Biomarkers for diagnosis of acute appendicitis in adults (Protocol)

Marshall MJ, Smart NJ, Hyde C, Winyard PG, Shaw AM, Daniels IR

This is a reprint of a Cochrane protocol, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2015, Issue 3

http://www.thecochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Biomarkers for diagnosis of acute appendicitis in adults

Morwena Jane Marshall[1], Neil J Smart[2], Chris Hyde[3], Paul Graham Winyard[4], Andrew M Shaw[5], Ian R Daniels[2]

[1]Exeter Health Services Research Unit, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK. [2]Colorectal Surgery, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK. [3]Institute of Health Research, University of Exeter Medical School, University of Exeter, Exeter, UK. [4]Experimental Medicine, University of Exeter Medical School, Exeter, UK. [5]Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

Contact address: Morwena Jane Marshall, Exeter Health Services Research Unit, Royal Devon & Exeter NHS Foundation Trust, Barrack Road, Exeter, Devon, EX2 5DW, UK. morwena.marshall@nhs.net.

Editorial group: Cochrane Colorectal Cancer Group.
Publication status and date: New, published in Issue 3, 2015.

Citation: Marshall MJ, Smart NJ, Hyde C, Winyard PG, Shaw AM, Daniels IR. Biomarkers for diagnosis of acute appendicitis in adults. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No.: CD011592. DOI: 10.1002/14651858.CD011592.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

The primary objective is to assess the diagnostic accuracy of biomarkers in acute appendicitis in adults.

Secondary objectives are to assess the diagnostic accuracy of biomarkers: in simple and complicated (gangrenous or perforated) acute appendicitis; in males and females; in women of reproductive age; in pregnancy; and at different time intervals from onset of pain.

BACKGROUND

Target condition being diagnosed

Acute appendicitis is the most common cause of emergency abdominal surgery (Humes 2006). The diagnosis remains a significant challenge due to varying symptoms and signs which may mimic other conditions. A delayed diagnosis can lead to significant complications such as perforation, abscess formation, peritonitis and death. The standard treatment is surgical removal of the inflamed appendix. Conservative management with antimicrobial therapy for early uncomplicated appendicitis is controversial due to increased failure rates (Varadhan 2010; Varadhan 2012). The accepted therapeutic role of antimicrobials is limited to the perioperative phase (Andersen 2005), in advanced disease such as appendiceal abscess or phlegmon (Andersson 2007b), and in remote medical environments (Campbell 2004). An incorrect diagnosis leads to removal of a normal appendix,known as a 'negative appendicectomy', in up to 20% of operations (National Surgical Research Collaborative 2013). Negative appendicectomies expose patients to the unnecessary risk of surgical complications with health economic implications (Flum 2002; Simpson 2008). Women of reproductive age are at higher risk of negative appendicectomy due to the increased prevalence of alternative diagnoses that mimic the presentation of appendicitis in this group, such as mid-cycle pain, ovarian cyst accidents, and pelvic inflammatory disease (Rothrock 1995).

Index test(s)

Blood biomarkers currently used for the diagnosis of acute appendicitis include C-reactive protein (CRP), white cell count (WCC)

and neutrophil count (absolute value or percentage of WCC) (Andersson 2004).

Emerging biomarkers under validation in clinical studies of acute appendicitis include Interleukin (IL)-1 to IL-10 (Paajanen 2002), bilirubin (Burcharth 2013), procalcitonin (Yu 2013) and calprotectin (Thuijls 2011).

Experimental diagnostic biomarkers have also been reported in the literature in small numbers and include pancreatic stone protein (Tschuor 2012), D-lactate (Duzgun 2006), D-dimer (Kaya 2012), fibrinogen (Kahramanca 2013), serum amyloid A (Schellekens 2013), mean platelet volume (Albayrak 2011), phospholipase A2 (Grönroos 1994), leucocyte elastase (Eriksson 1995), lactoferrin (Thuijls 2011), plasma total-oxidant capacity (Ozdogan 2006), adenosine deaminase (Öztürk 2008), lipopolysaccharide binding protein (Braˇ nescu 2012), and nuclear factor-kappaB (Pennington 2000).

