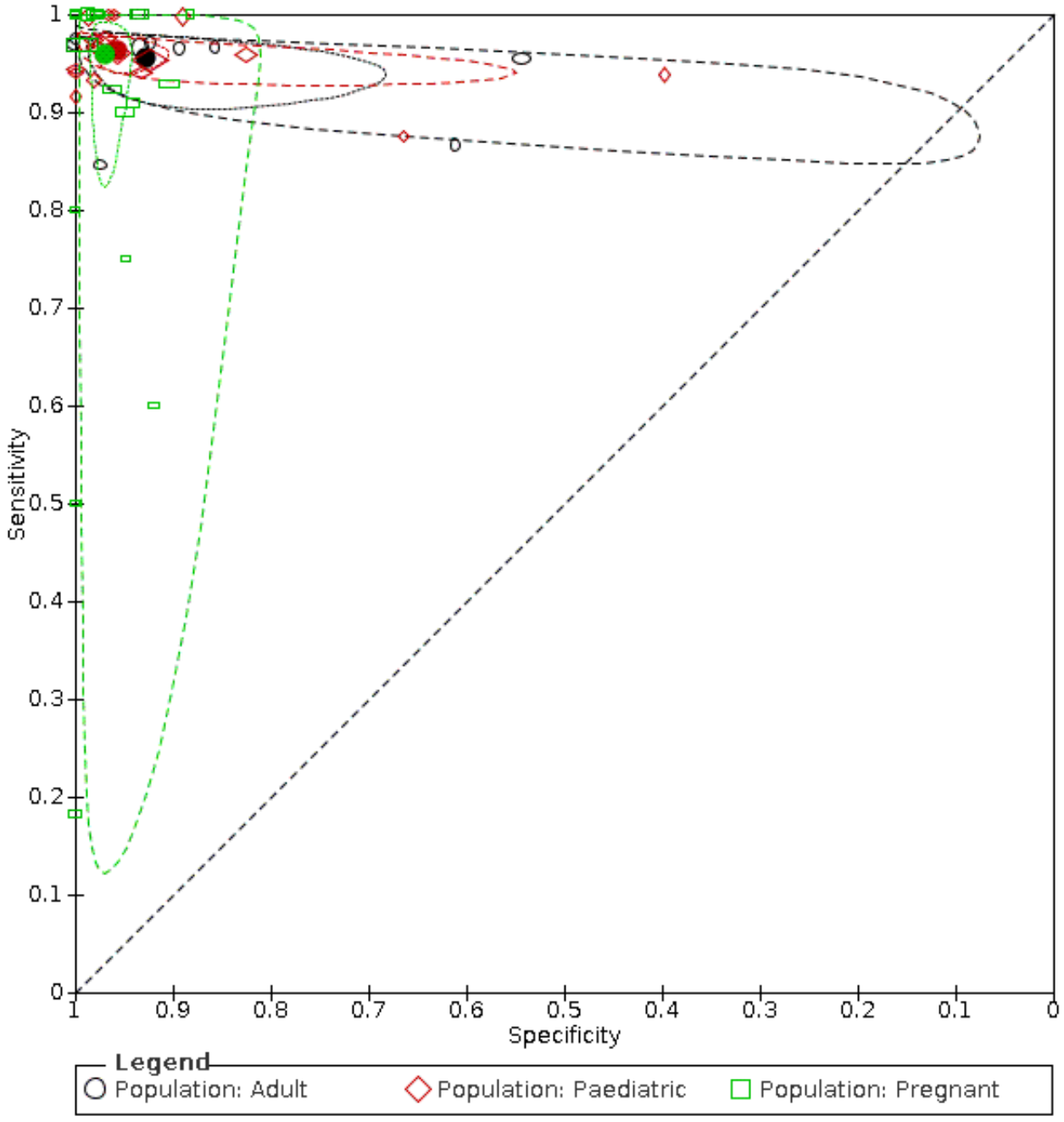
Figure 6. Summary ROC plot of MRI for appendicitis. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circle represents the summary sensitivity and specificity. This summary point is surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).



At the median pre-test appendicitis prevalence of 0.25, the post-test probability following a positive and a negative MRI result was 0.85 (95% CI 0.80 to 0.93) and 0.02 (95% CI 0.01 to 0.02), respectively. Likewise, at the minimum pre-test prevalence (0.06), the post-test probabilities were 0.53 (95% CI 0.62 to 0.71) and 0.00 (95% CI 0.00 to 0.00), respectively. At the maximum pre-test prevalence (0.88), the post-test probabilities were 0.99 (95% CI 0.99 to 1.00) and 0.26 (95% CI 0.20 to 0.33), respectively.

We performed the outlined subgroup analyses, with the following results (Figure 7):

Figure 7. Summary ROC plot: subgroup analyses in populations of adults, children, and pregnant women. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).



- In adults (9 studies, 1088 participants), the summary sensitivity and summary specificity were 0.96 (95% CI 0.93 to 0.97) and 0.93 (95% CI 0.80 to 0.98).
- In children (17 studies, 2794 children), the summary sensitivity and summary specificity were 0.96 (95% CI 0.95 to 0.97) and 0.96 (95% CI 0.92 to 0.98).
- In pregnant women (21 studies, 2282 women), the summary sensitivity and summary specificity were 0.96 (95% CI 0.88 to 0.99) and 0.97 (95% CI 0.95 to 0.98).

Investigations of heterogeneity

Primary study estimates of sensitivity and specificity were homogeneous on visual inspection of the ROC plot (Figure 6). We found no statistical evidence in meta-regression analyses that field of view (Figure 8), slice thickness (Figure 9), MRI sequences (Figure 10), use of contrast enhancement (Figure 11), or scan time (Figure 12) affected summary estimates of sensitivity and specificity (Table 1).

Figure 8. Summary ROC plot: analysis of effect of field of view. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).

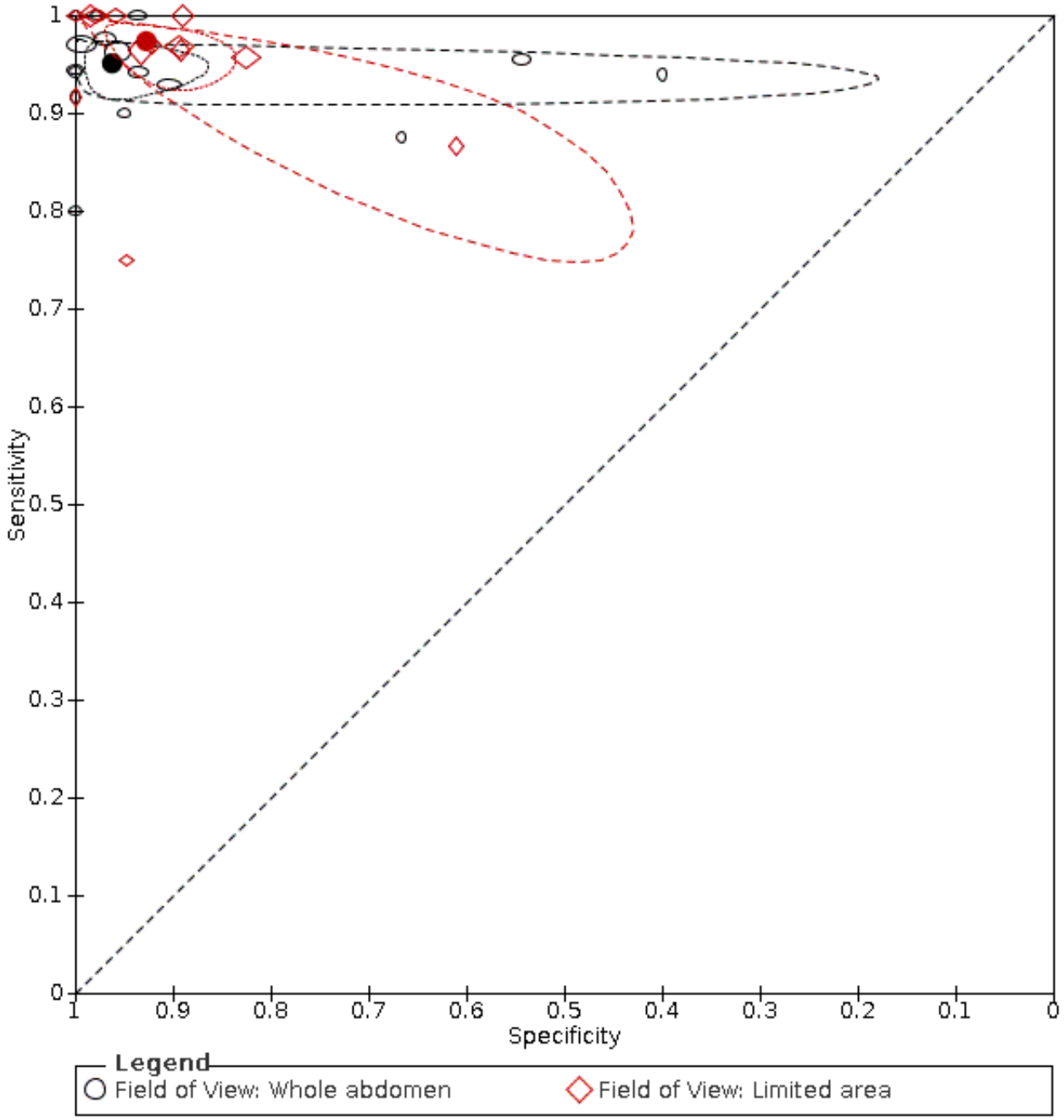


Figure 9. Summary ROC plot: analysis of effect of slice thickness. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).

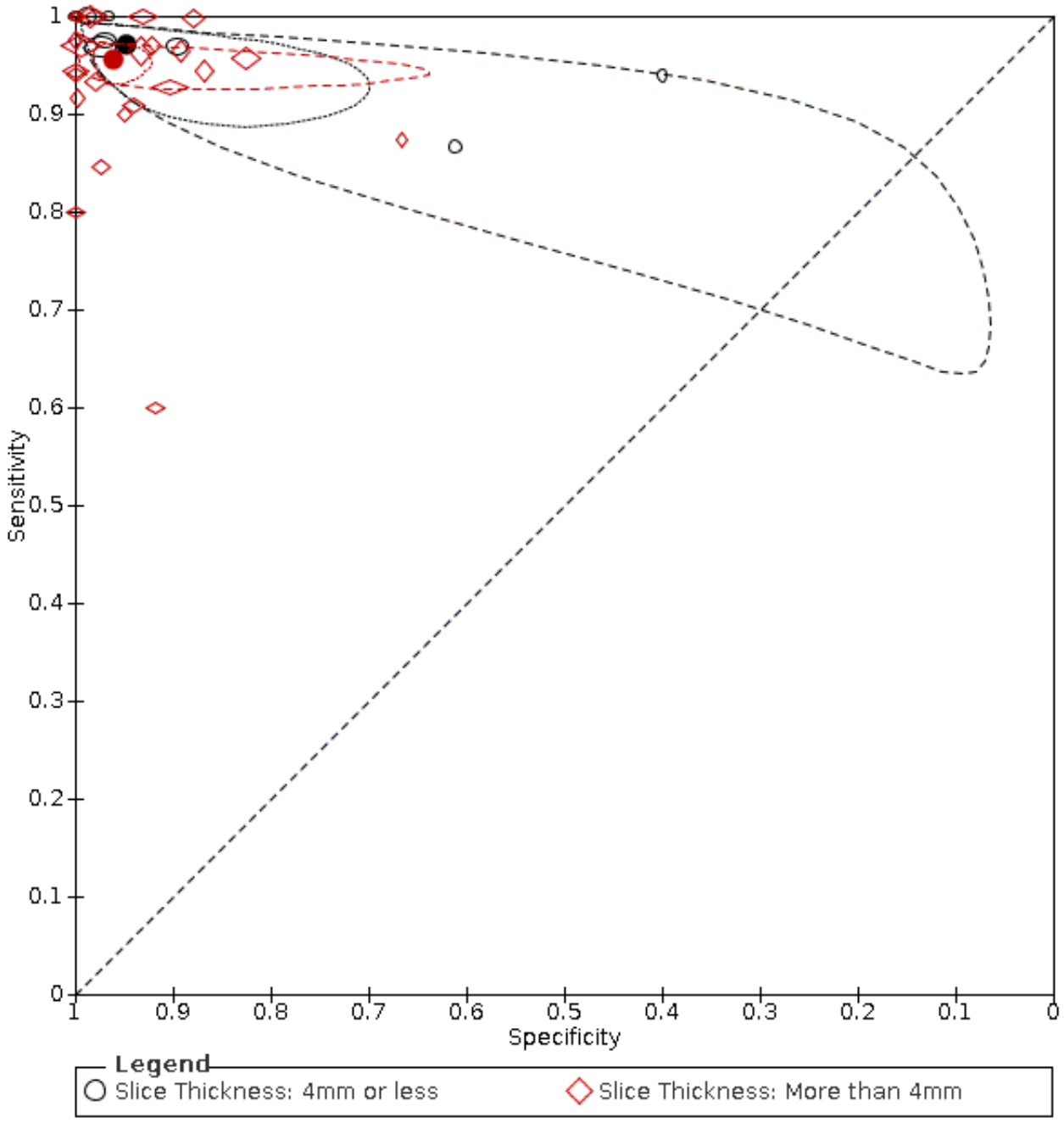


Figure 10. Summary ROC plot: analysis of the effect of MRI sequences. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).

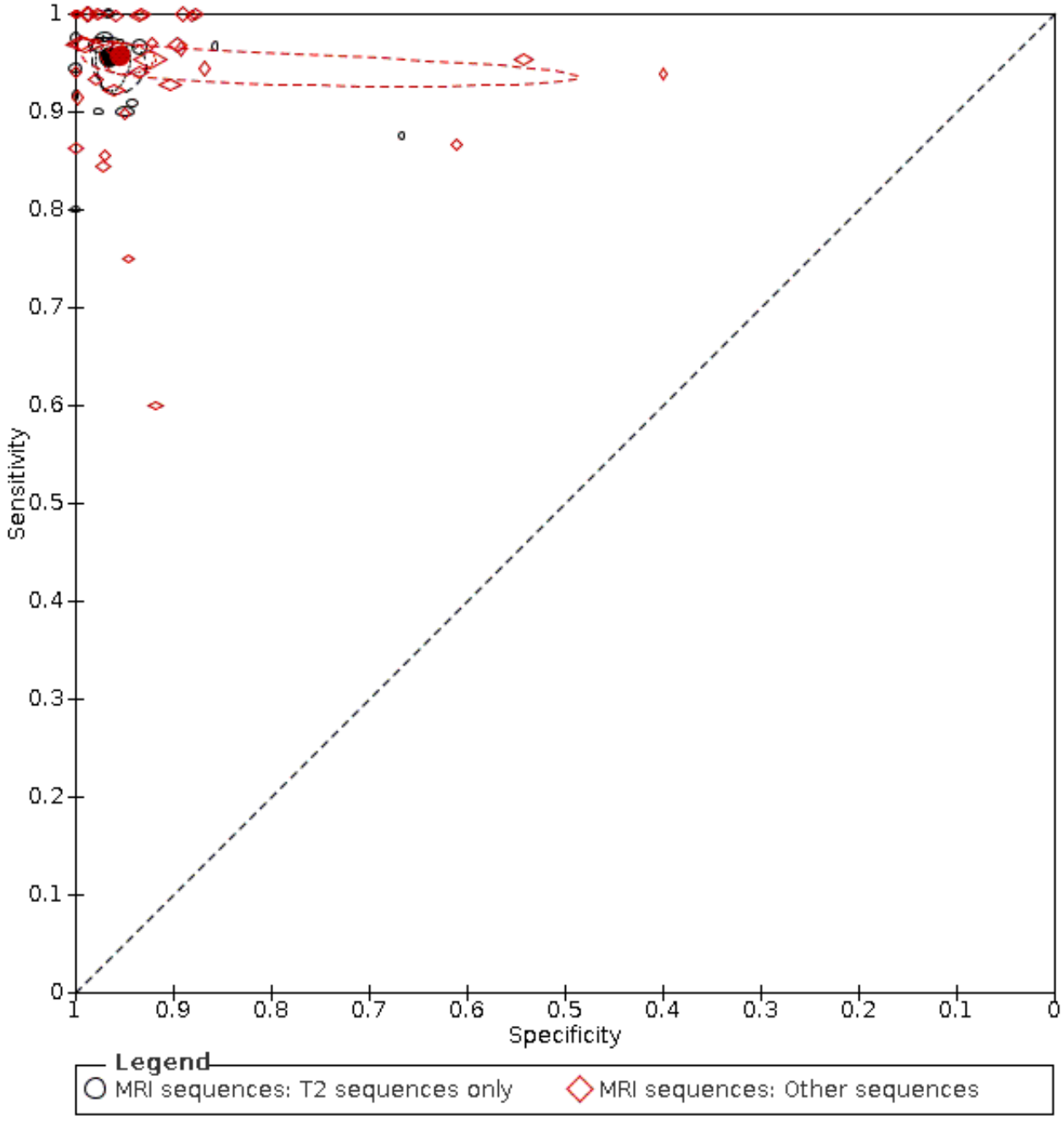


Figure 11. Summary ROC plot: analysis of the effect of contrast enhancement. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line).