Clinical pathway

Biomarkers form part of the pre-operative diagnostic work-up for patients with suspected acute appendicitis and feature in some clinical scoring systems for appendicitis. Current biomarkers attempt to identify the presence of inflammation and, together with a suggestive history and examination, may inform the decision to perform imaging or surgery in those with suspected appendicitis.

Alternative test(s)

Standard diagnostic practice involves history taking, clinical examination and often a period of active observation to elicit progression of symptoms (Andersson 2007a). Clinical scoring systems may stratify risk at this stage. However, these have been poorly validated and have not been widely adopted (Wilasrusmee 2014). Biochemical markers of inflammation, conventionally WCC and CRP, are employed to help inform the diagnosis.

Imaging modalities such as ultrasonography (US) (Van Randen 2008), computerised tomography (CT) and magnetic resonance imaging (MRI) may be utilised (Cobben 2009; Rud 2012). However these modalities have cost implications and varying availability. The diagnostic accuracy of US is operator-dependent and performances reported from high volume centres have not been reproduced in other healthcare settings (Toorenvliet 2010). A meta-analysis directly comparing graded compression US with CT in 671 predominantly adult patients gave an estimated sensitivity and specificity of 78% and 83%, respectively for US (Van Randen 2008). The poor sensitivity of US in adults as compared to children is in part due to increased adipose tissue with age (Doria 2006).

CT has the advantage of being less operator-dependent and interpretation is not impaired by bowel gas pattern, adipose tissue or patient discomfort. In adults, the sensitivity of CT is higher than US, with an estimated sensitivity and specificity of 91% and 90% in the aforementioned meta-analysis (Van Randen 2008). A disadvantage of CT is the associated ionising radiation exposure, which is especially undesirable in young patients, with concerns of increasing lifetime risk of cancer (Brenner 2007), and is relatively contraindicated in pregnancy.

The use of MRI in the diagnosis of appendicitis is a relatively recent advance and is not widely implemented. MRI protocols for appendicitis and training for interpretation of scans are under development (Cobben 2009). However, MRI is currently recommended by the American College of Radiology in suspected appendicitis in pregnancy when an US examination is negative or inconclusive and where ionising radiation should be avoided (Leeuwenburgh 2012).

The reference standard for appendicitis is histological examination of the excised appendix, giving a dichotomous marker with which to compare pre-operative diagnostic tests.

Rationale

Appendicitis remains the most common cause of abdominal surgical emergency and is a significant diagnostic challenge (Humes 2006). The reported negative appendicectomy rate is up to 20% and increases to 30% in females of reproductive age (National Surgical Research Collaborative 2013). Many studies have reported on the diagnostic accuracy of biomarkers in appendicitis and that of emerging novel biomarkers (Hallan 1997; Andersson 2004; Giordano 2013; Yu 2013). It is important to summarise objectively the conclusions of this rapidly expanding field for the practicing clinician, to keep information up-to-date, and to obtain performance criteria against which emerging biomarkers may be compared. *The Cochrane Library* contains a review of the diagnostic accuracy of CT in appendicitis (Rud 2012), but there is currently no protocol for the evaluation of biomarkers in the diagnosis of appendicitis. Biomarkers remain an important element in the diagnostic armamentarium of the clinician; particularly when imaging modalities are not available or radiation exposure may be undesirable, such as in young or pregnant patients.

OBJECTIVES

The primary objective is to assess the diagnostic accuracy of biomarkers in acute appendicitis in adults.

Secondary objectives

Secondary objectives are to assess the diagnostic accuracy of biomarkers: in simple and complicated (gangrenous or perforated) acute appendicitis; in males and females; in women of reproductive age; in pregnancy; and at different time intervals from onset of pain.

METHODS

Criteria for considering studies for this review

Types of studies

We will include prospective observational clinical studies measuring the accuracy of blood biomarkers relative to the reference standards for the diagnosis of appendicitis. We will exclude retrospective and diagnostic case-control studies, case reports and reviews. Studies must include a histological confirmation of acute appendicitis in operatively managed patients. The minimum dataset sufficient to construct a two-by-two contingency table of true positive, true negative, false positive and false negative rates, or data sufficient to derive these values, must be available for meta-analysis.

Participants

Participants will include adults with clinically suspected acute appendicitis. Clinically suspected acute appendicitis comprises those patients admitted to hospital with right iliac fossa pain in keeping with appendicitis, and patients undergoing appendicectomy for suspected appendicitis.

Adults are considered separately from children due to varying normal ranges of biomarkers between the two groups. The definition of adult varies in the literature from over 14 to over 18 years of age. Therefore studies that include patients younger than 14 will be excluded. It is important to evaluate the diagnostic test on the same population on which the test is applied in clinical practice as patient spectrum affects test performance characteristics (Bentley 2012).

We will exclude studies conducted in immunocompromised patients due to a theoretical alteration in biomarker response in this group. Populations from countries in which neglected tropical diseases (NTDs) are endemic will be excluded, due to the differing prevalence of pathology in these regions, particularly an increased presentation of parasitic infections and tuberculosis (Appendix 1). We will also exclude studies specifically conducted on patients with systemic conditions causing abdominal pain, such as porphyria, sickle cell anaemia and Familial Mediterranean Fever.

Index tests

The index tests will include all blood biomarkers used for the diagnosis of acute appendicitis. If combinations of biomarkers are reported in sufficient numbers, such as raised white cell count (WCC) and C-reactive protein (CRP); or raised WCC or CRP, then each defined combination will be considered separately as an index test. Similarly, if serial measurements of biomarkers are combined in a defined manner in sufficient numbers, each defined combination shall be considered as an index test. Where combinations or serial measurements of biomarkers are reported, we shall endeavour to extract individual data where available, for example on the performance of isolated index tests or admission tests.

Conventional blood biomarkers for the diagnosis of acute appendicitis include CRP, WCC and neutrophil count.

Experimental diagnostic biomarkers have also been reported in the literature and include IL-1 to IL-10 (Paajanen 2002), bilirubin (Burcharth 2013), procalcitonin (Yu 2013), calprotectin (Thuijls 2011), pancreatic stone protein (Tschuor 2012), D-lactate (Duzgun 2006), D-dimer (Kaya 2012), fibrinogen, (Kahramanca 2013), serum amyloid A (Schellekens 2013), mean platelet volume (Albayrak 2011), phospholipase A2 (Grönroos 1994), leucocyte elastase (Eriksson 1995), lactoferrin (Thuijls 2011), plasma total-oxidant capacity (Ozdogan 2006), adenosine deaminase (Öztürk 2008), lipopolysaccharide binding protein (Braˇ nescu 2012), and nuclear factor-kappaB (Pennington 2000).

The conventional biomarkers have predefined positivity thresholds derived from population normal ranges. The positivity thresholds from emerging biomarkers, however, are likely to be poorly defined and may therefore contribute to heterogeneity.

CRP is an acute-phase protein with a large dynamic range in response to non-specific inflammation. Sensitive immunoassays for CRP have been routinely employed in clinical practice since the mid-1990s and are widely reported in acute appendicitis. In health, the median concentration is 0.8 mg/L, the 90^{th} centile is 3.0 mg/L and the 99^{th} centile is 10 mg/L (Pepys 2003).

Target conditions

The target condition is acute appendicitis. This may be categorised into simple or complicated (gangrenous or perforated) based on histological examination.

Reference standards

The following reference standards will be used:

1. Histological examination of the excised appendix or intra-operative assessment of a non-excised appendix in operatively managed patients.
2. Computerised tomography (CT) diagnosis of non-operatively managed patients or magnetic resonance imaging (MRI) diagnosis in non-operatively managed pregnant patients.
3. Clinical discharge diagnosis in those who did not receive an operation or imaging. This may include clinical follow-up where provided, such as readmissions, telephone or clinic follow-up.