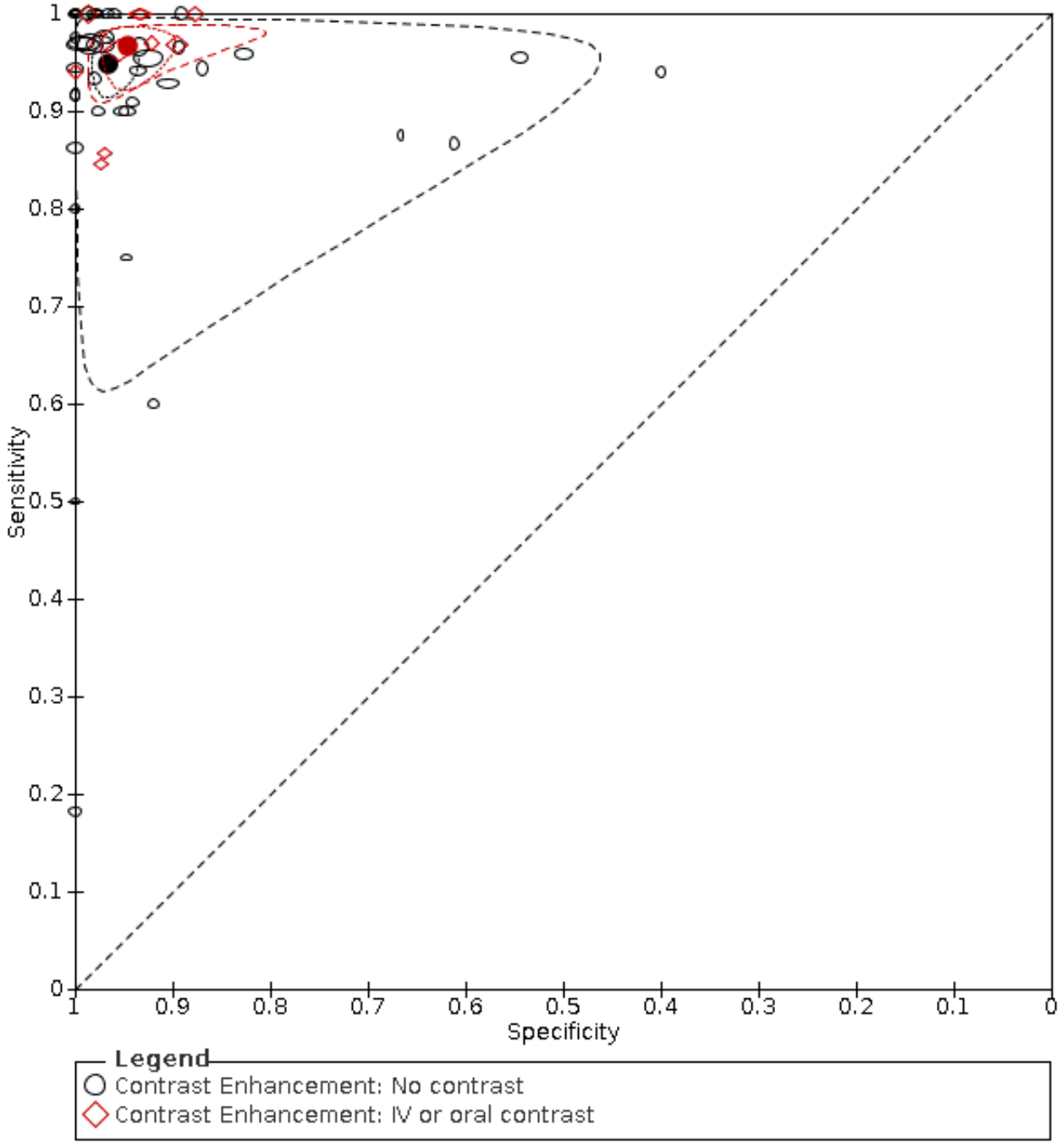
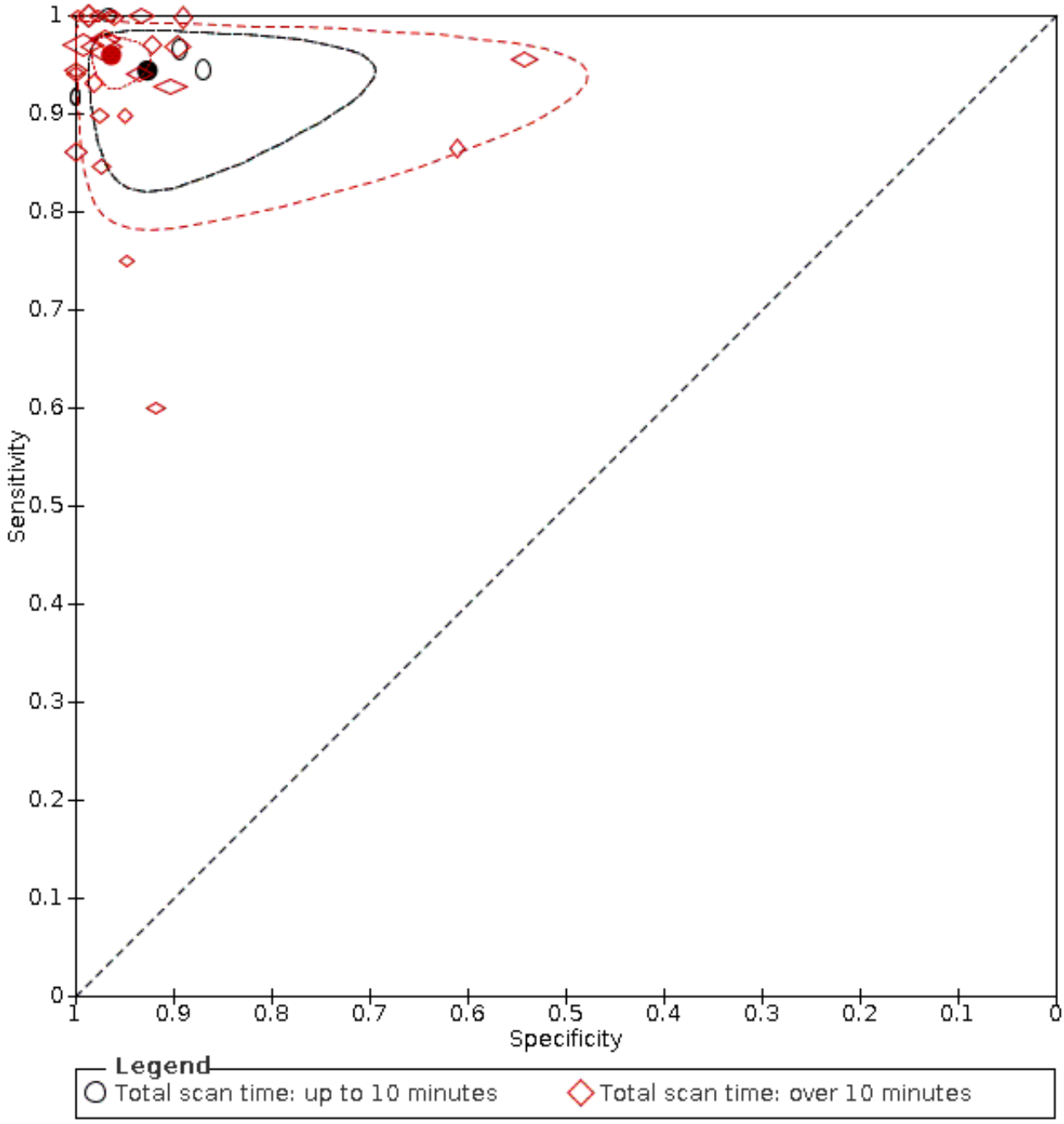


Figure 12. Summary ROC plot: analysis of the effect of scan time. The hollow symbols represent pairs of sensitivity and specificity from included studies. The symbol size is scaled according to the sample size of the study. The solid circles represent summary sensitivity and specificity. These summary points are surrounded by a 95% confidence region (dotted line) and a 95% prediction region (interrupted line). Confidence and prediction regions overlap for the 'up to 10 minutes' category due to extremely low variances of the random effects for logit sensitivity and logit specificity.



The effect of intravenous contrast and MRI sequences was investigated in three paired studies. Kinner and colleagues prospectively recruited 48 patients who underwent intravenous contrast-enhanced and unenhanced MRI (Kinner 2017). Sensitivity was higher for contrast-enhanced than for unenhanced MRI (0.94 versus 0.86), but the difference was not statistically significant. Specificity did not differ (0.94). In Lyons 2016, MRIs were reassessed in 89 patients who had undergone MRI with unenhanced and intravenous contrast-enhanced sequences. Sensitivity and specificity were higher for contrast-enhanced than for unenhanced

sequences: 1.0 versus 0.87 and 0.92 versus 0.79, respectively. In Rosines 2014, five radiologists reassessed MRIs from 49 patients who had undergone MRI with the following sequences: unenhanced and intravenous contrast-enhanced T1-weighted, T2-weighted, and balanced steady-state free precession. Mean sensitivity and specificity were 0.94 and 1.0 for contrast-enhanced T1-weighted sequences; 0.88 and 0.97 for T2-weighted sequences; and 0.81 and 0.94 for balanced steady-state free precession.

Sensitivity analyses

Influence of methodological quality

The sensitivity analyses are reported in Table 2. Summary sensitivity and specificity for 14 studies (2096 participants) with low risk of bias for domain 1 were 0.96 (95% CI 0.94 to 0.97) and 0.96 (95% CI 0.90 to 0.99), respectively. Summary sensitivity and specificity for 26 studies (3272 participants) with low risk of bias for domain 2 were 0.96 (95% CI 0.95 to 0.97) and 0.95 (95% CI 0.92 to 0.97), respectively. Summary sensitivity and specificity for 6 studies (819 participants) with low risk of bias for domain 3 were 0.96 (95% CI 0.92 to 0.98) and 0.94 (95% CI 0.90 to 0.97), respectively. A similar sensitivity analysis was not feasible for domain 4 because only one study was at low risk of bias. These sensitivity analyses demonstrate that our assessment of methodological quality did not influence the summary estimates. Likewise, when we excluded three studies at low risk of bias for domains 1, 2, and 3 (Herliczek 2013; Koning 2014; Thieme 2014), summary estimates did not change.

Other sensitivity analyses

A substantial proportion of the included studies had low numbers of participants. As estimates of sensitivity and specificity can be extreme in such studies due to chance variation, we performed a sensitivity analysis that excluded 13 studies with fewer than 40 participants. Summary sensitivity and specificity in the remaining 45 studies (7111 participants) were 0.96 (95% CI 0.94 to 0.97) and 0.96 (95% CI 0.95 to 0.97), respectively. These estimates were almost identical to the overall results, hence studies with a low number of participants do not appear to influence the overall results.

We looked closely at eight studies with outlying results defined as sensitivity or specificity below 0.7. A likely explanation for the low sensitivity in Donlon 2015, Patel 2017, Vu 2009, and Aguilera 2018 was the play of chance due to very low numbers of participants with appendicitis in these studies (n = 3, n = 5, n = 2, and n = 11, respectively). Likewise, the low specificity in Chabanova 2011, Dibble 2017, and Didier 2017 was probably related to low numbers of participants without appendicitis in these studies (n = 18, n = 3, and n = 5, respectively). By contrast, the low specificity in Shin 2017 did not appear to be explained by the number of participants without appendicitis (n = 103), but rather the use of a novel MRI sign (the T1 bright appendix sign), which was used in isolation to exclude appendicitis, without considering other criteria for appendicitis. We performed a sensitivity analysis by excluding these eight studies with outlying results. Summary sensitivity of the remaining 50 studies (7090 participants) was 0.96 (95% CI 0.95 to 0.97), and summary specificity was 0.97 (95% CI 0.95 to 0.97). Hence, the influence of studies with outlying results was marginal.

MRI was used as a second-line imaging test following negative or equivocal US in 12 studies (Amitai 2016; Dibble 2017; Dillman 2016; Fonseca 2014; Herliczek 2013; Konrad 2015; Lyons 2016; Martin 2017; Masselli 2011; Ramalingam 2015; Rosines 2014; Vu 2009). Summary estimates of sensitivity and specificity for these 12 studies (705 participants) were 0.96 (95% CI 0.90 to 0.98) and 0.98 (95% CI 0.94 to 0.99), respectively. These estimates were marginally higher than the summary estimates in the overall results.

The sensitivity analyses above were not preplanned in the protocol (see Differences between protocol and review).

DISCUSSION

Summary of main results

This review builds on the results of previously published meta-analyses on the same topic, confirming that MRI appears to be a highly accurate test for diagnosing appendicitis. However, the methodological quality of the included studies was generally poor due to inadequate and incomplete follow-up in participants who did not have surgery.

The results of our meta-analysis are summarised in Summary of findings 1. The meta-analysis of 58 studies with 7462 participants reported a summary sensitivity and summary specificity of MRI for appendicitis of 0.95 (95% CI 0.94 to 0.97) and 0.96 (95% CI 0.95 to 0.97), respectively. Summary estimates of sensitivity and specificity only differed slightly between subgroups of unselected adult participants, paediatric participants, and pregnant women. None of the MRI protocol variables, nor the risk of bias across QUADAS domains, influenced summary estimates of sensitivity and specificity in any meaningful way.

Strengths and weaknesses of the review

We employed a comprehensive literature search and review methods recommended by Cochrane. Three study team members independently identified 59 relevant studies, from which data were extracted. We performed a thorough quality assessment, which yielded different results to previous meta-analyses. We updated our search results during the review; as a result more studies are included in this meta-analysis than in any published review to date (Barger 2010; Blumenfeld 2011; Duke 2016; Kave 2019; Moore 2016; Repplinger 2016). We further identified and included unpublished studies in the grey literature, such as Batool 2016, Donlon 2015, Hotchkiss 2011. Other studies, usually conference abstracts, on the diagnostic accuracy of appendicitis were retrieved but lacked the raw data required for a 2 x 2 table (including, but not limited to, Aronberg 2017, Bernbeck 2015, and Byott 2016; see Characteristics of excluded studies). We also explored the accuracy of MRI across different populations (adults, children, and pregnant women), and assessed a range of MRI protocol variables that could potentially affect diagnostic accuracy.

The significant limitation to the review was the overall methodological weakness of the included studies and low standards of reporting. Although concern for applicability was low, risk of bias was high in our assessment. Essentially, this means that whilst these studies were conducted within a relevant clinical setting using typical patients with suspected appendicitis, the summary estimates may not be representative of the accuracy of MRI for diagnosing appendicitis in clinical practice. In our view, the finding that summary estimates did not change when three studies with low risk of bias were excluded, Herliczek 2013; Koning 2014; Thieme 2014, does not imply that the impact of low methodological

quality was negligible. On the contrary, the paucity of studies with low risk of bias for all domains prevented us from assessing whether potential bias from low methodological quality impacted the summary estimates.

Despite the retrospective design of most studies, the QUADAS-2 tool suggests low potential for selection bias if a consecutive sample of patients are enrolled. The applicability of patient selection was deemed as low concern when patients with clinically suspected appendicitis who had MRI were recruited for the study. The main methodological problems related to the reference standard and flow and timing domains. The reference standard was considered inadequate due to insufficient follow-up in participants who did not have surgery. The majority of studies were retrospective, and follow-up was usually based on case note review. This is problematic for two reasons. First, an alternative diagnosis (e.g. diverticulitis, pelvic inflammatory disease, ureter stone) may rule out appendicitis is some, but not all, participants who did not have surgery, particularly when the alternative diagnosis is non-specific abdominal pain. Second, review of case notes will not capture cases of missed appendicitis when patients seek treatment elsewhere. The numbers of false-negative MRI results in studies was generally very low (median 1, interquartile range 0 to 2), and the numbers of study participants was also low in a substantial proportion of the included studies. Hence, even one or two participants with negative MRI results misclassified as true-negatives could have a substantial influence on sensitivity. The influence on specificity would tend to be less pronounced, unless studies are small and prevalence is high.

In the flow and timing domain, the problems were differential verification and insufficient reporting in prospective studies on the proportion of participants who had follow-up as planned. In most studies, the majority of participants with a positive MRI result had surgery, whereas most participants with a negative MRI result had follow-up because it was considered unethical to expose these participants to surgery that was potentially harmful and unlikely to be necessary. In our view, the likely consequence of low-quality follow-up and loss to follow-up was partial verification. Unfortunately, a sensitivity analysis of the effect of this was unfeasible due to a lack of studies with low risk of bias for domain 4. Partial verification has been associated with higher estimates of sensitivity in diagnostic accuracy studies in general (Whiting 2013), and we believe that it is reasonable to suspect that a similar association could exist in this review.

Another limitation relates to the low number of studies that compared MRI protocols using a paired or a randomised study design. We identified three studies that compared the accuracy of protocols with and without intravenous contrast-enhanced sequences using a paired study design. Sensitivity in these studies was generally higher for intravenous contrast-enhanced sequences. A corresponding difference was not demonstrated in our meta-regression analysis on the effect of contrast enhancement, which also included oral contrast enhancement. However, the results of the meta-regression analyses should be interpreted cautiously, because comparisons in such analyses are subject to confounding by other factors such as population characteristics and study methods. Hence, although no differences in sensitivity or specificity were demonstrated for MRI protocol variables in heterogeneity analyses, this does not rule out the existence of such differences.

A limitation in the planning of the review was that we did not consider complicated appendicitis a separate target condition. However, only three studies investigated the accuracy of MRI in distinguishing between simple and complicated appendicitis (appendicitis with perforation or abscess formation) (Church 2016; Leeuwenburgh 2014; Rosenbaum 2017). This is relevant in light of trends towards non-operative management of simple appendicitis and the consequent necessity of imaging to rule out features of complicated appendicitis. Church 2016 reported sensitivity 0.87 and specificity 0.74 for MRI in separating complicated from simple appendicitis in 135 participants who had an appendicectomy. A similar study reported sensitivity 0.82 and specificity 0.85 (Rosenbaum 2017). Leeuwenburgh 2014 reported much lower values of sensitivity and specificity (0.57 and 0.86, respectively).

Applicability of findings to the review question

Participants were recruited in acute, emergency settings. The majority of studies were conducted in teaching hospitals. MRI scanners are not present within all hospitals, and when they are, emergency or out-of-hours MRI services may be limited. This has prevented enrolment of up to 60% of patients in some studies (Leeuwenburgh 2014). Our concern for applicability related to patient selection was generally low (51/59, 86%), as participants were largely included due to suspected appendicitis based on history, clinical examination, blood tests, and urinalysis without inappropriate exclusion criteria. Nevertheless, prevalence of appendicitis varied widely (5% to 100%), which reflects the higher and lower risk of appendicitis dependent on the selection criteria used in the primary studies. The prevalence of appendicitis was highest (62% to 100%) in studies where patients scheduled for surgery were recruited (Chabanova 2011; Hormann 1998; Zhu 2012), and lowest in pregnant women, reflecting a low threshold for MRI in this population. Our analyses demonstrated little to no variation in accuracy across subgroups of adults, children, pregnant women, and participants who had MRI subsequent to a negative or equivocal US.

There was low concern for applicability of the index test (48/59, 81%). Study results were homogeneous despite the wide variation in MRI scanner generation, MRI sequences, field of view, slice thickness, use of contrast enhancement, or subspecialty interest.

There was low concern for applicability of the reference standard (55/59, 93%) of histology or adequate follow-up because it reflected clinical practice. However, as stated above, the quality of the reference standard employed in most retrospective studies was compromised by poor standards of follow-up. Whilst case note review may have been employed by retrospective studies on pragmatic grounds, it is an inadequate methodology to exclude appendicitis. The potential bias introduced by inadequate and incomplete follow-up was limited to participants that did not have appendicitis on MRI. Participants with appendicitis on MRI underwent surgery with appendicectomy and histological examination of the resected appendix. By contrast, participants without signs of appendicitis on MRI were unlikely to have undergone surgery, and the reference standard in these participants consisted of follow-up. The negative predictive value of MRI for appendicitis could therefore be particularly biased from inadequate and incomplete follow-up. This finding is important for the interpretation of negative MRI findings in clinical practice.

AUTHORS' CONCLUSIONS

Implications for practice

Magnetic resonance imaging (MRI) appears to be highly accurate in confirming and excluding acute appendicitis in adults, children, and pregnant women regardless of protocol, in keeping with results from previous meta-analyses (Barger 2010; Blumenfeld 2011; Duke 2016; Kave 2019; Moore 2016; Repplinger 2016). However, the methodological quality of the included studies was generally low due to incomplete and low standards of follow-up, so summary estimates of sensitivity and specificity may be biased. Due to the very low number of high-quality studies, we could not assess the impact and direction of potential bias. Studies comparing MRI protocols were few, and although we found no influence of MRI protocol variables on the summary estimates of accuracy, our results do not rule out that some MRI protocols are more accurate than others.

Implications for research

Based on the findings of this review, we consider the following issues most important for future research.

Methodological quality

In the design and conduct of future studies, the priority would be to ensure adequacy of follow-up method and duration. As differential verification in this area of diagnostic research appears inevitable, follow-up should aim to reliably rule out appendicitis in patients that do not have surgery. We believe that a follow-up period of seven to 31 days is sufficiently long to capture missed cases and sufficiently short so that new events are not captured. Follow-up should ideally be performed by clinicians who were not part of the surgical team and who were blinded to the MRI report. To improve feasibility and ensure compliance, follow-up could be performed over email or telephone. When loss to follow-up does occur, it must be quantified within the study manuscript. Other steps to improve reporting should comply with STARD (Standards for Reporting Diagnostic accuracy studies) guidelines (Bossuyt 2015). Patient enrolment should ideally be consecutive or random, with clear eligibility and appropriate exclusion criteria.

MRI protocol

Diagnostic accuracy of MRI did not vary by MRI protocol variables including field of view, slice thickness, MRI sequence, use of contrast enhancement, or scan time (Table 1). However, as stated, these findings do not necessarily imply that all protocols are equally accurate. In this context, it is notable that higher sensitivity was found with intravenous contrast-enhanced sequences in three paired studies compared to unenhanced protocols. However, abbreviated T2-only imaging protocols shortened scan time to less than 5 minutes (Bayraktutan 2014; Israel 2008; Johnson 2012; Zhu 2012), whilst maintaining diagnostic accuracy. Rosines and colleagues found that T2 weighted sequences and sequences with balanced steady-state free precession did not provide additional accuracy compared to intravenous contrast-enhanced T1 weighted sequences in a study in children (Rosines 2014). Other studies have shown 100% scan completion in paediatric patients aged 4 to 17 with no sedation (Johnson 2012), and 95% completion in sedated infants less than one year old (Bayraktutan 2014).

Future studies should address the issue of the optimal MRI protocol. Given the wide number of protocol variables and the predominance of single-centre retrospective studies, further trial design should highlight the need for abbreviated, quick T2-only protocols, which would maximise scan completion rates. In children this means the weighing of accuracy against scan time and the need for contrast enhancement. In pregnant women this means weighing accuracy with unknown risks to the fetus from different amounts of radiofrequency energy that different protocols may give, as well as the recommendation that gadolinium contrast not be used in pregnancy. Such studies should have a paired design with prospective data collection where two or more MRI protocols are evaluated in the same study population to minimise potential confounding, or alternatively they should have a randomised design.

Simple versus complicated appendicitis

Non-operative management of simple appendicitis, with, CODA 2020, D'Souza 2014, Sallinen 2016, or without, Park 2017, antibiotics, continues to accumulate in the literature. Imaging confirmation of simple appendicitis (i.e. no sign of abscess or perforation) is required prior to non-operative management with antibiotics (CODA 2020; Salminen 2015; Vons 2011). At present, there are only three studies that investigate the ability of MRI to distinguish simple from complicated appendicitis (Church 2016; Leeuwenburgh 2014; Rosenbaum 2017). If the decision about non-operative management is to be based on MRI findings, then there is a need for more studies that evaluate the accuracy of MRI in differentiating simple from complicated appendicitis. The design of such studies should also address the challenges of difficulties in disease verification when studies include antibiotic therapy as a treatment arm; histology as a reference standard may only be possible when patients fail antibiotic treatment and undergo surgery.

ACKNOWLEDGEMENTS

We would like to thank Sys Johnsen (Information Specialist for the Cochrane Colorectal Cancer Group) for her help in devising the search terms. Scott Steele and Henning Keinke Andersen (former Managing Editor of the Cochrane Colorectal Cancer Group) performed the initial peer review; we employed their many excellent comments in the editing of this manuscript. The Cochrane DTA Editorial team in Birmingham also made helpful comments on the methodological section of the protocol. Lisa Winer's work as copy editor was extremely helpful. Finally, we would like to thank Anthony Thavenirathan for his considerable input when drafting the protocol, filtering search results, and extracting data.

REFERENCES

References to studies included in this review

Aggarwala 2018 *{published data only}*

Aggarwala S, Choi J, Nol J, Jones A, Moscova M, Wong P. The role of a rapid protocol abdominal MRI for patients with clinically suspected appendicitis. *Journal of Medical Imaging and Radiation Oncology* 2018;**62**:115.

Aguilera 2018 *{published data only}*

Aguilera F, Gilchrist B, Farkas D. Accuracy of MRI in diagnosing appendicitis during pregnancy. *American Surgeon* 2018;**84**:1326-8.

Amitai 2016 *{published data only}*

Amitai MM, Eldad Katorza M, Larisa Guranda M, Sara Apter M, Orith Portnoy M, Yael Inbar M. Role of emergency magnetic resonance imaging for the workup of suspected appendicitis in pregnant women. *Israel Medical Association Journal* 2016;**18**(10):600-4.

Aspelund 2014 *{published data only}*

Aspelund G, Fingeret A, Gross E, Kessler D, Keung C, Thirumoorthi A, et al. Ultrasonography/MRI versus CT for diagnosing appendicitis. *Pediatrics* 2014;**133**:586-93.

Avcu 2013 *{published data only}*

Avcu S, Cetin FA, Arslan H, Kemik O, Dulger AC. The value of diffusion-weighted imaging and apparent diffusion coefficient quantification in the diagnosis of perforated and nonperforated appendicitis. *Diagnostic and Interventional Radiology (Ankara, Turkey)* 2013;**19**(2):106-10. [PMID: 23266968]

Batool 2016 *{published data only}*

Batool S, Thomson G, Ortiz C, Bhandal S, Gupta P. Diagnostic accuracy of MR imaging for evaluation of acute appendicitis: a prime prospective study of Canadian pediatric population. *Pediatric Radiology* 2016;(Suppl):S125.

Bayraktutan 2014 *{published data only}*

Bayraktutan U, Oral A, Kantarci M, Demir M, Ogul H, Yalcin A, et al. Diagnostic performance of diffusion-weighted MR imaging in detecting acute appendicitis in children: comparison with conventional MRI and surgical findings. *Journal of Magnetic Resonance Imaging* 2014;**39**:1518-24.

Burke 2015 *{published data only}*

Burke LMB, Bashir MR, Miller FH, Siegelman ES, Brown M, Alobaidy M. Magnetic resonance imaging of acute appendicitis in pregnancy: a 5-year multiinstitutional study. *American Journal of Obstetrics and Gynecology* 2015;**213**(5):693.e1-6.

Burns 2018 *{published data only}*

Burns M, Hague C, Vos P, Tiwari P, Wiseman S. Utility of magnetic resonance imaging for the diagnosis of appendicitis during pregnancy: a Canadian experience. *Canadian Association of Radiologists Journal* 2017;**68**:392-400.

Chabanova 2011 *{published data only}*

Chabanova E, Balslev I, Achiam M, Nielsen YW, Adamsen S, Gocht-Jensen P, et al. Unenhanced MR imaging in adults with clinically suspected acute appendicitis. *European Journal of Radiology* 2011;**79**:206-10.

Cobben 2004 *{published data only}*

Cobben LP, Groot I, Haans L, Blickman JG, Puylaert J. MRI for clinically suspected appendicitis during pregnancy. *American Journal of Roentgenology* 2004;**183**:671-5.

Cobben 2009 *{published data only}*

Cobben L, Groot I, Kingma L, Coerkamp E, Puylaert J, Blickman J. A simple MRI protocol in patients with clinically suspected appendicitis: results in 138 patients and effect on outcome of appendectomy. *European Radiology* 2009;**19**:1175-83.

Corkum 2018 *{published data only}*

Corkum K, Oyetunji T, Grabowski J, Rigsby C, Lautz T. Absolute neutrophil count as a diagnostic guide for the use of MRI in the workup of suspected appendicitis in children. *Journal of Pediatric Surgery* 2018;**54**(7):1359-64.

des Plantes 2016 *{published data only}*

des Plantes CZ, van Veen M, van der Palen J, Klaase J, Gielkens H, Geelkerken R. The effect of unenhanced MRI on the surgeons? Decision-making process in females with suspected appendicitis. *World Journal of Surgery* 2016;**40**(12):2881-7.

Dibble 2017 *{published data only}*

Dibble E, Swenson D, Cartagena C, Baird G, Herliczek T. Effectiveness of a staged US and unenhanced MR imaging algorithm in the diagnosis of pediatric appendicitis. *Radiology* 2018;**286**:1022-9.

Didier 2017 *{published data only}*

Didier R, Hopkins K, Coakley F, Krishnaswami S, Spiro D, Foster B. Performance characteristics of magnetic resonance imaging without contrast agents or sedation in pediatric appendicitis. *Pediatric Radiology* 2017;**47**(10):1312-20.

Dillman 2016 *{published data only}*

Dillman JR, Gadepalli S, Sroufe NS, Davenport MS, Smith EA, Chong ST. Equivocal pediatric appendicitis: unenhanced MR imaging protocol for nonsedated children? A clinical effectiveness study. *Radiology* 2016;**279**(1):216-25.

Donlon 2015 *{published data only}*

Donlon N, Cormick P, Larkin J, Mehigan B. The use of MRI to diagnose appendicitis during pregnancy. *Colorectal Disease* 2015;**17**:51.

Fonseca 2014 *{published data only}*

Fonseca AL, Schuster KM, Kaplan LJ, Maung AA, Lui FY, Davis KA. The use of magnetic resonance imaging in the diagnosis of suspected appendicitis in pregnancy: shortened length of stay without increase in hospital charges. *JAMA Surgery* 2014;**149**:687-93.

Herliczek 2013 *{published data only}*

Herliczek TW, Swenson DW, Mayo-Smith WW. Utility of MRI after inconclusive ultrasound in pediatric patients with suspected appendicitis: retrospective review of 60 consecutive patients. *American Journal of Roentgenology* 2013;**200**:969-73.

Heverhagen 2012 *{published data only}*

Heverhagen JT, Pfestroff K, Heverhagen AE, Klose KJ, Kessler K, Sitter H. Diagnostic accuracy of magnetic resonance imaging: a prospective evaluation of patients with suspected appendicitis. *Journal of Magnetic Resonance Imaging* 2012;**35**:617-23.

Hormann 1998 *{published data only}*

Hormann M, Paya K, Eibenberger K, Dorffner R, Lang S, Kreuzer S. MR imaging in children with nonperforated acute appendicitis: value of unenhanced MR imaging in sonographically selected cases. *American Journal of Roentgenology* 1998;**171**:467-70.

Hotchkiss 2011 *{published data only}*

Hotchkiss J, Roberge E, Graham J. The diagnostic accuracy of MRI in the evaluation of acute right lower quadrant pain during pregnancy. *Emergency Radiology* 2011;**18**:453-81.

Imler 2017 *{published data only}*

Imler D, Keller C, Sivasankar S, Wang NE, Vasanawala S, Bruzoni M. Magnetic resonance imaging versus ultrasound as the initial imaging modality for pediatric and young adult patients with suspected appendicitis. *Academic Emergency Medicine* 2017;**24**(5):569-77.

Incesu 1997 *{published data only}*

Incesu L, Coskun A, Selcuk MB, Akan H, Sozubir S, Bernay F. Acute appendicitis: MR imaging and sonographic correlation. *American Journal of Roentgenology* 1997;**168**:669-74.

Inci 2011 *{published data only}*

Inci E, Kilickesmez O, Hocaoglu E, Aydin S, Bayramoglu S, Cimilli T. Utility of diffusion-weighted imaging in the diagnosis of acute appendicitis. European Radiology 2011;**21**:768-75.

Israel 2008 *{published data only}*

Israel GM, Malguria N, McCarthy S, Copel J, Weinreb J. MRI vs. ultrasound for suspected appendicitis during pregnancy. *Journal of Magnetic Resonance Imaging* 2008;**28**:428-33.

Jang 2011 *{published data only}*

Jang KM, Kim SH, Choi D, Lee SJ, Rhim H, Park MJ. The value of 3D T1-weighted gradient-echo MR imaging for evaluation of the appendix during pregnancy: preliminary results. *Acta Radiologica* 2011;**52**:825-8.

Johnson 2012 *{published data only}*

Johnson AK, Filippi CG, Andrews T, Higgins T, Tam J, Keating D, et al. Ultrafast 3-T MRI in the evaluation of children with acute lower abdominal pain for the detection of appendicitis. *American Journal of Roentgenology* 2012;**198**:1424-30.

Kearl 2016 *{published data only}*

Kearl LY, Claudius I, Behar S, Cooper J, Dollbaum R, Hardasmalani M. Accuracy of magnetic resonance imaging and ultrasound for appendicitis in diagnostic and nondiagnostic studies. *Academic Emergency Medicine* 2016;**23**:179-85.

Kennedy 2018 *{published data only}*

Kennedy TM, Thompson AD, Choudhary AK, Caplan RJ, Schenker KE, DePiero AD. Utility of applying white blood cell cutoffs to non-diagnostic MRI and ultrasound studies for suspected pediatric appendicitis. *American Journal of Emergency Medicine* 2019;**37**(9):1723-8.

Khalil 2018 *{published data only}*

Khalil S, Tripathee N, Skaug M, Shiralkar K, Surabhi V. Accuracy of unenhanced MR imaging in the detection of perforated acute appendicitis in pediatric patients. *Abdominal Radiology* 2018;**43**:1880.

Kinner 2017 *{published data only}*

Kinner S, Pickhardt PJ, Riedesel EL, Gill KG, Robbins JB, Kitchin DR. Diagnostic accuracy of MRI versus CT for the evaluation of acute appendicitis in children and young adults. *American Journal of Roentgenology* 2017;**209**(4):911-9.

Koning 2014 *{published data only}*

Koning JL, Naheedy JH, Kruk PG. Diagnostic performance of contrast-enhanced MR for acute appendicitis and alternative causes of abdominal pain in children. *Pediatric Radiology* 2014;**44**:948-55.

Konrad 2015 *{published data only}*

Konrad J, Grand D, Lourenco A. MRI: first-line imaging modality for pregnant patients with suspected appendicitis. *Abdominal Imaging* 2015;**40**(8):3359-64.

Kulaylat 2015 *{published data only}*

Kulaylat AN, Moore MM, Engbrecht BW, Brian JM, Khaku A, Hollenbeak CS. An implemented MRI program to eliminate radiation from the evaluation of pediatric appendicitis. *Journal of Pediatric Surgery* 2015;**50**:1359-63.

Leeuwenburgh 2014 *{published data only}*

Leeuwenburgh MM, Wiezer MJ, Wiarda BM, Bouma WH, Phoa SS, Stockmann HB. Accuracy of MRI compared with ultrasound imaging and selective use of CT to discriminate simple from perforated appendicitis. *British Journal of Surgery* 2014;**101**:e147-55.

Lyons 2016 *{published data only}*

Lyons GR, Renjen P, Askin G, Giambrone AE, Beneck D, Kovanlikaya A. Diagnostic utility of intravenous contrast for MR imaging in pediatric appendicitis. *Pediatric Radiology* 2017;**47**(4):398-403.

Martin 2017 *{published data only}*

Martin J, Mathison D, Mullan P, Otero H. Secondary imaging for suspected appendicitis after equivocal ultrasound: time to disposition of MRI compared to CT. *Emergency Radiology* 2017;**25**:161-8.

Masselli 2011 *{published data only}*

Masselli G, Brunelli R, Casciani E, Polettini E, Bertini L, Laghi F. Acute abdominal and pelvic pain in pregnancy: MR imaging

as a valuable adjunct to ultrasound? *Abdominal Imaging* 2011;**36**:596-603.

Meesa 2011 *{published data only}*

Meesa I, Mammen L. MRI of pregnant women with abdominal pain and suspected appendicitis: diagnostic accuracy and outcomes. *American Journal of Roentgenology* 2011;**196 (5 Suppl)**:A176.

Moore 2012 *{published data only}*

Moore MM, Gustas CN, Choudhary AK, Methratta ST, Hulse MA, Geeting G. MRI for clinically suspected pediatric appendicitis: an implemented program. *Pediatric Radiology* 2012;**42**:1056-63.

Nitta 2005 *{published data only}*

Nitta N, Takahashi M, Furukawa A, Murata K, Mori M, Fukushima M. MR imaging of the normal appendix and acute appendicitis. *Journal of Magnetic Resonance Imaging* 2005;**21**:156-65.

Orth 2014 *{published data only}*

Orth RC, Guillerman RP, Zhang W, Masand P, Bisset GS 3rd. Prospective comparison of MR imaging and US for the diagnosis of pediatric appendicitis. *Radiology* 2014;**272**:233-40.

Oto 2005 *{published data only}*

Oto A, Ernst RD, Shah R, Koroglu M, Chaljub G, Gei AF. Right-lower-quadrant pain and suspected appendicitis in pregnant women: evaluation with MR imaging - initial experience. *Radiology* 2005;**234**:445-51.

Ozdemir 2018 *{published data only}*

Ozdemir O, Metin Y, Metin N, Bilir O, Yavasi O, Kupeli A. Added value of diffusion-weighted MR imaging to non-enhanced CT in the evaluation of acute appendicitis. *Iranian Journal of Radiology* 2018;**15**(1):1-9.

Patel 2017 *{published data only}*

Patel D, Fingard J, Winters S, Low G. Clinical use of MRI for the evaluation of acute appendicitis during pregnancy. *Abdominal Radiology* 2017;**42**(7):1857-63.

Pedrosa 2009 *{published data only}*

Pedrosa I, Lafornara M, Pandharipande PV, Goldsmith JD, Rofsky NM. Pregnant patients suspected of having acute appendicitis: effect of MR imaging on negative laparotomy rate and appendiceal perforation rate. *Radiology* 2009;**250**:749-57.

Petkovska 2016 *{published data only}*

Petkovska I, Martin DR, Covington MF, Urbina S, Duke E, Daye ZJ. Accuracy of unenhanced MR imaging in the detection of acute appendicitis: single-institution clinical performance review. *Radiology* 2016;**279**(2):451-60.

Ramalingam 2015 *{published data only}*

Ramalingam V, LeBedis C, Kelly JR, Uyeda J, Soto JA, Anderson SW. Evaluation of a sequential multi-modality imaging algorithm for the diagnosis of acute appendicitis in the pregnant female. *Emergency Radiology* 2015;**22**:125-32.

Rapp 2013 *{published data only}*

Rapp EJ, Naim F, Kadivar K, Davarpanah A, Cornfeld D. Integrating MR imaging into the clinical workup of pregnant patients suspected of having appendicitis is associated with a lower negative laparotomy rate: single-institution study. *Obstetrical and Gynecological Survey* 2013;**68**:617-8.

Repplinger 2018 *{published data only}*

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

Rosines 2014 *{published data only}*

Rosines LA, Chow DS, Lampl BS, Chen S, Gordon S, Mui LW. Value of gadolinium-enhanced MRI in detection of acute appendicitis in children and adolescents. *American Journal of Roentgenology* 2014;**203**(5):W543-8.

Shin 2017 *{published data only}*

Shin I, An C, Lim JS, Kim M-J, Chung YE. T1 bright appendix sign to exclude acute appendicitis in pregnant women. *European Radiology* 2017;**27**(8):3310-6.

Theilen 2015 *{published data only}*

Theilen LH, Mellnick VM, Longman RE, Tuuli MG, Odibo AO, MacOnes GA. Utility of magnetic resonance imaging for suspected appendicitis in pregnant women. *American Journal of Obstetrics and Gynecology* 2015;**212**:345.e1-6.

Thieme 2014 *{published data only}*

Thieme ME, Leeuwenburgh MM, Valdehueza ZD, Bouman DE, de Bruin IG, Schreurs WH. Diagnostic accuracy and patient acceptance of MRI in children with suspected appendicitis. *European Radiology* 2014;**24**:630-7.