Histological evidence of inflammation of the appendix represents the 'gold standard' diagnosis for acute appendicitis. However, not all patients presenting with clinically suspected appendicitis will undergo appendicectomy. In order to include non-operatively managed patients (true negative and false negative) in the analysis, clinical and CT reference standards will also be used.

The clinical reference standard is not well defined in the literature. It may include discharge diagnosis, re-admission to hospital or telephone or clinic follow-up.

Search methods for identification of studies

Electronic searches

We will search *The Cochrane Library* (Appendix 2), MEDLINE (Ovid) (Appendix 3), EMBASE (Ovid) (Appendix 4) and the ISI Web of Knowledge (Appendix 5) using indexing terms and free text to capture the index tests and target disease as shown in the example outlined in Appendix 2. We will adapt the search terms to identify all papers included in existing systematic reviews of biomarkers in appendicitis. Relevant international surgical conference proceedings will be searched to identify abstracts. The search will be limited by date from 1990 to the current day in order to reduce heterogeneity in assays available for CRP. No language restrictions will be applied.

Searching other resources

The reference lists of retrieved articles will be searched manually for relevant studies.

Data collection and analysis

Selection of studies

Two clinical review authors (MJM and NJS) will independently assess the eligibility of the titles and abstracts of all papers identified by the search strategy, using the predefined inclusion and exclusion criteria (see Criteria for considering studies for this review). Those papers considered eligible will then be read in full and assessed for inclusion. Any discrepancies will be resolved by consensus opinion on review of the full article, if necessary by a third clinical review author (IRD).

Data extraction and management

Two review authors (MJM and NJS) will independently extract data using a standardised Excel spreadsheet. Data extracted will include study design, participant characteristics, outcome measures and methodological quality. Specific data extracted will include patient presentation, selection, age, gender and pregnancy status; prevalence of appendicitis, perforation and negative appendicectomy; reference standard used; and biomarker diagnostic performance characteristics. We will seek outstanding data from the publishing authors.

Assessment of methodological quality

We will use the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) criteria to assess the quality of selected studies (Whiting 2011). Assessment will be carried out by two authors and disagreements will be resolved by consensus between senior authors. In Appendix 3 we provide the precise criteria we expect to use to make judgements about the risk of bias and problems with applicability.

Statistical analysis and data synthesis

The statistical analysis will be undertaken by MJM, advised and supported by CH and his colleagues in the Statistics Group of the University of Exeter Medical School. The general approach will be that suggested in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010).

For all included studies, we will enter the data in the two-by-two tables into Review Manager 5 software (Review Manager 5), which will allow the sensitivities, specificities and their 95% confidence intervals (CIs) to be presented in forest plots and receiver operating characteristic (ROC) space. These presentations will be used to explore the included study results, considering each biomarker individually, focusing on the test threshold in common use in clinical practice. The interaction between test accuracy results, study characteristics and methodological study quality will also be considered. In this, the first round of this systematic review, we will be focusing on clarifying the accuracy of individual tests and combinations thereof where available. Given the scale of this task, we will not extend the review to consider comparisons of tests. A further justification for this is that there is currently no consensus about the most clinically relevant alternative diagnostic strategies which should sensibly be compared. We believe an initial review focused on individual tests and combinations will inform such a debate, and that comparing alternative biomarkers is premature.

As well as accuracy, we will tabulate the number of uninterpretable results reported in each included study for each biomarker.

Where appropriate, particularly in terms of the number of included studies (minimum of four studies), we will attempt meta-analysis of each pair of sensitivity and specificity results from each included study for each biomarker. We will use the Hierarchical Summary ROC (HSROC) model implemented in SAS (SAS) or WinBUGs (WinBUGS), depending on software availability in the host institution at the time we conduct the analyses. Although the bivariate method may be appropriate for some biomarkers where the threshold is well defined, it will not be appropriate for the newer biomarkers. We thus prefer an approach which can be used consistently across all tests. The HSROC will also allow exploration of heterogeneity by incorporating co-variates into the model provided there is a sufficient number of included studies (see below).