Tsai 2017 *{published data only}*

Tsai R, Raptis C, Fowler KJ, Owen JW, Mellnick VM. MRI of suspected appendicitis during pregnancy: interradiologist agreement, indeterminate interpretation and the meaning of non-visualization of the appendix. *British Journal of Radiology* 2017;**90**(1079):20170383.

Vu 2009 *{published data only}*

Vu L, Ambrose D, Vos P, Tiwari P, Rosengarten M, Wiseman S. Evaluation of MRI for the diagnosis of appendicitis during pregnancy when ultrasound is Inconclusive. *Journal of Surgical Research* 2009;**156**:145-9.

Zhu 2012 *{published data only}*

Zhu B, Zhang B, Li M, Xi S, Yu D, Ding Y. An evaluation of a superfast MRI sequence in the diagnosis of suspected acute appendicitis. *Quantitative Imaging in Medicine and Surgery* 2012;**2**:280.

References to studies excluded from this review

Abdeen 2019 *{published data only}*

Abdeen N, Naz F, Linthorst R, Khan U, Dominguez PC, Koujok K, Bettoli M, Shenouda N. Clinical impact and cost-effectiveness

of noncontrast MRI in the evaluation of suspected appendiceal abscesses in children. *Journal of Magnetic Resonance Imaging* 2019;**49**(7):e241-9.

Ali 2011 *{published data only}*

Ali C, Shankar S. Acute care imaging in pregnancy. *Emergency Radiology* 2011;**18**(6):474.

Al-Katib 2016 *{published data only}*

Al-Katib S, Sokhandon F, Farah M. MRI for appendicitis in pregnancy: is seeing believing? clinical outcomes in cases of appendix nonvisualization. *Abdominal Radiology* 2016;**41**(12):2455-9.

Armstrong 2011 *{published data only}*

Armstrong L, Pandey T, Jambhekar K. MRI of the acute abdomen in pregnancy. In: Emergency Radiology. Vol. 18. 2011:6.

Aronberg 2017 *{published data only}*

Aronberg R, Richer E, Garver K. Efficacy and feasibility of MRI in the diagnosis of pediatric appendicitis: A pilot program in the community hospital setting. *Pediatric Radiology* 2017;**47**:S80.

Bernbeck 2015 *{published data only}*

Bernbeck M, Barth R, Lewis A, Dannenberg B, Imler D, Vasanawala S. Diffusion and post contrast MRI for evaluation of acute appendicitis: The Stanford experience. *Pediatric Radiology* 2015;**45**:1-246.

Birjawi 2009 *{published data only}*

Birjawi GA, Nassar LJ, Atweh LA, Akel S, Haddad MC. Emergency abdominal radiology: the acute abdomen. *The Lebanese Medical Journal* 2009;**57**(3):178-212. [PMID: 19938402]

Bracken 2018 *{published data only}*

Bracken R, Repplinger M, Markhardt B, Park J, Kim N, Kitchin D, Robbins J, Ziemlewicz T, Pickhardt P, Reeder S. Diagnostic accuracy of community versus academic radiologists' mr interpretations for the detection of acute appendicitis. *Emergency Radiology* 2018;**25**(5):567.

Brandon 2015 *{published data only}*

Brandon L, Hazel K, Kavanagh G, Ismail M, Keohane J, Sengupta S. A review of preoperative Imaging in appendectomy in a single referral centre in Ireland. *Gastroenterology* 2015;**148**(4):S-604.

Brian 2017 *{published data only}*

Brian J, Moore M. MRI for appendicitis in pediatric patients. *Applied Radiology* 2017;**46**(9):18-24.

Brook 2007 *{published data only}*

Brook O, Slotzky M, Deutsch M, Goldsher D. The role of magnetic resonance imaging in the differential diagnosis of acute right lower quadrant pain during pregnancy. *The Israel Medical Association Journal* 2007;**9**(12):883-4.

Byott 2016 *{published data only}*

Byott S, Harris I. Rapid acquisition axial and coronal T2 HASTE MR in the evaluation of acute abdominal pain. *European Journal of Radiology* 2016;**85**:286-90.

Cantineau 2009 *{published data only}*

Cantineau AE, Gordijn SJ, Hofker HS, de Groot JC, Zeeman GG. Appendicitis during pregnancy. *Dutch Journal of Medicine* 2009;**153**(1-2):20-4.

Church 2016 *{published data only}*

Church JT, Coughlin MA, Antunez AG, Smith EA, Bruch SW. Creating a diagnostic algorithm for complicated appendicitis. *Journal of the American College of Surgeons* 2016;**33**(9):1007-12.

Claudius 2015 *{published data only}*

Claudius I, Kearl YL, Behar S, Cooper J, Dollbaum R, Hardasmalani M. How definitive Is ultrasound for the diagnosis of appendicitis in children and is confirmatory advanced imaging necessary? *Annals of Emergency Medicine* 2015;**66**(4):S118.

Cobben 2009a *{published data only}*

Cobben LP, Bakker OJ, Puylaert JB, Kingma LM, Blickman JG. Imaging of patients with clinically suspected appendicitis in the Netherlands: conclusions of a survey. *British Journal of Radiology* 2009;**82**(978):482-5.

Corkum 2017 *{published data only}*

Corkum K, Oyetunji T, Grabowski J, Lautz T. An algorithm for the workup of suspected appendicitis in the era of MRI. *Journal of the American College of Surgeons* 2017;**225**(4):133.

Dabir 2010 *{published data only}*

Dabir D, Furst G, Blondin D. MRI in acute appendicitis in pregnancy [MRT bei akuter Appendizitis in der Schwangerschaft.]. *RöFo - Advances in X-ray and Imaging Technology* 2010;**182**(11):1010-1. [PMID: 20563964]

Díaz 2008 *{published data only}*

Cabrera SD, Díaz-Feijoo B, Gil-Moreno A, Martínez-Palones JM, Xercavins J. Mucocele of the appendix. *Journal of Minimally Invasive Gynaecology* 2008;**15**(2):130-1.

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. In: Pediatric Radiology. Vol. 46. 2016.

Dibble 2018 *{published data only}*

Dibble E, Swenson D, Cartagena C, Baird G, Herliczek T. Effectiveness of a staged US and unenhanced MR imaging algorithm in the diagnosis of pediatric appendicitis. *Radiology* 2018;**286**(3):1022-9.

Dillman 2015 *{published data only}*

Dillman J, Gadapelli S, Sroufe N, Smith E, Chong S, Mazza M. Implementation of a rapid noncontrast MRI protocol for use in children with equivocal appendicitis: A clinical effectiveness study. In: Pediatric Radiology. Vol. 45. 2015:S100-1.

Epifanio 2016 *{published data only}*

Epifanio M, Antonio de Medeiros Lima M, Corrêa P, Baldisserotto M. An imaging diagnostic protocol in children

with clinically suspected acute appendicitis. *American Surgeon* 2016;**82**(5):390-6.

Gottgens 2014 *{published data only}*

Göttgens K, Lahaye M, Lambregts D, Mutsaers E, Essers B, Bouvy N, Breukink S, Beets-Tan R. Does inclusion of imaging in the work up of patients with clinically suspected appendicitis reduce the rate of unnecessary surgical procedures? *Colorectal Disease* 2014;**16**:6.

Harringa 2016 *{published data only}*

Harringa J, Kinner S, Riedesel E, Gill K, Ziemlewicz T, Robbins J, Kitchin D, Pickhardt P, Reeder S, Repplinger M. 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):S147-8.

Harringa 2019 *{published data only}*

Harringa J, Bracken R, Davis J, Mao L, Kitchin D, Robbins J. Prospective evaluation of MRI compared with CT for the etiology of abdominal pain in emergency department patients with concern for appendicitis. *Journal of Magnetic Resonance Imaging* 2019;**50**(5):1651-8.

Herliczek 2012 *{published data only}*

Herliczek T, Swenson D. MRI of right lower quadrant pain in pediatric patients with inconclusive appendix us: Initial experience. *Pediatric Radiology* 2012;**42**:S244-5.

Hormann 2002 *{published data only}*

Hormann M, Puig S, Prokesch S, Partik B, Helbich T. MR imaging of the normal appendix in children. *European Radiology* 2002;**12**(9):2313-6.

How 2014 *{published data only}*

How R, Wikiel KJ, Hamad GG. Right lower quadrant pain during pregnancy. *JAMA Surgery* 2014;**149**(5):489-90.

Hutton 2015 *{published data only}*

Hutton MDC, Strugnell M, Hopkins I, Murray IA. Gastrointestinal: An abnormal MRI. Remember the history. *Journal of Gastroenterology and Hepatology* 2015;**30**(4):647.

Inci 2011a *{published data only}*

Inci E, Hocaoglu E, Aydin S, Palabiyik F, Cimilli T, Turhan AN, Aygün E. Efficiency of unenhanced MRI in the diagnosis of acute appendicitis: comparison with Alvarado scoring system and histopathological results. *European Journal of Radiology* 2011;**80**(2):253-8.

Inoue 2020 *{published data only}*

Inoue A, Furukawa A, Nitta N, Takaki K, Ohta S, Murata K. Optimization of pulse sequences in ultrafast magnetic resonance imaging for the diagnosis of acute abdominal pain caused by gastrointestinal disease. *Acta Radiologica Open* 2020;**9**(8):2058460120949246.

Kearl 2014 *{published data only}*

Kearl L, Claudius I, Behar S, Hardasmalani M, Rose E, Santillanes G. Magnetic resonance imaging for the diagnosis of pediatric appendicitis. *Annals of Emergency Medicine* 2014;**64**(4):S104-5.

Kelly 2017 *{published data only}*

Kelly D, Patel F, Gill M, Bhowmick A. Appendicectomy: The case for pre-operative imaging. *Colorectal Disease* 2017;**19**:60.

Koning 2014a *{published data only}*

Koning JL, Naheedy JH, Kruk PG. Diagnostic performance of contrast-enhanced MR for acute appendicitis and alternative causes of abdominal pain in children. *Pediatric Radiology* 2014;**44**(8):948-55.

Koning 2014b *{published data only}*

Koning J, Naheedy J, Kruk P, Hauschildt J. Contrast enhanced magnetic resonance evaluation of acute appendicitis in the pediatric population: efficacy of a novel imaging protocol. In: Pediatric Radiology. Vol. 44. 2014:948-55.

Kruger 2019 *{published data only}*

Krüger PC, Mentzel HJ. Radiological evaluation of acute abdomen in children. *Der Radiologe* 2019;**59**(2):146-153.

Leeuwenburgh 2013 *{published data only}*

Leeuwenburgh MM, Wiarda BM, Wiezer MJ, Vrouenraets BC, Gratama JW, Spilt A, Richir MC, Bossuyt PM, Stoker J, Boermeester MA, OPTIMAP Study Group. 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.

Lescheid 2015 *{published data only}*

Leschied J, Dillman J, Smith E, Strouse P, Gadapelli S, N Sroufe N. Suspected acute appendicitis in children: MRI appearances, alternative diagnoses, and lessons learned. In: Pediatric Radiology. Vol. 45. 2015:S186.

Long 2011 *{published data only}*

Long SS, Long C, Lai H, Macura KJ. Imaging strategies for right lower quadrant pain in pregnancy. *American Journal of Roentgenology* 2011;**196**(1):4-12.

Lyons 2017 *{published data only}*

Lyons GR, Renjen P, Askin G, Giambrone AE, Beneck D, Kovanlikaya A. Diagnostic utility of intravenous contrast for MR imaging in pediatric appendicitis. *Pediatric Radiology* 2017;**47**(4):398-403.

Martin 2016 *{published data only}*

Martin JF, Mathison DJ, Mullan PC, Otero HJ. Applicability of emergent MRI for suspected pediatric appendicitis with equivocal ultrasound results. In: Pediatric Radiology. Vol. 46. 2016:S127.

Meesa 2011a *{published data only}*

Meesa IR, Mammen L. MRI of pregnant women with abdominal pain and suspected appendicitis: Diagnostic accuracy and outcomes. In: American Journal of Roentgenology. Vol. 196 (5S). 2011:A176.

Mittal 2019 *{published data only}*

Mittal M. Appendicitis: role of MRI. *Paediatric Emergency Care* 2019;**35**(1):63-6.

Modgil 2006 *{published data only}*

Modgil G, Cooke DI, Newbury L. Appendiceal appearances: the great imitator. *Archives of Disease in Childhood* 2006;**91**(4):333.

Naz 2018 *{published data only}*

Naz F, Abdeen N. Rapid non-contrast MR protocol for evaluation of abdominal and pelvic abscesses in children. *Pediatric Radiology* 2018;**48**:S166-7.

Nitta 2005a *{published data only}*

Nitta N, Takahashi M, Furukawa A, Murata K, Mori M, Fukushima M. MR imaging of the normal appendix and acute appendicitis. *Journal of Magnetic Resonance Imaging* 2005;**21**(2):156-65.

Olympia 2016 *{published data only}*

Olympia RP, Hariharan N, Chen BY. Implementation of a MRI program for the evaluation of pediatric right lower quadrant abdominal pain: Effect on emergency department length of stay in adolescent females. *Annals of Emergency Medicine* 2016;**68**(4):S36.

Orth 2013 *{published data only}*

Orth RC, Guillerman RP, Zhang W, Masand P, Bisset GS. Prospective comparison of MRI and ultrasound for the diagnosis of pediatric appendicitis. *Pediatric Radiology* 2013;**43**:S298-9.

Oto 2009 *{published data only}*

Oto A, Ernst RD, Ghulmiyyah LM, Nishino TK, Hughes D, Chaljub G, Saade G. MR imaging in the triage of pregnant patients with acute abdominal and pelvic pain. *Abdominal Imaging* 2009;**34**(2):243-50.

Pedrosa 2009a *{published data only}*

Pedrosa I, Lafornara M, Pandharipande PV, Goldsmith JD, Rofsky NM. Pregnant patients suspected of having acute appendicitis: effect of MR imaging on negative laparotomy rate and appendiceal perforation rate. *Radiology* 2009;**250**(3):749-57.

Repplinger 2014 *{published data only}*

Repplinger MD, Pickhardt PJ, 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;**64**(4):S21.

Repplinger 2015 *{published data only}*

Repplinger M, Pickhardt P, Kitchin D, Robbins J, Ziemlewicz T, Reeder S. Prospective evaluation of contrast-enhanced magnetic resonance imaging for suspected appendicitis. *Academic Emergency Medicine* 2015;**22**:S402-3.

Rosenbaum 2017 *{published data only}*

Rosenbaum DG, Askin G, Beneck DM, Kovanlikaya A. Differentiating perforated from non-perforated appendicitis on contrast-enhanced magnetic resonance imaging. *Pediatric Radiology* 2017;**47**(11):1483-90.

Saunders 2016 *{published data only}*

Saunders D, Tolan D, Lambie H. Could MRI become the first-line investigation in suspected appendicitis? *American Journal of Roentgenology.* 2016;**207**(5):W103.

Shin 2018 *{published data only}*

Shin I, Chung YE, An C, Lee HS, Kim H, Lim JS, Kim MJ. Optimisation of the MR protocol in pregnant women with suspected acute appendicitis. *European Radiology* 2018;**28**(2):514-21.

Singh 2007 *{published data only}*

Singh A, Danrad R, Hahn PF, Blake MA, Mueller PR, Novelline RA. MR imaging of the acute abdomen and pelvis: acute appendicitis and beyond. *Radiographics* 2007;**27**(5):1419-31.

Spalluto 2012 *{published data only}*

* Spalluto LB, Grand DJ. MRI of acute appendicitis in the pregnant patient. *Medicine & Health Rhode Island* 2012;**95**(2):39-40.

Steinkeler 2008 *{published data only}*

Steinkeler JA, Woodfield CA. MR imaging of acute appendicitis in pregnancy. *Medicine & Health Rhode Island* 2008;**91**(6):196. [PMID: 18610808]

Steinkeler 2017 *{published data only}*

Steinkeler JA, Lee KS. MRI in pregnancy: gastrointestinal and genitourinary pathology. *Applied Radiology* 2017;**46**(5):23-8.

Stiefelhagen 2009 *{published data only}*

Stiefelhagen P. Difficult differential diagnosis. Appendicitis or right quadrant diverticulitis? MMW Fortschritte der Medizin 2009;**151**(46):18. [PMID: 20043386]

Stoker 2008 *{published data only}*

Stoker J. Magnetic resonance imaging and the acute abdomen. *British Journal of Surgery* 2008;**95**(10):1193-4.

Theilen 2014 *{published data only}*

Theilen L, Mellnick V, Longman R, Macones G, Cahill A. Rate of and risk factors for non-visualization of the appendix with magnetic resonance imaging (MRI) in pregnant women with suspected appendicitis. *American Journal of Obstetrics & Gynecology* 2014;**210**(1):S105-6.

Thompson 2015 *{published data only}*

Thompson G, Martin DA, Killam R, Eccles R, Brindle M, Gupta P. Magnetic resonance imaging provides useful diagnostic information following equivocal ultrasound in children with suspected appendicitis. *Academic Emergency Medicine* 2015;**22**:S325.

Trout 2015 *{published data only}*

Trout A, Wallihan D, Towbin A, Depinet H, Von Allmen D, Podberesky D. MR as a rapid secondary assessment for suspected appendicitis following equivocal sonography:

Feasibility and patient preference assessment. *Pediatric Radiology* 2015;**45**:S100.

Tseng 2018 *{published data only}*

Tseng J, Cohen T, Melo N, Alban RF. Imaging utilization affects negative appendectomy rates in appendicitis: An ACS-NSQIP study. *American Journal of Surgery* 2018;**217**(6):1094-8.

Warner 2020 *{published data only}*

Warner J, Desoky S, Tiwari HA, Morello F, Gilbertson D, Udayasankar U. Unenhanced MRI of the abdomen and pelvis in the comprehensive evaluation of acute atraumatic abdominal pain in children. *American Journal of Roentgenology* 2020;**215**(5):1218-28.

Wolfe 2007 *{published data only}*

Wolfe G, Oto A. MR evaluation of acute abdominal pain in pregnant patients. *Applied Radiology* 2007;**36**(9):20.

References to studies awaiting assessment

Covelli 2019 *{published data only}*

Covelli JD, Madireddi SP, May LA, Costello JE, Lisanti CJ, Carlson CL. MRI for pediatric appendicitis in an adult-focused general hospital: A clinical effectiveness study-challenges and lessons learned. American Journal of Roentgenology 2019;**212**(1):180-7.

Davis 2020 *{published data only}*

Davis J, Chima M, Kasmire K. Radiation-free diagnosis of pediatric appendicitis: accuracy of point-of-care ultrasonography and magnetic resonance Imaging. *Pediatric Emergency Care* 2020:Epub ahead of print. [PMID: 32925703]

Donlon 2019 *{published data only}*

Donlon NE, Kelly ME, Davern M, Sheppard A, Nugent T, Durand M. Should MRI be the imaging modality of choice in suspected appendicitis during pregnancy? *Irish Medical Journal* 2019;**112**(10):1018. [PMID: 32311244]

Heye 2020 *{published data only}*

Heye P, Saavedra JSM, Victoria T, Laje P. Accuracy of unenhanced, non-sedated MRI in the diagnosis of acute appendicitis in children. *Journal of Pediatric Surgery* 2020;**55**(2):253-6.

James 2020 *{published data only}*

James, Karl, Duffy, Patrick, Kavanagh, Richard G. Fast acquisition abdominal MRI study for the investigation of suspected acute appendicitis in paediatric patients. Insights into Imaging 2020;**11**(1):78.

Jung 2021 *{published data only}*

Jung JY, Na JU, Han SK, Choi PC, Lee JH, Shin DH. Differential diagnoses of magnetic resonance imaging for suspected acute appendicitis in pregnant patients. World Journal of Emergency Medicine 2018-03-15;**9**(1):26-32.

Kalimullina 2019 *{published data only}*

Kalimullina DS, Egorova EA, Lezhnev DA, Bazhin AV, Merkusheva MM. Magnetic resonance imaging for diagnosis of acute appendicitis in pregnant women in multi-field hospital. *Khirurgiia (Mosk)* 2019;(7):45-51.

Kashmire 2020 *{published data only}*

Kashmire K, Davis J, Chima M. Accurate, radiation-free evaluation of pediatric appendicitis. *Academic Emergency Medicine* 2020;**27**(S1):S7-S316.

Kennedy 2019 *{published data only}*

Kennedy TM, Thompson AD, Choudhary AK, Caplan RJ, Schenker KE, DePiero AD. Utility of applying white blood cell cutoffs to non-diagnostic MRI and ultrasound studies for suspected pediatric appendicitis. *American Joural of Emergency Medicine* 2019;**37**(9):1723-8.

Lukenaite 2020 *{published data only}*

Lukenaite B, Luksaite-Lukste R, Mikalauskas S, Strupas K, Poskus T. MRI reduces the rate of diagnostic laparoscopy in pregnant patients with suspected acute appendicitis. *Gastroenterology* 2020;**158**(6):S1535.

Marie 2019 *{published data only}*

Marie E, Alhashmi G, Ricci A, Man C, Doria A. How can criteria for interpretation of MRI examinations of appendicitis influence diagnostic accuracy? *Pediatric Radiology* 2019;**49**(1):1-245.

Mushtaq 2019 *{published data only}*

Mushtaq R, Desoky SM, Morello F, Gilbertson-Dahdal D, Gopalakrishnan G, Leetch A, Vedantham S, Kalb B, Martin DR, Udayasankar UK. First-line diagnostic evaluation with MRI of children suspected of having acute appendicitis. *Radiology* 2019;**291**(1):170-7.

Roh 2019 *{published data only}*

Roh, Albert T, Xiao, Zhibo, Cheng, Joseph Y. Conical ultrashort echo time (UTE) MRI in the evaluation of pediatric acute appendicitis. Abdominal Radiology 2019;**44**(1):22-30.

Sincavage 2019 *{published data only}*

Sincavage J, Buonpane C, Benyamen B, Benya E, Lautz T, Helenowski I. Alvarado scores predict additive value of magnetic resonance imaging in workup of suspected appendicitis in children. *Journal of Surgical Research* 2019;**244**:42-9.

Additional references

Akbar 2010

Akbar F, Yousuf M, Morgan RJ, Maw A. Changing management of suspected appendicitis in the laparoscopic era. Annals of The Royal College of Surgeons of England 2010;**92**(1):65-8. [PMID: 20056066]

Andersen 2005

Andersen BR, Kallehave FL, Andersen HK. Antibiotics versus placebo for prevention of postoperative infection after appendicectomy. *Cochrane Database of*

Systematic Reviews 2005, Issue 3. Art. No: CD001439. [DOI: 10.1002/14651858.CD001439.pub2]

Bakker 2010

Bakker OJ, Go PM, Puylaert JB, Kazemier G, Heij HA. Guideline on diagnosis and treatment of acute appendicitis: imaging prior to appendectomy is recommended. *Nederlands Tijdschrift Voor Geneeskunde* 2010;**154**:A303. [PMID: 21262032]

Barger 2010

Barger RL, Nandalur KR. Diagnostic performance of magnetic resonance imaging in the detection of appendicitis in adults: a meta-analysis. *Academic Radiology* 2010;**17**(10):1211-6. [PMID: 20634107]

Bhangu 2013

Bhangu A, Richardson C, Torrance A, Pinkney T, Battersby C, Beral D. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. *British Journal of Surgery* 2013;**100**(9):1240-52. [PMID: 23842836]

Bhangu 2020

Bhangu A, RIFT Study Group on behalf of the West Midlands Research Collaborative. Evaluation of appendicitis risk prediction models in adults with suspected appendicitis. *British Journal of Surgery* 2020;**107**(1):73-86.

Birchard 2005

Birchard KR, Brown MA, Hyslop WB, Firat Z, Semelka RC. MRI of acute abdominal and pelvic pain in pregnant patients. *American Journal of Roentgenology* 2005;**184**(2):452-8. [DOI: 10.2214/ajr.184.2.01840452]

Blumenfeld 2011

Blumenfeld YJ, Wong AE, Jafari A, Barth RA, El-Sayed YY. MR imaging in cases of antenatal suspected appendicitis--a meta-analysis. *Journal of maternal-fetal & neonatal medicine* 2011;**24**(3):485-8.

Bossuyt 2015

Bossuyt P, Reitsma J, Bruns D, Gatsonis C, Glasziou P, Irwig L. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal* 2015;**351**:826-32.

Boyd-Carson 2019

Boyd-Carson H, Doleman B, Herrod P, Anderson I, Williams J, Lund J. Association between surgeon special interest and mortality after emergency laparotomy. *British Journal of Surgery* 2019;**106**(7):940-8.

Brenner 2007

Brenner DJ, Hall EJ. Computed tomography - an increasing source of radiation exposure. *The New England Journal of Medicine* 2007;**357**(22):2277-84. [PMID: 18046031]

Cheng 2015

Cheng Y, Zhou S, Zhou R, Lu J, Wu S, Xiong X. Abdominal drainage to prevent intra-peritoneal abscess after open appendectomy for complicated appendicitis. *Cochrane Database of Systematic Reviews* 2015, Issue 2. Art. No: CD010168. [DOI: 10.1002/14651858.CD010168.pub2]

CODA 2020

CODA Collaborative. A randomized trial comparing antibiotics with appendectomy for appendicitis. *New England Journal of Medicine* 2020;**383**(20):1907-19.

D'Souza 2013

D'Souza N, Karim D, Sunthareswaran R. Bilirubin; a diagnostic marker for appendicitis. *International Journal of Surgery* 2013;**11**(10):1114-7. [PMID: 24080115]

D'Souza 2014

D'Souza N, Nugent K. Appendicitis. Systematic review 408. BMJ Clinical Evidence 2014:0408. [PMID: 25486014]

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. [PMID: 25499245]

D'Souza 2018

D'Souza N, Marsden M, Bottomley S, Nagarajah N, Scutt F, Toh S. Cost-effectiveness of routine imaging of suspected appendicitis. *Annals of the Royal College of Surgeons of England* 2018;**100**(1):47-51.

Dill 2008

Dill T. Contraindications to magnetic resonance imaging: non-invasive imaging. *Heart* 2008;**94**(7):943-8. [PMID: 18552230]

Di Saverio 2020

Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, Boermeester M et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World journal of emergency surgery* 2020;**15**:1-42.

Drake 2012

Drake FT, Florence MG, Johnson MG, Jurkovich GJ, Kwon S, Schmidt Z. Progress in the diagnosis of appendicitis: a report from Washington state's surgical care and outcomes assessment program. *Annals of Surgery* 2012;**256**(4):586-94. [PMID: PMID: 22964731]

Duke 2016

Duke E, Kalb B, Arif-Tiwari H, Daye ZJ, Gilbertson-Dahdal D, Keim SM. A systematic review and meta-analysis of diagnostic performance of MRI for evaluation of acute appendicitis. *American Journal of Roentgenology* 2016;**206**(3):508-17.

Florie 2006

Florie J, Wasser MN, Arts-Cieslik K, Akkerman EM, Siersema PD, Stoker J. Dynamic contrast-enhanced MRI of the bowel wall for assessment of disease activity in Crohn's disease. *American Journal of Roentgenology* 2006;**186**(5):1384-92. [DOI: 10.2214/AJR.04.1454]

Giordano 2013

Giordano S, Paakkonen M, Salminen P, Gronroos JM. Elevated serum bilirubin in assessing the likelihood of perforation in acute appendicitis: a diagnostic meta-analysis. *International Journal of Surgery* 2013;**11**(9):795-800. [PMID: 23732757]

Gronroos 1999

Gronroos JM, Gronroos P. Leucocyte count and C-reactive protein in the diagnosis of acute appendicitis. *The British Journal of Surgery* 1999;**86**(4):501-4. [PMID: 10215824]

Guller 2011

Guller U, Rosella L, McCall J, Brugger LE, Candinas D. Negative appendicectomy and perforation rates in patients undergoing laparoscopic surgery for suspected appendicitis. *The British Journal of Surgery* 2011;**98**(4):589-95. [PMID: 21259233]

Hall 2002

Hall MJ, Owings MF. National hospital discharge survey: advance data from vital and health statistics. National Center for Health Statistics 2002.

Harnoss 2017

Harnoss J, Zelienka I, Probst P, Grummich K, Maller-Lantzsch C. Antibiotics versus surgical therapy for uncomplicated appendicitis: systematic review and meta-analysis of controlled trials. *Annals of Surgery* 2017;**265**(5):889-900.

Health and Social Care Information Centre 2012

Health and Social Care Information Centre. Main operations summaries, hospital episode statistics, admitted patient care. Health and Social Care Information Centre 2012.

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 of England* 2012;**94**(7):476-80. [PMID: 23031764]

Jones 2012

Jones GE, Kreckler S, Shah A, Stechman MJ, Handa A. Increased use of laparoscopy in acute right iliac fossa pain, is it good for patients? *Colorectal Disease* 2012;**14**(2):237-42. [DOI: 10.1111/j.1463-1318.2011.02576.x]

Kave 2019

Kave M, Parooie F, Salarzaei M. Pregnancy and appendicitis: a systematic review and meta-analysis on the clinical use of MRI in diagnosis of appendicitis in pregnant women. *World Journal of Emergency Surgery* 2019;**14**(1):37.

Kim 2012

Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP. Low-dose abdominal CT for evaluating suspected appendicitis. *New England Journal of Medicine* 2012;**366**:1596-1605. [DOI: 10.1056/NEJMoa1110734]

Kim 2017

Kim HJ, Jeon BG, Hong CK, et al. 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 and Hepatology* 2017;**2**(11):793-804.

Lee 2020

Lee KH, Lee S, Park JH, et al. Risk of hematologic malignant neoplasms from abdominopelvic computed tomographic radiation in patients who underwent appendectomy.. *JAMA* 2021;**156**(4):343-51.

Leeuwenburgh 2012

Leeuwenburgh MM, Wiarda BM, Bipat S, Nio CY, Bollen TL, Kardux JJ. Acute appendicitis on abdominal MR images: training readers to improve diagnostic accuracy. *Radiology* 2012;**264**(2):455-63. [PMID: 22700556]

Leyendecker 2008

Leyendecker JR, Barnes CE, Zagoria RJ. MR Urography: techniques and clinical applications. *RadioGraphics* 2008;**28**(1):23-46. [DOI: 10.1148/rg.281075077]

Liu 2011

Liu K, Fogg L. Use of antibiotics alone for treatment of uncomplicated acute appendicitis: a systematic review and meta-analysis. *Surgery* 2011;**150**(4):673-83. [PMID: 22000179]

Ma 2010

Ma KW, Chia NH, Yeung HW, Cheung MT. If not appendicitis, then what else can it be? A retrospective review of 1492 appendectomies. *Hong Kong Medical Journal* 2010;**16**(1):12-7. [PMID: 20124568]

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Martin 2005

Martin DR, Danrad R, Herrmann K, Semelka RC, Hussain SM. Magnetic resonance imaging of the gastrointestinal tract. *Topics in Magnetic Resonance Imaging* 2005;**16**(1):77-98. [PMID: 16314698]

Moore 2016

Moore MM, Kulaylat AN, Hollenbeak CS, Engbrecht BW, Dillman JR, Methratta ST. Magnetic resonance imaging in pediatric appendicitis: a systematic review. *Pediatric Radiology* 2016;**46**(6):928-39.

Park 2017

Park H C, Kim M J, Lee B H. Randomized clinical trial of antibiotic therapy for uncomplicated appendicitis. *British Journal of Surgery* 2017;**104**(13):1785-90.

Pena 1999

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**(3 pt 1):440-6. [PMID: 10469767]

Phillips 2009

Phillips AW, Jones AE, Sargen K. Should the macroscopically normal appendix be removed during laparoscopy for acute right iliac fossa pain when no other explanatory pathology is found? *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques* 2009;**19**(5):392-4. [PMID: 19851267]

Rankey 2008

Rankey D, Leach JL, Leach SD. Emergency MRI utilization trends at a tertiary care academic medical center: baseline data. *Academic Radiology* 2008;**15**(4):438-43.

Rao 1998

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. *The New England Journal of Medicine* 1998;**338**(3):141-6. [PMID: 9428814]

Rehman 2011

Rehman H, Rao Ahsan M, Ahmed I. Single incision versus conventional multi-incision appendicectomy for suspected appendicitis. *Cochrane Database of Systematic Reviews* 2011, Issue 7. Art. No: CD009022. [DOI: 10.1002/14651858.CD009022.pub2]

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**(10):982-90. [DOI: 10.1016/j.jclinepi.2005.02.022]

Repplinger 2016

Repplinger MD, Levy JF, Peethumnongsin E, Gussick ME, Svenson JE, Golden SK. Systematic review and meta-analysis of the accuracy of MRI to diagnose appendicitis in the general population. *Journal of magnetic resonance imaging: JMRI* 2016;**43**(6):1346-54.

RevMan 2014 [Computer program]

Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Roberts 2008

Roberts JK, Behravesh M, Dmitrewski J. Macroscopic findings at appendicectomy are unreliable: implications for laparoscopy and malignant conditions of the appendix. *International Journal of Surgical Pathology* 2008;**16**(4):386-90. [PMID: 18387986]

Rothrock 1995

Rothrock SG, Green SM, Dobson M, Colucciello SA, Simmons CM. Misdiagnosis of appendicitis in nonpregnant women of childbearing age. *The Journal of Emergency Medicine* 1995;**13**(1):1-8. [PMID: 7782616]

Rud 2019

Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jorgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *The Cochrane Database of Systematic Reviews* 2019;(11). [PMID: 31743429]

Sallinen 2016

Sallinen V, Akl EA, You JJ, Agarwal A, Shoucair S, Vandvik PO. Meta-analysis of antibiotics versus appendicectomy for non-perforated acute appendicitis. *The British Journal of Surgery* 2016;**103**(6):656-67. [PMID: 26990957]

Salminen 2015

Salminen P, Paajanen H, Rautio T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: The APPAC randomized clinical trial. *JAMA* 2015;**313**(23):2340-8.

Sauerland 2010

Sauerland S, Jaschinski T, Neugebauer EAM. Laparoscopic versus open surgery for suspected appendicitis. *Cochrane Database of Systematic Reviews* 2010, Issue 10. Art. No: CD001546. [DOI: 10.1002/14651858.CD001546.pub3]

Seetahal 2011

Seetahal SA, Bolorunduro OB, Sookdeo TC, Oyetunji TA, Greene WR, Frederick W. Negative appendectomy: a 10-year review of a nationally representative sample. *American Journal of Surgery* 2011;**201**(4):433-7. [DOI: 10.1016/j.amjsurg.2010.10.009]

Sempere 2005

Sempere GJ, Martinez SV, Medina CE, Benages A, Tome TA, Canelles P. MRI evaluation of inflammatory activity in Crohn's disease. *American Journal of Roentgenology* 2005;**184**(6):1829-35. [DOI: 10.2214/ajr.184.6.01841829]

Sengupta 2009

Sengupta A, Bax G, Paterson-Brown S. White cell count and C-reactive protein measurement in patients with possible appendicitis. Annals of The Royal College of Surgeons of England 2009;**91**(2):113-5. [PMID: 19102827]

Sippola 2020

Sippola S, Virtanen J, Tammilehto V, Granroos J, Hurme S, Niiniviita H. The accuracy of low-dose computed tomography protocol in patients with suspected acute appendicitis: The OPTICAP Study. *Annals of surgery* 2020;**271**(2):332-8. [PMID: 30048324]

Slotboom 2014

Slotboom T, Hamminga JT, Hofker HS, Heineman E, Haveman JW, Apple Study Group. Intraoperative motive for performing a laparoscopic appendectomy on a postoperative histological proven normal appendix. *Scandinavian Journal of Surgery* 2014;**103**(4):245-8. [DOI: 10.1177/1457496913519771]

Sohaib 2007

Sohaib SAA, Reznek RH. MR imaging in ovarian cancer. *Cancer Imaging* 2007;**7**(Special issue A):S119-29. [DOI: 10.1102/1470-7330.2007.9046]

Stata 2015 [Computer program]

Stata Statistical Software. Version Release 14. College Station, TX: StataCorp LP, 2015.

Stoker 2010

Stoker Jaap. MRI of the Gastrointestinal Tract. Springer Science & Business Media, 2010.

Takwoingi 2013

Takwoingi Y. Meta-analysis of test accuracy studies in Stata: a bivariatemodel approach. Version 1.0. Available from: http://srdta.cochrane.org/ November 2013.

Teh 2000

Teh SH, O'Ceallaigh S, McKeon JGK, O'Donohoe MK, Tanner WA, Keane FBV. Should an appendix that looks normal be removed at diagnostic laparoscopy for acute right iliac fossa pain? *European Journal of Surgery* 2000;**166**(5):388-9. [PMID: 10881950]

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**(7):537-46. [PMID: 15466771]

The Royal College of Surgeons of England 2014

The Royal College of Surgeons of England. Emergency General Surgery - Commissioning Guide. 2014.

Tkacz 2009

Tkacz JN, Anderson SA, Soto J. MR imaging in gastrointestinal emergencies. *Radiographics* 2009;**29**(6):1767-80. [PMID: 19959520]

Tseng 2019

Tseng J, Cohen T, Melo N, Alban RF. Imaging utilization affects negative appendectomy rates in appendicitis: an ACS-NSQIP study. *American Journal of Surgery* 2019;**217**(6):1094-8.

U.S. Department of Health and Human Services 2015

Medical X-ray Imaging - what are the radiation risks from CT? U.S. Department of Health and Human Services 2015.

Van Rossem 2015

Van Rossem CC, Bolmers MDM, Schreinemacher MHF, Van Geloven AAW, Bemelman WA. Prospective nationwide outcome audit of surgery for suspected acute appendicitis. *British Journal of Surgery* 2015;**103**(1):144-51. [DOI: 10.1002/bjs.9964]

Vons 2011

Vons C, Barry C, Maitre S, Pautrat K, Leconte M, Costaglioli B. Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: an open-label, non-inferiority, randomised controlled trial. *Lancet* 2011;**377**(9777):1573-9. [PMID: 21550483]

Whiting 2011

Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Whiting 2013

Whiting Penny F, Rutjes Anne WS, Westwood Marie E, Mallett Susan, QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *Journal of Clinical Epidemiology* 2013;**66**(10):1093-104.