We will interpret the results and prepare a summary of results table using Chapter 11 of the *Cochrane Handbook for Systematic Reviews*

of Diagnostic Test Accuracy as guidance (Bossuyt 2013).

Investigations of heterogeneity

The wider framework for investigation of heterogeneity will include aspects of the index test (specific manufacturer, version of test and threshold (if no established threshold in clinical practice)), target disorder (proportion of simple or complicated appendicitis), target population (age and prior symptoms/investigation) and study quality (nature of reference standard).

Heterogeneity will be explored in the first instance through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC plots. We will then include possible explanators of the heterogeneity as co-variates in the HSROC model. Of the potential sources of heterogeneity listed, we think that the precise nature of the biomarker and the nature of the reference standard will be, a priori, the most important influences on heterogeneity, and so we will preferentially examine the influence of these where the number of included studies limits our ability to fully explore potential sources of heterogeneity.

Sensitivity analyses

Where appropriate and if sufficient data are available, we will explore the sensitivity of any summary accuracy estimates to aspects of study quality using the number of domains where risk of bias was identified in the QUADAS-2 quality assessment. Provisionally, we will consider the effect of excluding studies where the reference standard domain was judged as having a high or unclear risk of bias, as this is considered the most relevant source of bias. We will then consider the effect of excluding studies where any two or more domains (patient selection, index test, reference standard, flow and timing) were judged as having a high or unclear risk of bias.

Assessment of reporting bias

Quantitative methods for exploring reporting bias are not well established for studies of diagnostic test accuracy. Specifically, we will not consider funnel plots of the diagnostic odds ratio versus the standard error of this estimate.

ACKNOWLEDGEMENTS

The authors thank the Royal Devon and Exeter National Health Service (NHS) Foundation Trust, the University of Exeter, and the University of Exeter Medical School for their support.

REFERENCES

Additional references

Albayrak 2011
Albayrak Y, Albayrak A, Albayrak F, Yildirim R, Aylu B, Uyanik A. Mean platelet volume: a new predictor in confirming acute appendicitis diagnosis. *Clinical and Applied Thrombosis/Hemostasis* 2011;**17**:362–6.

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. [DOI: 10.1002/14651858.CD001439.pub2]

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

Andersson 2007a
Andersson RE. The natural history and traditional management of appendicitis revisited: spontaneous resolution and predominance of prehospital perforations imply that a correct diagnosis is more important than an early diagnosis. *World Journal of Surgery* 2007;**31**(1):86–92.

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

Bentley 2012
Bentley T, Catanzaro A, Ganiats T. Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis. *BMC Research Notes* 2012;**5**:563.

Bossuyt 2013
Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11: Interpreting results and drawing conclusions. In: Deeks JJ, Bossuyt PM, Gatsonis C editor (s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.9*. The Cochrane Collaboration, 2013. Available from: http://srdta.cochrane.org/.

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

Brănescu 2012
Brănescu C, Șerban D, Șavlovschi C, Dascălu A, Kraft A. Lipopolysaccharide binding protein (LBP)-an inflammatory marker of prognosis in the acute appendicitis. *Journal of Medicine and Life* 2012;**5**:342.

Burcharth 2013
Burcharth J, Pommergaard H, Rosenberg J, Gögenur I. Hyperbilirubinemia as a predictor for appendiceal perforation: a systematic review. *Scandinavian Journal of Surgery* 2013;**102**:55–60.

Campbell 2004
Campbell MR, Johnston SL, Marshburn T, Kane J, Lugg D. Nonoperative treatment of suspected appendicitis in remote medical care environments: implications for future spaceflight medical care. *Journal of the American College of Surgeons* 2004;**198**:822–30.

Centers for Disease Control and Prevention 2013
Centers for Disease Control and Prevention. Global Overlap of Six of the Common NTDs. http://www.cdc.gov/globalhealth/ntd/diseases/ntd-worldmap-table.html Last updated 22 November, 2013. Last accessed 20th Oct 2014.

Cobben 2009
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.

Doria 2006
Doria AS, Moineddin R, Kellenberger CJ, Epelman M, Beyene J, Schuh S, et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. *Radiology-Radiological Society of North America* 2006;**241**: 83–94.

Duzgun 2006
Duzgun AP, Bugdayci G, Sayin B, Ozmen MM, Ozer MV, Coskun F. Serum D-lactate: a useful diagnostic marker for acute appendicitis. *Hepato-gastroenterology* 2006;**54**: 1483–6.

Eriksson 1995
Eriksson S, Granström L, Olander B, Pira U. Leucocyte elastase as a marker in the diagnosis of acute appendicitis. *European Journal of Surgery = Acta Chirurgica* 1995;**161**: 901–5.

Flum 2002
Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. *Archives of Surgery* 2002;**137**:799.

Giordano 2013
Giordano S, Pääkkönen M, Salminen P, Grönroos 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.

Grönroos 1994
Grönroos J, Forsström J, Irjala K, Nevalainen TJ. Phospholipase A2, C-reactive protein, and white blood cell count in the diagnosis of acute appendicitis. *Clinical Chemistry* 1994;**40**:1757–60.

Hallan 1997
Hallan S, Åsberg A. The accuracy of C-reactive protein in diagnosing acute appendicitis-a meta-analysis. *Scandinavian Journal of Clinical & Laboratory Investigation* 1997;**57**: 373–80.

Humes 2006
Humes D, Simpson J. Acute appendicitis. *BMJ* 2006;**333**: 530.

Kahramanca 2013
Kahramanca S, Seker G, Azili C, Guzel H, Kucukpinar T, Kulacoglu IH. The role of fibrinogen, mean platelet volume and platelet distribution width in diagnosis of acute appendicitis. *European Surgical Research* 2013;**50**:139.

Kaya 2012
Kaya B, Sana B, Eris C, Karabulut K, Bat O, Kutanis R. The diagnostic value of D-dimer, procalcitonin and CRP in acute appendicitis. *International Journal of Medical Sciences* 2012;**9**:909.

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

Macaskill 2010
Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C editor(s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

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

Ozdogan 2006
Ozdogan M, Devay AO, Gurer A, Ersoy E, Devay SD, Kulacoglu H, et al. Plasma total anti-oxidant capacity correlates inversely with the extent of acute appendicitis: a case control study. *World Journal of Emergency Surgery* 2006; **1**:6.

Paajanen 2002
Paajanen H, Mansikka A, Laato M, Ristamäki R, Pulkki K, Kostiainen S. Novel serum inflammatory markers in acute appendicitis. *Scandinavian Journal of Clinical & Laboratory Investigation* 2002;**62**:579–84.

Pennington 2000
Pennington C, Dunn J, Li C, Ha T, Browder W. Nuclear factor kappaB activation in acute appendicitis: a molecular marker for extent of disease?. *American Surgeon* 2000;**66**: 914–8.

Pepys 2003
Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *Journal of Clinical Investigation* 2003;**111**:1805–12.

Review Manager 5
The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rothrock 1995
Rothrock SG, Green SM, Dobson M, Colucciello SA, Simmons CM. Misdiagnosis of appendicitis in nonpregnant women of childbearing age. *Journal of Emergency Medicine* 1995;**13**:1–8.

Rud 2012
Rud B, Olafsson L, Vejborg TS, Wilhelmsen M, Reitsma JB, Rappeport ED, et al. Diagnostic accuracy of computed tomography for appendicitis in adults. *Cochrane Database of Systematic Reviews* 2012, Issue 7. [DOI: 10.1002/14651858.CD009977]

SAS
World Headquarters SAS Institute Inc. 100 SAS Campus Drive, Cary, USA. SAS 9.4. World Headquarters SAS Institute Inc. 100 SAS Campus Drive, Cary, USA, 2013.