Wild 2013

Wild JRL, Abdul N, Ritchie JE, Rud B, Freels S, Nelson RL. Ultrasonography for diagnosis of acute appendicitis. *Cochrane Database of Systematic Reviews* 2013, Issue 2. Art. No: CD010402. [DOI: 10.1002/14651858.CD010402]

Wilms 2011

Wilms IMHA, De Hoog DENM, De Visser DC, Janzing HMJ. Appendectomy versus antibiotic treatment for acute appendicitis. *Cochrane Database of Systematic Reviews* 2011, Issue 11. Art. No: CD008359. [DOI: 10.1002/14651858.CD008359.pub2]

Yu 2013

Yu CW, Juan LI, Wu MH, Shen CJ, Wu JY, Lee CC. Systematic review and meta-analysis of the diagnostic accuracy of procalcitonin, C-reactive protein and white blood cell count for suspected acute appendicitis. *British Journal of Surgery* 2013;**100**(3):322-9. [DOI: 10.1002/bjs.9008]

Zanardi 2003

Zanardi R, Del Frate C, Zuiani C, Bazzocchi M. Staging of pelvic endometriosis based on MRI findings versus laparoscopic classification according to the American Fertility Society. *Abdominal Imaging* 2003;**28**(5):733-42. [DOI: 10.1007/s00261-003-0005-2]

References to other published versions of this review

D'Souza 2016

D'Souza N, Thaventhiran A, Beable R, Higginson A, Rud B. Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis. *Cochrane Database of Systematic Reviews* 2016;(1).

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

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

Aggarwala 2018

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 52 Females: 37 Mean age: unclear (range 10 to 39 years) Inclusion criteria: suspected appendicitis Exclusion criteria: unclear Setting: university hospital, Australia, 2017 to 2018
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 30 days Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Aggarwala 2018 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Unclear		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	Unclear		
Could the patient flow have introduced bias?		Unclear risk	

Aguilera 2018

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 52 Females: 52 Mean age: unclear (median 25) Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: unclear Setting: university hospital, USA, 2014 to 2016
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis

Aguilera 2018 *(Continued)*

	Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: “electronic medical record" for a period of between 1.4 to 3.6 years Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Aguilera 2018 *(Continued)*

Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Amitai 2016

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 49 Females: 49 Mean age: unclear Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: unclear Setting: university hospital, Israel, 2007 to 2013		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: “surgical and gynaecological follow up outcomes”. No duration given. Number of participants who were excluded from the analysis: unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Amitai 2016 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Aspelund 2014

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 397 Females: unclear Mean age: unclear Inclusion criteria: • Paediatric patients (< 18 years) with suspected appendicitis and equivocal ultrasound Exclusion criteria: • Over 18 years • Imaging performed to exclude other disease • Initial imaging elsewhere Setting: university hospital, USA, 2008 to 2012
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: telephone survey Number of participants who were excluded from the analysis: 237/397 participants not followed up by telephone; these participants were still included in analysis.
Comparative	
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		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

Aspelund 2014 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	No		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	No		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Avcu 2013

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 55 Females: 26 Mean age: 35.6 Inclusion criteria: consecutive patients with suspected appendicitis

Avcu 2013 *(Continued)*

	Exclusion criteria: patients diagnosed with appendicitis that refused surgery Setting: university hospital, Turkey, 2009 to 2010
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: unclear Number of participants who were excluded from the analysis: 5 (8.3%)
Comparative	
Notes	No details on clinical follow-up described in the methods, and limited to "observation period" in the results.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		

Avcu 2013 *(Continued)*

Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	Unclear	
Could the patient flow have introduced bias?		High risk

Batool 2016

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 100 Females: 56 Mean age: unclear Inclusion criteria: children with suspected appendicitis Exclusion criteria: unclear Setting: university hospital, Canada, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: unclear
Flow and timing	Type and length of follow-up: unclear Number of participants who were excluded from the analysis: unclear
Comparative	

Batool 2016 *(Continued)*

Notes	Conference abstract		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Unclear		
all -ve MRI had surgery or F/U	Unclear		

Batool 2016 *(Continued)*

choice of reference standard independent of MRI	Unclear
Could the patient flow have introduced bias?	High risk

Bayraktutan 2014

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 45 Females: 19 Mean age: 7 Inclusion criteria: • Paediatric patients with a clinical diagnosis of acute appendicitis • Paediatric patients with suspected appendicitis • Paediatric patients with an appendix that could not be visualised on ultrasound Exclusion criteria: • Patients with claustrophobia Setting: university hospital, Turkey, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up between 2 weeks and 2 months Number of participants who were excluded from the analysis: 2
Comparative	
Notes	No details on clinical follow-up described in the methods.

Methodological quality

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

Bayraktutan 2014 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Burke 2015

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 709

Burke 2015 *(Continued)*

	Females: 709 Mean age: 27.5 Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: absence of pathological confirmation Setting: multicentre, USA, 2009 to 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: unclear Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	No follow-up described.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Burke 2015 *(Continued)*

Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Burns 2018

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 63 Females: 63 Mean age: 31 Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: unclear Setting: university hospital, Canada, 2006 to 2012
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: medical records

Burns 2018 *(Continued)*

	Number of participants who were excluded from the analysis: unclear		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		

Burns 2018 *(Continued)*

all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Chabanova 2011

Study characteristics		
Patient Sampling	Type of study: prospective	
Patient characteristics and setting	Sample size: 48 Females: 29 Mean age: 37.1 Inclusion criteria: • Clinical diagnosis of appendicitis • Scheduled for appendicectomy Exclusion criteria: • Pregnancy • Age < 18 years • Contraindications for MRI Setting: university hospital, study dates unclear	
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5	
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or operative findings	
Flow and timing	Type and length of follow-up: histology or operative findings only (no follow-up, all participants underwent surgery) Number of participants who were excluded from the analysis: unclear	
Comparative		
Notes	There were applicability concerns regarding patient selection, as only patients undergoing surgery were included.	
Methodological quality		
Item	**Authors' judgement**	**Risk of bias** **Applicability concerns**

Chabanova 2011 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			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?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Cobben 2004

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 12 Females: 12 Mean age: 28 Inclusion criteria: pregnant women with suspected appendicitis undergoing ultrasonography Setting: district hospital, the Netherlands, 2000 to 2003
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up until delivery Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	No details on performance of clinical follow-up described in the methods.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Cobben 2004 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Cobben 2009

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 138 Females: 80 Mean age: unclear (range 6 to 80) Inclusion criteria: • Suspected appendicitis Exclusion criteria: • Claustrophobia • MRI scanner not working Setting: district hospital, the Netherlands, 2005 to 2006
Index tests	Index test: MRI

Cobben 2009 *(Continued)*

	Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up in outpatients within 1 week, follow-up period of 2 years Number of participants who were excluded from the analysis: 4
Comparative	
Notes	No description of how follow-up was performed beyond 1 week

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Cobben 2009 *(Continued)*

DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Corkum 2018

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 135 Females: unclear Mean age: 11.5 Inclusion criteria: children aged 5 to 18 with suspected appendicitis undergoing USS Exclusion criteria: no absolute neutrophil count available Setting: university hospital, USA, 2015 to 2016
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 30-day readmission data Number of participants who were excluded from the analysis: 16/128 (12.5%)
Comparative	
Notes	

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

Corkum 2018 *(Continued)*

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

des Plantes 2016

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 112 Females: 112 Mean age: 22 Inclusion criteria: • Female patients with suspected appendicitis Exclusion criteria: • No informed consent • Pregnancy • Contraindication for MRI Setting: university hospital, the Netherlands, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 4-month telephone follow-up Number of participants who were excluded from the analysis: 16/128 (12.5%)
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

des Plantes 2016 *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Dibble 2017

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 77 Females: unclear Mean age: 11.5 Inclusion criteria: paediatric patients (< 18 years) with suspected appendicitis Exclusion criteria: patients only undergoing MRI without USS Setting: university hospital, USA, 2011 to 2012

Dibble 2017 *(Continued)*

Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: medical records for 9 to 45 months Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Dibble 2017 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	Unclear	
Could the patient flow have introduced bias?		High risk

Didier 2017

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 98 Females: 60 Mean age: 11 Inclusion criteria: • Patients aged 4 to 18 with suspected appendicitis • Alvarado score >= 4 Exclusion criteria: • Pregnancy • Inability to tolerate MRI Setting: university hospital, USA, 2013 to 2015
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: medical records for 3 months Number of participants who were excluded from the analysis: unclear
Comparative	

Didier 2017 *(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?	Unclear		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		

Didier 2017 *(Continued)*

choice of reference standard independent of MRI	Unclear		
Could the patient flow have introduced bias?		High risk	

Dillman 2016

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 103 Females: 56 Mean age: 11.5 Inclusion criteria: paediatric patients (< 18 years) with suspected appendicitis and equivocal ultrasound Exclusion criteria: MRI scan not completed Setting: university hospital, USA, 2013 to 2014		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: case note review to exclude readmission within 30 days Number of participants who were excluded from the analysis: 3		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

Dillman 2016 *(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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	No		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	No		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Donlon 2015

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 29 Females: 29 Mean age: unclear Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: unclear

Donlon 2015 *(Continued)*

	Setting: Ireland, 2008 to 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review to exclude readmission Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	Conference abstract

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		

Donlon 2015 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Unclear	
>95% histo or F/U	Unclear	
all +ve MRI had surgery or F/U	Unclear	
all -ve MRI had surgery or F/U	Unclear	
choice of reference standard independent of MRI	Unclear	
Could the patient flow have introduced bias?		High risk

Fonseca 2014

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 31 Females: 31 Mean age: unclear Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: CT scan performed during same admission Setting: university hospital, USA, 2000 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: “surgical impression” and histology, or follow-up
Flow and timing	Type and length of follow-up: until delivery
Comparative	
Notes	
Methodological quality	

Fonseca 2014 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	No		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Herliczek 2013

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 60 Females: 32 Mean age: 13.4 Inclusion criteria: paediatric patients aged 7 to 17 undergoing appendix MRI examinations within 24 hours of an inconclusive ultrasound scan Setting: university hospital, USA, 2009 to 2012
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: chart review for an average of 14.5 months to exclude hospital readmission Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	Of the 48 participants who did not undergo appendicectomy, 29 were observed for an average of 1.4 days, and 19 were discharged from the emergency department without observation after MRI.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Herliczek 2013 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Heverhagen 2012

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 52 Females: 21 Mean age: 44.7 Inclusion criteria: • Patients with suspected appendicitis Exclusion criteria: • Underwent immediate surgery without preoperative MRI

Heverhagen 2012 *(Continued)*

	• Discharged from emergency department without the need for further evaluation Setting: university hospital, Germany, 2008
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: telephone follow-up at 1 month Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	Potential applicability concerns regarding patient selection, as patients with suspected appendicitis who were immediately discharged or underwent surgery did not undergo MRI.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		

Heverhagen 2012 *(Continued)*

Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	No		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Hormann 1998

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 20 Females: 14 Mean age: 12 Inclusion criteria: • Paediatric patients aged 7 to 16 years awaiting surgery for clinically diagnosed acute appendicitis • Preoperative ultrasound scan performed Setting: university hospital, Austria, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: histology only (no follow-up, all participants underwent surgery)

Hormann 1998 *(Continued)*

	Number of participants who were excluded from the analysis: none
Comparative	
Notes	Applicability concerns regarding patient selection, as only patients undergoing surgery were included. No patients with a negative MRI scan were included.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Hormann 1998 *(Continued)*

>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Hotchkiss 2011

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 44 Females: 454 Mean age: unclear Inclusion criteria: pregnant women with acute right lower quadrant pain Setting: USA, study dates unclear		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: 30 days Number of participants who were excluded from the analysis: unclear		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	

Hotchkiss 2011 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			High
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Unclear		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Unclear		
all -ve MRI had surgery or F/U	Unclear		
choice of reference standard independent of MRI	Unclear		
Could the patient flow have introduced bias?		High risk	

Imler 2017

Study characteristics	
Patient Sampling	Type of study: randomised
Patient characteristics and setting	Sample size: 37 Females: 26

Imler 2017 *(Continued)*

	Mean age: 13.5 Inclusion criteria: • Patients between 2 and 30 years old Exclusion criteria: • Surgical review prior to recruitment • Appendicitis not a differential diagnosis • Contraindication to MRI or ultrasound • Claustrophobia • Imaging protocol not followed Setting: university teaching hospital, USA, 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 7-day telephone follow-up Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Imler 2017 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Incesu 1997

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 60 Females: 33 Mean age: 20 Inclusion criteria: patients with suspected appendicitis Exclusion criteria: did not undergo both ultrasound and MRI Setting: university hospital, Turkey, 1994 to 1996
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis

Incesu 1997 *(Continued)*

	Reference standard: histology or laparoscopic findings, or follow-up
Flow and timing	Type and length of follow-up: follow-up until discharge from hospital Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Incesu 1997 *(Continued)*

Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
>95% histo or F/U	Yes
all +ve MRI had surgery or F/U	Yes
all -ve MRI had surgery or F/U	Yes
choice of reference standard independent of MRI	No
Could the patient flow have introduced bias?	High risk

Inci 2011

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 85 Females: 40 Mean age: 26.5 Inclusion criteria: • Adult patients with a clinical diagnosis of acute appendicitis Exclusion criteria: • Claustrophobia • Metal prostheses Setting: university hospital, Turkey, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up of minimum 1 month Number of participants who were excluded from the analysis: 4
Comparative	
Notes	No details on clinical follow-up described in the methods.
Methodological quality	

Inci 2011 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Israel 2008

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 33 Females: 33 Mean age: 25.6 Inclusion criteria: pregnant women with a clinical diagnosis of suspected appendicitis Exclusion criteria: did not undergo both ultrasound and MRI Setting: university hospital, USA, 2004 to 2006		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: chart review to exclude readmission at the hospital Number of participants who were excluded from the analysis: 3		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		

Israel 2008 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Jang 2011

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 18 Females: 18 Mean age: 31.7 Inclusion criteria: pregnant women with suspected appendicitis Setting: university hospital, Korea, 2008 to 2010
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5

Jang 2011 *(Continued)*

Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 1-month clinical follow-up Number of participants who were excluded from the analysis: none
Comparative	
Notes	No description of how clinical follow-up was performed or quantification of loss to follow-up

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Jang 2011 *(Continued)*

DOMAIN 4: Flow and Timing	
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
>95% histo or F/U	Yes
all +ve MRI had surgery or F/U	Yes
all -ve MRI had surgery or F/U	Yes
choice of reference standard independent of MRI	No
Could the patient flow have introduced bias?	High risk

Johnson 2012

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 42 Females: unclear Mean age: unclear Inclusion criteria: • Paediatric patients aged 4 to 17 years with suspected appendicitis Exclusion criteria: • Clinically unstable patients • History of acute trauma • Positive urine pregnancy test • Chronic medical conditions (e.g. inflammatory bowel disease) Setting: university teaching hospital, USA, study dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up for 6 months Number of participants who were excluded from the analysis: unclear
Comparative	

Johnson 2012 *(Continued)*

Notes	No description of how clinical follow-up was performed or quantification of loss to follow-up

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		

Johnson 2012 *(Continued)*

choice of reference standard independent of MRI	No
Could the patient flow have introduced bias?	High risk

Kearl 2016

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 192 Females: unclear Mean age: 14.8 Inclusion criteria: • Paediatric patients (aged 3 to 21 years) with suspected appendicitis and equivocal ultrasound Exclusion criteria: • Postsurgical evaluation • Evaluation for another condition • Incarceration • Self-discharge • Duplicate medical records Setting: teaching hospital, USA, 2010 to 2013
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: telephone follow-up or case note review Number of participants who were excluded from the analysis: 13 (6.3%) lost to follow-up
Comparative	
Notes	Telephone follow-up > 1 year

Methodological quality

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

Kearl 2016 *(Continued)*

Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Kennedy 2018

Study characteristics

Kennedy 2018 *(Continued)*

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 612 Females: 353 Mean age: 11.7 Inclusion criteria: • Patients <= 18 years with suspected appendicitis Exclusion criteria: • Missing MRI reports by attending radiologist • Missing pathology reports • Inconclusive final diagnosis • MRI study terminated due to movement Setting: university hospital, USA, 2014 to 2017
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: unclear Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Kennedy 2018 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	Unclear		
Could the patient flow have introduced bias?		High risk	

Khalil 2018

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 568 Females: unclear Mean age: unclear Inclusion criteria: • Patients under 18 years of age with suspected appendicitis Exclusion criteria: • Pregnant patients

Khalil 2018 *(Continued)*

	• 18 years of age or older Setting: university hospital, USA, 2014 to 2017
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology only
Flow and timing	Type and length of follow-up: unclear Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		

Khalil 2018 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		High
DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Unclear	
>95% histo or F/U	Unclear	
all +ve MRI had surgery or F/U	Unclear	
all -ve MRI had surgery or F/U	Unclear	
choice of reference standard independent of MRI	Unclear	
Could the patient flow have introduced bias?		High risk

Kinner 2017

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 230 Females: 28 Mean age: 17.1 Inclusion criteria: • Aged 12 to 20 years with suspected appendicitis undergoing CT scan Exclusion criteria: • Standard MRI contraindications • Inability to consent Setting: university hospital, USA, 2012 to 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: telephone call, 1 month

Kinner 2017 *(Continued)*

	Number of participants who were excluded from the analysis: unclear		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		

Kinner 2017 *(Continued)*

all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	Yes		
Could the patient flow have introduced bias?		High risk	

Koning 2014

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 364 Females: 223 Mean age: 11.3 Inclusion criteria: paediatric patients with suspected appendicitis Setting: paediatric hospital, USA, 2012 to 2013		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: until discharge from emergency department, case note review to exclude readmission Number of participants who were excluded from the analysis: none		
Comparative			
Notes	No quantification of loss to follow-up		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	

Koning 2014 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Konrad 2015

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 114 Females: 114

Konrad 2015 *(Continued)*

	Mean age: unclear Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: lack of clinical data Setting: university teaching hospital, USA, 2009 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review to exclude readmission, duration unclear Number of participants who were excluded from the analysis: unclear
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		

Konrad 2015 *(Continued)*

Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Kulaylat 2015

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 510 Females: 23 Mean age: 11.3 Inclusion criteria: paediatric patients (< 18 years) with suspected appendicitis Exclusion criteria: imaging at referring hospital prior to transfer Setting: university hospital, USA, 2011 to 2013
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review to exclude readmission Number of participants who were excluded from the analysis: none

Kulaylat 2015 *(Continued)*

Comparative
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
>95% histo or F/U	Unclear		
all +ve MRI had surgery or F/U	Yes		

Kulaylat 2015 *(Continued)*

all -ve MRI had surgery or F/U	Unclear
choice of reference standard independent of MRI	No
Could the patient flow have introduced bias?	High risk

Leeuwenburgh 2014

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 223 Females: 138 Mean age: 38 Inclusion criteria: • Adult patients with suspected appendicitis Exclusion criteria: • Claustrophobia • Technical failure Setting: multicentre, the Netherlands, 2010
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: clinical follow-up for 3 months Number of participants who were excluded from the analysis: none
Comparative	
Notes	No description of how 3-month clinical follow-up was performed. No quantification of loss to follow-up

Methodological quality

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

Leeuwenburgh 2014 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Lyons 2016

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 112

Lyons 2016 *(Continued)*

	Females: 65 Mean age: 12.7 Inclusion criteria: 21 years of age or younger with suspected appendicitis Exclusion criteria: pregnancy Setting: university hospital, USA, dates unclear
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: medical notes, unclear duration Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Lyons 2016 *(Continued)*

Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	Unclear		
Could the patient flow have introduced bias?		High risk	

Martin 2017

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 30 Females: 18 Mean age: 12.2 Inclusion criteria: • 5 to 18 years of age undergoing ultrasound for right lower quadrant pain or suspected appendicitis • Equivocal ultrasound Monday to Friday 5 am to 5 pm Exclusion criteria: • Pregnancy Setting: university hospital, USA, 2014 to 2015
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis

Martin 2017 *(Continued)*

	Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 7 days Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		

Martin 2017 *(Continued)*

Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Masselli 2011

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 40 Females: 40 Mean age: 28 Inclusion criteria: • Pregnant women with suspected appendicitis • Equivocal ultrasound Setting: university teaching hospital, Italy, 2006 to 2010		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: follow-up until delivery Number of participants who were excluded from the analysis: none		
Comparative			
Notes	Follow-up until delivery of baby: length unknown. No quantification of loss to follow-up		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Masselli 2011 *(Continued)*

Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Meesa 2011

Study characteristics

Meesa 2011 *(Continued)*

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 46 Females: 46 Mean age: 28 Inclusion criteria: pregnant women with suspected appendicitis Setting: multicentre, USA, 2008 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review to exclude readmission for surgery Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Meesa 2011 *(Continued)*

DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Moore 2012

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 208 Females: 119 Mean age: 11.2 Inclusion criteria: paediatric patients (3 to 17 years) with suspected appendicitis Setting: university hospital, USA, 2009 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review to exclude readmission for surgery

Moore 2012 *(Continued)*

	Number of participants who were excluded from the analysis: none
Comparative	
Notes	Follow-up of case notes to look for alternative diagnosis or the absence of readmission for up to 30 days

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Unclear		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		

Moore 2012 *(Continued)*

>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Nitta 2005

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 37 Females: 19 Mean age: 37.1 Inclusion criteria: adult patients with suspected appendicitis Setting: university hospital, Japan, study dates unclear		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: CT or ultrasound to confirm the negative diagnosis of acute appendicitis. Clinical follow-up to exclude readmission for surgery Number of participants who were excluded from the analysis: none		
Comparative			
Notes	No description of how clinical follow-up was performed		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Avoids Inappropriate Exclusion	Unclear		

Nitta 2005 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Orth 2014

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 81

Orth 2014 *(Continued)*

	Females: 47 Mean age: 12.3 Inclusion criteria: • Patients with suspected appendicitis undergoing ultrasound Exclusion criteria: • < 4 years of age • Acute distress • Contraindications to MRI • CT or US performed at a referral facility for review at the time of US interpretation • Declined consent Setting: academic paediatric hospital, 2012 to 2013
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review, telephone to general practitioner or parent/guardian in 88% (45/51). Duration unclear Number of participants who were excluded from the analysis: 6 (11.8%)
Comparative	
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		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		

Orth 2014 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Oto 2005

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 23 Females: 23 Mean age: 24.7 Inclusion criteria: pregnant women with suspected appendicitis Setting: university hospital, USA, 2001 to 2003
Index tests	Index test: MRI

Oto 2005 *(Continued)*

	Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Oto 2005 *(Continued)*

DOMAIN 4: Flow and Timing	
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
>95% histo or F/U	Yes
all +ve MRI had surgery or F/U	Yes
all -ve MRI had surgery or F/U	Yes
choice of reference standard independent of MRI	No
Could the patient flow have introduced bias?	High risk

Ozdemir 2018

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 94 Females: 46 Mean age: 38.1 Inclusion criteria: • Suspected appendicitis undergoing CT scan Exclusion criteria: • Excessive motion artefact • Appendix not visible • Other diagnosis • Pregnant • Under 16 years • Clinically unstable/poor co-operation/claustrophobia Setting: university hospital, Turkey, 2014 to 2017
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: 3-month follow-up Number of participants who were excluded from the analysis: unclear
Comparative	

Ozdemir 2018 *(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?	Yes		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		

Ozdemir 2018 *(Continued)*

choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Patel 2017

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 42 Females: 42 Mean age: 25.5 Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: patients with MRI performed at an outside hospital Setting: university teaching hospital, USA, 2008 to 2015		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: chart review up to 6 months postdischarge Number of participants who were excluded from the analysis: unclear		
Comparative			
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Patel 2017 *(Continued)*

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		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Pedrosa 2009

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 148 Females: 148 Mean age: 29 Inclusion criteria:

Pedrosa 2009 *(Continued)*

	• Pregnant women with suspected appendicitis who had undergone MRI Exclusion criteria: • Clinical charts not available • Cessation of participation in the study Setting: university hospital, USA, 2002 to 2007
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			

Pedrosa 2009 *(Continued)*

Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Petkovska 2016

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 403 Females: 23 Mean age: 41.5 Inclusion criteria: patients with suspected appendicitis Exclusion criteria: younger than 3 years or older than 50 years Setting: university hospital, USA, 2012 to 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: telephone follow-up, case note review, consensus panel assessment if no follow-up data

Petkovska 2016 *(Continued)*

	Number of participants who were excluded from the analysis: 35 (10.7%) with no follow-up data. Still included after consensus panel assessment
Comparative	
Notes	70.2% (229/326) participants were contacted for telephone follow-up at > 8 weeks. 19.0% (62/326) had a clinical follow-up note excluding appendicitis. 10.7% (35/326) could not be reached for telephone follow-up and had no clinical follow-up. A consensus panel reviewed each patient's notes, and all were assessed as negative for appendicitis.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Petkovska 2016 *(Continued)*

Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	No		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Ramalingam 2015

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 102 Females: 102 Mean age: 26.2 Inclusion criteria: pregnant women with suspected appendicitis with a negative or equivocal ultrasound Setting: university hospital, USA, 2007 to 2012		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: until discharge Number of participants who were excluded from the analysis: 1 (1%)		
Comparative			
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		

Ramalingam 2015 *(Continued)*

Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	No		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Rapp 2013

Study characteristics			

Rapp 2013 *(Continued)*

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 217 Females: 217 Mean age: 26 Inclusion criteria: retrospective analysis of patients who underwent laparotomy after MRI scan for suspected appendicitis Exclusion criteria: patients undergoing caesarean section Setting: university hospital, USA, 1996 to 2013
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Rapp 2013 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Repplinger 2018

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 198 Females: 114 Mean age: 31.6 Inclusion criteria: patients over 12 years of age undergoing CT for suspected appendicitis Setting: university hospital, USA, 2012 to 2014
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis

Repplinger 2018 *(Continued)*

	Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: case note review until discharge, telephone follow-up 1-month postdischarge Number of participants who were excluded from the analysis: 6 (lost to follow-up)
Comparative	
Notes	Potential selection bias, as 118 patients were enrolled out of 1224 eligible patients during the study period

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			

Repplinger 2018 *(Continued)*

Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
>95% histo or F/U	Yes
all +ve MRI had surgery or F/U	Yes
all -ve MRI had surgery or F/U	Yes
choice of reference standard independent of MRI	Yes
Could the patient flow have introduced bias?	High risk

Rosines 2014

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 49 Females: 24 Mean age: 12.9 Inclusion criteria: • Paediatric patients (aged 7 to 19 years) with suspected appendicitis and equivocal ultrasound scan Exclusion criteria: • Not scanned using all sequences • Motion degradation during scan Setting: university hospital, USA, 2010 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: until discharge Number of participants who were excluded from the analysis: 15 (23%)
Comparative	
Notes	
Methodological quality	

Rosines 2014 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Shin 2017

Study characteristics			
Patient Sampling	Type of study: retrospective		
Patient characteristics and setting	Sample size: 125 Females: 125 Mean age: 30.6 Inclusion criteria: • Pregnant women with suspected appendicitis Exclusion criteria: • MRI with protocol other than that used for appendicitis • Previous appendicectomy • Lost to follow-up Setting: university teaching hospital, Korea, 2008 to 2015		
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5		
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up		
Flow and timing	Type and length of follow-up: follow-up period of "at least 2 weeks" Number of participants who were excluded from the analysis: 8 (6%) lost to follow-up		
Comparative			
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index Test (All tests)			

Shin 2017 *(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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Theilen 2015

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 171 Females: 171 Mean age: unclear Inclusion criteria: • Pregnant women with suspected appendicitis

Theilen 2015 *(Continued)*

	Exclusion criteria: • Not pregnant • MRI for other disease processes (Crohn's flare, known psoas abscess, small bowel obstruction, diverticulitis, hydronephrosis) Setting: university hospital, USA, 2007 to 2012
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: until discharge Number of participants who were excluded from the analysis: none
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		

Theilen 2015 *(Continued)*

Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No	
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	No	
Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?	High risk	

Thieme 2014

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 104 Females: 57 Mean age: 12 Inclusion criteria: • Paediatric patients (4 to 18 years) with suspected appendicitis Exclusion criteria: • Recent abdominal surgery (< 6 weeks before inclusion) • Contraindications to MRI Setting: teaching hospital, the Netherlands, 2009
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up

Thieme 2014 *(Continued)*

Flow and timing	Type and length of follow-up: clinical follow-up of at least 3 months. Family physicians contacted if outside institution. Number of participants who were excluded from the analysis: 15 (23%)
Comparative	
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		
Avoids Inappropriate Exclusion	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		

Thieme 2014 *(Continued)*

Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	Yes	
choice of reference standard independent of MRI	No	
Could the patient flow have introduced bias?		High risk

Tsai 2017

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 233 Females: 233 Mean age: 28.4 Inclusion criteria: • Pregnant women with suspected appendicitis Exclusion criteria: • Appendiceal inflammation deemed secondary to non-appendiceal origin • Incorrect MRI protocol • Imaging deemed non-diagnostic by radiologists Setting: university hospital, USA, 2003 to 2015
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology only
Flow and timing	Type and length of follow-up: until discharge Number of participants who were excluded from the analysis: unclear
Comparative	
Notes	
Methodological quality	

Tsai 2017 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	No		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Vu 2009

Study characteristics	
Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 19 Females: 19 Mean age: 31 Inclusion criteria: consecutive pregnant patients with suspected appendicitis Setting: teaching hospital, Canada, 2004 to 2008
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or follow-up
Flow and timing	Type and length of follow-up: until delivery Number of participants who were excluded from the analysis: 1 (5%)
Comparative	
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		
Avoids Inappropriate Exclusion	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
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		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Vu 2009 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	No		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
>95% histo or F/U	Yes		
all +ve MRI had surgery or F/U	Yes		
all -ve MRI had surgery or F/U	Yes		
choice of reference standard independent of MRI	No		
Could the patient flow have introduced bias?		High risk	

Zhu 2012

Study characteristics	
Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 41 Females: 23 Mean age: 41.5 Inclusion criteria: patients with suspected appendicitis Setting: university hospital, China, 2009 to 2011
Index tests	Index test: MRI Index test criteria for positive diagnosis: see Appendix 5
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology

Zhu 2012 *(Continued)*

Flow and timing	Type and length of follow-up: histology only (no other follow-up as all patients underwent surgery) Number of participants who were excluded from the analysis: none
Comparative	
Notes	Applicability concerns regarding patient selection, as only patients undergoing surgery were included

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Avoids Inappropriate Exclusion	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			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		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference Standard			
Are the reference standards (histo or F/U) likely to correctly classify the target condition?	Yes		
Were the reference standard (histo or F/U) results interpreted without knowledge of the results of the index test	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		

Zhu 2012 *(Continued)*

Were all patients included in the analysis?	Yes	
>95% histo or F/U	Yes	
all +ve MRI had surgery or F/U	Yes	
all -ve MRI had surgery or F/U	No	
choice of reference standard independent of MRI	Yes	
Could the patient flow have introduced bias?		Low risk

CT: computed tomography
MRI: magnetic resonance imaging
US: ultrasound
USS: ultrasound scan

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

Study	Reason for exclusion
Abdeen 2019	Target condition not appendicitis
Ali 2011	Not diagnostic test accuracy study
Al-Katib 2016	Not diagnostic accuracy test study, character of cut-off and cut-off contingent on non-visualisation
Armstrong 2011	Not diagnostic test accuracy study
Aronberg 2017	Insufficient data for 2 x 2 table
Bernbeck 2015	Insufficient data for 2 x 2 table
Birjawi 2009	Not diagnostic test accuracy study
Bracken 2018	Duplicate data: Repplinger 2018 included
Brandon 2015	Not index test
Brian 2017	Not diagnostic test accuracy study
Brook 2007	Case report
Byott 2016	Insufficient data for 2 x 2 table (no response from authors to email seeking missing data)
Cantineau 2009	Case report
Church 2016	Retrospective examination of scans after appendicectomy for appendicitis
Claudius 2015	Duplicate data: Kearl 2016 included
Cobben 2009a	Not diagnostic test accuracy study
Corkum 2017	Duplicate data: Corkum 2018 included

Study	Reason for exclusion
Dabir 2010	Case report
Díaz 2008	Case report
Dibble 2016	Insufficient data for 2 x 2 table
Dibble 2018	Duplicate data: Dibble 2017 included
Dillman 2015	Duplicate data: Dillman 2016 included
Epifanio 2016	Insufficient data for 2 x 2 table
Gottgens 2014	Insufficient information for 2 x 2 table
Harringa 2016	Duplicate data of Kinner 2017
Harringa 2019	Target condition not appendicitis
Herliczek 2012	Duplicate data: Herliczek 2013 included
Hormann 2002	Not target condition
How 2014	Case report
Hutton 2015	Case report
Inci 2011a	Duplicate data: Inci 2011 included
Inoue 2020	Target condition not appendicitis
Kearl 2014	Duplicate data: Kearl 2016 included
Kelly 2017	Insufficient data for 2 x 2 table
Koning 2014a	Duplicate data: Koning 2014 included
Koning 2014b	Duplicate data: Koning 2014 included
Kruger 2019	Not diagnostic test accuracy study
Leeuwenburgh 2013	Duplicate data: Leeuwenburgh 2014 included
Lescheid 2015	Not diagnostic test accuracy study
Long 2011	Not diagnostic test accuracy study
Lyons 2017	Not diagnostic test accuracy study
Martin 2016	Insufficient data for 2 x 2 table
Meesa 2011a	Duplicate data: Meesa 2011 included
Mittal 2019	Not diagnostic test accuracy study
Modgil 2006	Case report

Study	Reason for exclusion
Naz 2018	Not diagnostic test accuracy study
Nitta 2005a	Not diagnostic test accuracy study
Olympia 2016	Not diagnostic test accuracy study
Orth 2013	Duplicate data: Orth 2014 included
Oto 2009	Not diagnostic test accuracy study
Pedrosa 2009a	Duplicate data: Repplinger 2018 included
Repplinger 2014	Duplicate data: Repplinger 2018 included
Repplinger 2015	Insufficient information for 2 x 2 table
Rosenbaum 2017	Not diagnostic test accuracy study
Saunders 2016	Letter
Shin 2018	Duplicate data: Shin 2017 included
Singh 2007	Not diagnostic test accuracy study
Spalluto 2012	Case report
Steinkeler 2008	Case report
Steinkeler 2017	Not diagnostic test accuracy study
Stiefelhagen 2009	Case report
Stoker 2008	Leading article
Theilen 2014	No reference standard
Thompson 2015	Insufficient information for 2 x 2 table
Trout 2015	No reference standard
Tseng 2018	Not diagnostic test accuracy study
Warner 2020	Not diagnostic test accuracy study
Wolfe 2007	Not diagnostic test accuracy study

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

Covelli 2019

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 528

Covelli 2019 *(Continued)*

	Female: 300 Mean age: 9.9 Inclusion criteria: age < 18, equivocal MRI, MRI report by non-paediatric radiologist Exclusion criteria: non-interpretable images (excessive motion), a known diagnosis of appendicitis in which surgery was not performed, studies read by paediatric radiologists Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology or operative findings
Flow and timing	Type and length of follow-up: outpatient follow-up > 1 month
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Davis 2020

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 209 Female: 104 Mean age: 10 years median age Inclusion criteria: ≤ 18 years Exclusion criteria: no saved images or no chart connected to images Setting: tertiary hospital, USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: notes checked for 90-day readmissions
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Donlon 2019

Patient Sampling	Type of study: retrospective

Donlon 2019 *(Continued)*

Patient characteristics and setting	Sample size: 29 Female: 29 Mean age: 29 years median Inclusion criteria: all consecutive pregnant patients with suspected appendicitis Exclusion criteria: unclear Setting: Ireland
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: unclear
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Heye 2020

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 350 Female: unclear Mean age: 12 years median Inclusion criteria: all unenhanced, non-sedated MRIs for suspected appendicitis Setting: paediatric hospital, USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: 30 days clinical follow-up
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

James 2020

Patient Sampling	Type of study: prospective
Patient characteristics and setting	Sample size: 52

James 2020 *(Continued)*

	Female: 34 Mean age: 11 Inclusion criteria: age 5 to 16 years (inclusive) with US for suspected appendicitis Exclusion criteria: previous abdominal surgery, behavioural disorder Setting: Ireland
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: patients who were treated non-surgically were followed for 6 months to assess for recurrent symptoms
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Jung 2021

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 46 Female: 46 Mean age: 31 years median Inclusion criteria: all pregnant patients with suspected appendicitis Exclusion criteria: unclear Setting: Korea
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: unclear
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Kalimullina 2019

Patient Sampling	Type of study: retrospective

Kalimullina 2019 *(Continued)*

Patient characteristics and setting	Sample size: 44 Female: 44 Mean age: 29.2 Inclusion criteria: pregnant women with suspected appendicitis Exclusion criteria: unclear Setting: Russia
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: unclear
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Kashmire 2020

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 208 Female: 103 Mean age: 10 years median Inclusion criteria: all paediatric patients who underwent US Exclusion criteria: unclear Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: unclear
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Kennedy 2019

Patient Sampling	Type of study: retrospective

Kennedy 2019 *(Continued)*

Patient characteristics and setting	Sample size: 612 Female: 353 Mean age: 11.7 Inclusion criteria: patients ≤ 18 years old with suspected appendicitis Exclusion criteria: missing MRI interpretations by attending radiologist, missing pathology reports, inconclusive final diagnoses, MRI studies terminated due to movement Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: follow-up encounter, unclear duration
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Lukenaite 2020

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 38 Female: 38 Mean age: 30.37 Inclusion criteria: all pregnant women admitted with suspected acute appendicitis Exclusion criteria: unclear Setting: Lithuania
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: unclear
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Marie 2019

Patient Sampling	Type of study: prospective

Marie 2019 *(Continued)*

Patient characteristics and setting	Sample size: 86 Female: unclear Mean age: 9.7 Inclusion criteria: children with suspected appendicitis undergoing US Exclusion criteria: unclear Setting: Canada
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: 1-month clinical follow-up
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Mushtaq 2019

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 403 Female: 235 Mean age: 13 years median Inclusion criteria: consecutive patients 18 years of age and younger Exclusion criteria: patients lost to follow-up Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: histology
Flow and timing	Type and length of follow-up: follow-up electronic records, unclear duration
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Roh 2019

Patient Sampling	Type of study: retrospective

Roh 2019 *(Continued)*

Patient characteristics and setting	Sample size: 84 Female: 47 Mean age: 11 Inclusion criteria: paediatric patients undergoing contrast-enhanced MRI for suspected appendicitis Exclusion criteria: incomplete exams that did not include all 3 sequences of interest were excluded Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: operative and histology reports
Flow and timing	Type and length of follow-up: 30-day chart review
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

Sincavage 2019

Patient Sampling	Type of study: retrospective
Patient characteristics and setting	Sample size: 112 Female: 60 Mean age: 12.4 years median Inclusion criteria: patients aged ≤ 18 years who received an indeterminate US for suspected appendicitis Exclusion criteria: unclear Setting: USA
Index tests	Index test: MRI
Target condition and reference standard(s)	Target condition: appendicitis Reference standard: unclear
Flow and timing	Type and length of follow-up: follow-up electronic records, unclear duration
Comparative	
Notes	We identified this study by a search update during the editorial process of the review.

MRI: magnetic resonance imaging
US: ultrasound

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 MRI	59	7482

ADDITIONAL TABLES

Table 1. Analyses of heterogeneity

Analyses	Number of studies (participants)	Sensitivity (95% CI)	Specificity (95% CI)	P value%
Overall	58 (7462)	0.95 (0.94 to 0.97)	0.96 (0.95 to 0.97)	-
MRI protocol - field of view	16 (1944)	0.95 (0.93 to 0.97)	0.96 (0.90 to 0.99)	LR Chi^2 = 2.37
• whole abdomen	13 (1326)	0.97 (0.94 to 0.99)	0.93 (0.87 to 0.96)	P = 0.31
• limited area				
Slice thickness	8 (1944)	0.97 (0.93 to 0.99)	0.95 (0.84 to 0.98)	LR Chi^2 = 1.58
• =< 4 mm	24 (2380)	0.96 (0.94 to 0.97)	0.96 (0.93 to 0.98)	P = 0.45
• > 4 mm				
MRI sequence	14 (1639)	0.96 (0.93 to 0.97)	0.97 (0.95 to 0.98)	LR Chi^2 = 1.16
• T2-weighted images only	36 (4788)	0.96 (0.94 to 0.97)	0.95 (0.93 to 0.97)	P = 0.56
• other				
Total scan time	5 (307)	0.94 (0.90 to 0.97)	0.93 (0.85 to 0.96)	LR Chi^2 = 0.42
• =< 10 min	26 (3144)	0.96 (0.94 to 0.97)	0.96 (0.94 to 0.98)	P = 0.81
• > 10 min				
Contrast enhancement	10 (1252)	0.97 (0.94 to 0.98)	0.95 (0.92 to 0.97)	LR Chi^2 = 1.64
• intravenous or oral contrast	41 (5767)	0.95 (0.92 to 0.97)	0.97 (0.94 to 0.98)	P = 0.44
• no contrast				

Abbreviations: MRI: magnetic resonance imaging, LR: likelihood ratio
All models fitted with equal variances for the random effects for logit(SN) and logit(SP) in the two groups.

Table 2. Sensitivity analyses

Sensitivity analysis	Number of studies (participants)	Summary estimates with 95% CI

Table 2. Sensitivity analyses *(Continued)*

		Sensitivity	Specificity
Exclusion of 8 studies with outlying results (sensitivity < 0.70 or specificity > 0.70)	50 (7090)	0.96 (0.95 to 0.97)	0.97 (0.95 to 0.97)
Low risk of bias for domain 1	14 (2096)	0.96 (0.94 to 0.97)	0.96 (0.90 to 0.99)
Low risk of bias for domain 2	26 (3272)	0.96 (0.95 to 0.97)	0.95 (0.92 to 0.97)
Low risk of bias for domain 3	6 (819)	0.96 (0.92 to 0.98)	0.94 (0.90 to 0.97)
Exclusion of 3 studies with low risk of bias for domains 1, 2, and 3	55 (6934)	0.95 (0.93 to 0.97)	0.96 (0.95 to 0.98)
Exclusion of 13 studies with fewer than 40 participants	45 (7111)	0.96 (0.94 to 0.97)	0.96 (0.95 to 0.97)
Retrospective study design	39 (5847)	0.95 (0.93 to 0.97)	0.96 (0.94 to 0.98)
US used before MRI in all participants	12 (705)	0.96 (0.90 to 0.98)	0.98 (0.94 to 0.99)
Overall	58 (7462)	0.95 (0.94 to 0.97)	0.96 (0.95 to 0.97)

Abbreviations: CI: confidence interval; MRI: magnetic resonance imaging; US: ultrasound

APPENDICES

Appendix 1. CENTRAL search strategy

CENTRAL in The Cochrane Library issue 1, 2021

1. “Appendicitis”

2. (Right near/2 (iliac fossa* or quadrant pain)):ti,ab,kw

3. “Appendix”

4. “Appendectomy”

5. (appendec* or appendicec* or appendicit*):ti,ab,kw

6. ((operat* or resect* or remov* or suger* or surgical or laparoscop* or acute) near/5 appendi*):ti,ab,kw

7. #1 or #2 or #3 or #4 or #5 or #6

8. “magnetic resonance” or “magnetic resonance imaging”

9. (MRI or MRIs):ti,ab,kw

10. (MR near/3 (imag* or scan*)):ti,ab,kw

11. #8 or #9 or #10

12. #7 and #11

Appendix 2. MEDLINE search strategy

Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present, 01 February 2021

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

8. Or/1-7

9. Magnetic resonance/ or magnetic resonance imaging/

10. (MRI or MRIs).tw,kf.

11. (MR adj3 (imag* or scan*)).tw,kf.

12. Or/9-11

13. 8 and 12

14. Exp animals/ not humans.sh.

15. 13 not 14

Appendix 3. Embase search strategy

Ovid Embase 1974 to 2021 Week 5

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. Nuclear magnetic resonance/ or nuclear magnetic resonance imagning/

9. (MRI or MRIs).tw,kw.

10. (MR adj3 (imag* or scan*)).tw,kw.

11. Or/8-10

12. 7 and 11

13. (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.)

14. 12 not 13

Appendix 4. QUADAS-2 guideline

QUADAS-2 assessment for MRI in appendicitis meta-analysis

Review question

What is the diagnostic accuracy of MRI for appendicitis?

Patients

All patients presenting to emergency department or the acute surgical team with suspected appendicitis (based on history and examination, or blood tests and urinalysis, or both).

Index test

MRI scan of the abdomen.

Reference standard

Appendicitis present: positive appendix histology.

Appendicitis not present: surgery resulting in negative appendix histology, or a normal appendix appearance intraoperatively with clinical follow-up. If no surgery, spontaneous resolution of symptoms with clinical follow-up.

Domain 1: Patient Selection

Risk of bias: Could the selection of patients have introduced bias?

Signalling question 1: Was a consecutive or random sample of patients enrolled?

- Yes: explicitly stated that enrolment was consecutive or random.
- No: above condition not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 2: Did the study describe explicit eligibility criteria for patients with suspected appendicitis?

- Yes: specific eligibility criteria on history, examination, observations, and baseline investigations described.
- No: no eligibility criteria described.
- Unclear: insufficient information available to answer yes or no.

Signalling question 3: Did the study avoid inappropriate exclusions?

- Yes if only the following patients were excluded.
  - Patients with very low clinical probability of appendicitis.
  - Peritonitic or septic patients too unwell for MRI.
  - Patients unable to undergo MRI due to:
    - unwillingness or inability to give consent (patient or guardian); or
    - inability to tolerate MRI (infants requiring intubation, claustrophobia).
- No: patients not meeting the above criteria were excluded.
- Unclear: insufficient information available to answer yes or no, or if consecutive or random sampling was stated but was inconsistent with other information in the study report

Risk of selection bias assessment

- High risk of bias: signalling questions 1, 2, and 3 are answered 'no'.
- Low risk of bias: signalling questions 1, 2, and 3 are answered 'yes'.
- Unclear risk of bias: insufficient information is reported to answer signalling questions 1, 2, and 3.

Applicability

Are there concerns that the included patients and setting do not match the review question?

- No concern
  - If patients are seen in the acute setting with a clinical history (migratory right iliac fossa pain, nausea, fevers, anorexia) and examination (rebound tenderness, tachycardia, low-grade pyrexia) consistent with appendicitis, with or without baseline investigations (blood tests and urinalysis) prior to MRI.
- High concern
  - Patients as above are not included, including stable patients with a high risk of appendicitis.
  - Patients from other settings (e.g. elective outpatient investigation) are included.
- Unclear
  - Insufficient information available.

Domain 2: Index test

Risk of bias: Could the conduct or interpretation of the index test have introduced bias?

Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?

MRI scans will routinely be performed prior to surgery. Reporting bias will only be present if the scan is reported after surgery, and the radiologist is aware of the operative findings.

- Yes if one of two conditions are met:
  - MRI scan reported prior to surgery; or
  - MRI scan reported:
    - following surgery with the radiologist blinded to the patient's operative findings; or
    - following conservative management with the radiologist blinded to the patient's clinical outcome.
- No
  - Neither condition met.
- Unclear
  - Insufficient information available to answer yes or no.

Signalling question 2: If a threshold was used, was it prespecified?

- Yes if pre-set criteria for MRI diagnosis of appendicitis are stated in the methodology.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Risk of index test bias assessment

- High risk of bias: signalling questions 1 or 2 are answered 'no'.
- Low risk of bias: signalling questions 1 or 2 are answered 'yes'.
- Unclear risk of bias: insufficient information is reported to answer signalling questions 1 or 2.

Applicability

Are there concerns that the index test or its conduct or interpretation differ from the review question?

The reproducibility of the index tests depends on several variables in its conduct and interpretation.

- Conduct
  - Sequences (e.g. T2 fast spin echo versus T1 gradient-recalled echo)
  - Region included in the scan (pelvis only versus abdomen and pelvis)
  - Slice thickness
  - Contrast (IV, oral, rectal)
  - Magnet strength
- Interpretation
  - Radiologist expertise and seniority

Domain 3: Reference standard

Risk of bias: Could the reference standard or its conduct or interpretation have introduced bias?

Signalling question 1: Is the reference standard likely to correctly classify the target condition?

- Yes if one the following conditions are met.
  - The diagnosis of appendicitis is based on histological analysis of the appendix specimen (all cases of macroscopic appendicitis at surgery should lead to appendicectomy).
  - A diagnosis excluding appendicitis not present is based on:
    - negative appendix histology;
    - a normal appearance to the appendix at surgery, with or without alternate pathology consistent with preoperative signs and symptoms.
      - This should be confirmed with treatment and symptom resolution or clinical follow-up for one month without recurrent symptoms or consequent appendicectomy.
  - Spontaneous resolution (i.e. without antibiotics) of symptoms and uneventful follow-up in patients that do not have surgery.
- No if none of the above conditions are met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

- Yes if the following conditions are met.

- If the appendix is removed, the histopathologist is blinded to results of the MRI scan.
- If the appendix is not removed, the clinician performing follow-up is blinded to the results of the MRI scan.

- No if neither of the above conditions met.
- Unclear: insufficient information available to answer yes or no.

Risk of reference test bias assessment

- High risk of bias: signalling questions 1 or 2 is answered 'no'.
- Low risk of bias: signalling questions 1 or 2 is answered 'yes'.
- Unclear risk of bias: insufficient information is reported to answer signalling questions 1 or 2.

Applicability

Are there concerns that the target condition as defined by the reference standard does not match the question?

- No concern
  - The study clearly aims to identify all cases of appendicitis.
- High concern
  - The review aims to identify subtypes of appendicitis or other conditions.
- Unclear
  - Insufficient information available.

Domain 4: Flow and timing

Risk of bias: Could the patient flow have introduced bias?

Signalling question 1: Did all patients receive a reference standard?

- Yes if the following conditions are met.
  - At least 95% of included patients had histological assessment (if appendicectomy performed) or clinical follow-up (if appendicectomy not performed).
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 2: Did all patients receive the same reference standard?

Patients are unlikely to have all received the same reference standard, as those with high risk of appendicitis would not undergo conservative management and clinical follow-up. Additionally, patients with low risk of appendicitis may not proceed to surgery. Nonetheless, different reference tests may introduce bias, since histological analysis is the reference standard.

- Yes if the following condition is met.
  - All patients had surgery with histological analysis of the appendix specimen.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 3: Did all patients with a positive MRI scan undergo surgery or clinical follow-up?

- Yes if the following condition is met.
  - All patients with a positive MRI scan underwent surgery or clinical follow-up.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 4: Did all patients with a negative MRI scan undergo surgery or clinical follow-up?

- Yes if the following condition is met.
  - All patients with a negative MRI scan underwent surgery or clinical follow-up.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 5: Was the choice of reference standard independent of the index test result?

- Yes if surgeons who decided on surgery or clinical follow-up were unaware of the MRI result.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Signalling question 6: Were all patients included in the analysis?

- Yes if the following condition is met.
  - At least 95% patients underwent surgery or clinical follow-up, or both.
- No if the above condition is not met.
- Unclear: insufficient information available to answer yes or no.

Risk of reference test bias assessment

- High risk of bias: signalling questions 1, 2, or 6 is answered 'no'.
- Low risk of bias: signalling questions 1, 2, or 6 is answered 'yes'.
- Unclear risk of bias: insufficient information is reported to answer signalling questions 1, 2, or 6.

Appendix 5. MRI criteria for appendicitis

Study	Appendix diameter	Wall thickening	Intraluminal fluid	Periappendiceal inflammation	Periappendiceal fluid	Appendicolith	Other findings
Aggarwala 2018							Not stated
Aguilera 2018							Not stated
Amitai 2016	Yes	Yes		Yes			High mural T2 signal secondary to mural oedema
Aspelund 2014							Not stated
Avcu 2013							High signal appendix lumen on DWI and low on ADC map
Batool 2016				Yes	Yes		Periappendiceal fat stranding
Bayraktutan 2014		Yes	Yes	Yes	Yes		Hyperintensity on DWI, hypointensity on ADC
Burke 2015	Yes			Yes	Yes	Yes	
Burns 2018	Yes	Yes		Yes			
Chabanova 2011	Yes	Yes		Yes	Yes		
Cobben 2004	Yes			Yes			Phlegmon or abscess
Cobben 2009	Yes			Yes			
Corkum 2018							Not stated
des Plantes 2016							Not stated
Dibble 2017							Not stated
Didier 2017	Yes	Yes	Yes		Yes		
Dillman 2016							Not stated

(Continued)

Donlon 2015							Not stated
Fonseca 2014							Not stated
Herliczek 2013	Yes	Yes	Yes	Yes		Yes	
Heverhagen 2012	Yes	Yes		Yes			
Hormann 1998	Yes		Yes	Yes	Yes		
Hotchkiss 2011							Not stated
Imler 2017							Not stated
Incesu 1997	Yes		Yes	Yes			Phlegmon
Inci 2011	Yes	Yes		Yes			
Israel 2008	Yes		Yes	Yes			
Jang 2011	Yes	Yes		Yes	Yes		Abscess
Johnson 2012	Yes	Yes		Yes	Yes		
Kearl 2016							Not stated
Kennedy 2018	Yes	Yes		Yes	Yes		
Khalil 2018							Not stated
Kinner 2017							Not stated
Koning 2014		Yes	Yes	Yes			Diffusion restriction
Konrad 2015	Yes				Yes		Abscess
Kulaylat 2015	Yes		Yes			Yes	
Leeuwenburgh 2014	Yes		Yes	Yes	Yes	Yes	Destruction of wall, abscess, extraluminal free air
Lyons 2016	Yes		Yes				

(Continued)

Masselli 2011							
Martin 2017	Yes			Yes	Yes		
Meesa 2011	Yes			Yes	Yes		
Moore 2012	Yes		Yes			Yes	
Nitta 2005	Yes	Yes		Yes			
Orth 2014	Yes	Yes		Yes		Yes	
Oto 2005	Yes	Yes			Yes		
Ozdemir 2018	Yes			Yes	Yes		Strong hyperintense signal on DWI, concomitant hypointensity on the ADC map
Patel 2017							Not stated
Pedrosa 2009	Yes		Yes	Yes			
Petkovska 2016		Yes	Yes	Yes			
Ramalingam 2015	Yes		Yes	Yes			
Rapp 2013	Yes	Yes	Yes	Yes	Yes		Abscess
Repplinger 2018	Yes	Yes	Yes	Yes		Yes	Increased appendix signal at DWI
Rosines 2014	Yes			Yes			
Shin 2017							T1 bright appendix sign
Theilen 2015							Not stated
Thieme 2014	Yes			Yes	Yes	Yes	Abscess, lymphadenopathy, restricted diffusion
Tsai 2017	Yes	Yes	Yes	Yes	Yes		

(Continued)

Vu 2009	Yes			Yes	
Zhu 2012	Yes	Yes	Yes	Yes	Abscess, free air

Abbreviations: ADC: apparent diffusion coefficient; DWI: diffusion-weighted imaging

WHAT'S NEW

Date	Event	Description
1 February 2021	New search has been performed	Searches updated.
13 March 2018	Amended	First draft review

HISTORY

Protocol first published: Issue 1, 2016

Date	Event	Description
20 November 2015	Feedback has been incorporated	Protocol revised according to editor's comments.
8 July 2015	Amended	Final version of protocol for editorial approval
30 June 2015	Amended	Final draft of protocol prepared for editorial approval.
31 January 2015	Amended	Started first draft of protocol

CONTRIBUTIONS OF AUTHORS

Nigel D'Souza conceived and co-ordinated the review and is the review guarantor.

Nigel D'Souza and Bo Rud designed the search strategies.

Nigel D'Souza, Anthony Thavenirathan, and Georgina Hicks extracted data.

Bo Rud performed the analyses.

Nigel D'Souza, Anthony Thavenirathan, and Bo Rud wrote the protocol.

Nigel D'Souza, Georgina Hicks, and Bo Rud wrote the review.

Richard Beable and Anthony Higginson reviewed and contributed to the protocol.

DECLARATIONS OF INTEREST

Nigel D'Souza: none
Georgina Hicks: none
Richard Beable: none
Antony Higginson: none
Bo Rud: none

SOURCES OF SUPPORT

Internal sources

- Sys Johnsen, Denmark

  Literature search expertise

External sources

- No sources of support provided

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In the protocol, we preplanned meta-regression analyses to investigate if summary sensitivity or summary specificity, or both, differed between children, adults, and pregnant women. In the review we decided to limit this investigation to subgroup analyses because we considered formal meta-regression analyses to be irrelevant, that is the summary estimates for the subgroups are relevant, not the potential difference between them.

In the review, we revised the definitions of categories for two covariates used in investigations of heterogeneity, as follows.

- For slice thickness, we used the categories ≤ 4 mm versus > 4 mm in the review; the categories ≤ 3 mm versus > 3 mm were preplanned in the protocol.
- For scan time, we used ≤ 10 min versus > 10 min in the review; the categories < 5 min, 5 to 20 min, > 20 min were preplanned in the protocol.

These revisions were necessary to avoid categories with few studies and to focus analyses.

In the review, we added sensitivity analyses to assess the influence of:

- small studies;
- use of MRI as a second-line imaging test following negative or equivocal ultrasound;
- outlying primary study results.

These analyses were not preplanned in the protocol. We also decided to assess the influence of methodological quality in sensitivity analyses rather than in meta-regression analyses as planned in the protocol.

INDEX TERMS

Medical Subject Headings (MeSH)

*Appendicitis [diagnostic imaging]; Magnetic Resonance Imaging; Retrospective Studies; Sensitivity and Specificity; Tomography, X-Ray Computed

MeSH check words

Adult; Child; Female; Humans; Pregnancy