Schellekens 2013
Schellekens DH, Hulsewé KW, Acker BA, Bijnen AA, Jaegere TM, Sastrowijoto SH, et al. Evaluation of the diagnostic accuracy of plasma markers for early diagnosis in patients suspected for acute appendicitis. *Academic Emergency Medicine* 2013;**20**:703–10.

Simpson 2008
Simpson J, Samaraweera A, Sara R, Lobo D. Acute appendicitis-a benign disease?. *Annals of The Royal College of Surgeons of England* 2008;**90**:313.

Thuijls 2011
Thuijls G, Derikx JP, Prakken FJ, Huisman B, Van Bijnen Ing AA, Van Heurn EL, et al. A pilot study on potential new plasma markers for diagnosis of acute appendicitis. *American Journal of Emergency Medicine* 2011;**29**:256–60.

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

Tschuor 2012
Tschuor C, Raptis DA, Limani P, Bächler T, Oberkofler CE, Breitenstein S, et al. The value of pancreatic stone protein in predicting acute appendicitis in patients presenting at the emergency department with abdominal pain. *BMC Gastroenterology* 2012;**12**:154.

Van Randen 2008
Van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester M. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. *Radiology* 2008;**249**:97–106.

Varadhan 2010
Varadhan KK, Humes DJ, Neal KR, Lobo DN. Antibiotic therapy versus appendectomy for acute appendicitis: a meta-analysis. *World Journal of Surgery* 2010;**34**:199–209.

Varadhan 2012
Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. *BMJ* 2012;**344**:e2156.

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

Wilasrusmee 2014
Wilasrusmee C, Anothaisintawee T, Poprom N, McEvoy M, Attia J, Thakkinstian A. Diagnostic scores for appendicitis: a systematic review of scores' performance. *British Journal of Medicine & Medical Research* 2014;**4**:711–30.

WinBUGS
MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Cambridge, UK. WinBUGS 1.4. MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Cambridge, UK, 2003.

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**:322–9.

Öztürk 2008
Öztürk Z, Köklü S, Erol M, Yilmaz F, Baar Ö, Yüksel O, et al. Serum adenosine deaminase levels in diagnosis of acute appendicitis. *Emergency Medicine Journal* 2008;**25**:583–5.

* *Indicates the major publication for the study*

APPENDICES

Appendix 1. List of excluded populations from countries where neglected tropical diseases are endemic

Afghanistan, Algeria, American Samoa, Angola, Antigua and Barbuda, Argentina, Azerbaijan, The Bahamas, Bangladesh, Barbados, Belize, Benin, Bhutan, Bolivia, Botswana, Brazil, Brunei, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, China, Colombia, Comoros, Democratic Republic of the Congo, Republic of the Congo, Cook Islands, Costa Rica, Cote d'Ivoire, Cuba, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, Fiji, French Guiana, French Polynesia, Gabon, The Gambia, Ghana, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, India, Indonesia, Iran, Iraq, Jamaica, Kenya, Kiribati, North Korea, Kyrgyzstan, Laos, Lesotho, Liberia, Libya, Madagascar, Malawi, Malaysia, Maldives, Mali, Marshall Islands, Mauritania, Mauritius, Mexico, Federated States of Micronesia, Morocco, Mozambique, Myanmar (Burma), Namibia, Nauru, Nepal, New Caledonia, Nicaragua, Niger, Nigeria, Niue, Oman, Pakistan, Palau, Panama, Papua New Guinea, Peru, Philippines, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, Sao Tome and Principe, Saudi Arabia, Senegal, Seychelles, Sierra Leone, Solomon Islands, Somalia, South Africa, Sri Lanka, Sudan, South Sudan, Suriname, Swaziland, Syria, Tajikistan, Tanzania, Thailand, Timor-Leste, Togo, Tonga, Trinidad and Tobago, Turkey, Tuvalu, Uganda, Uruguay, Vanuatu, Venezuela, Vietnam, Wallis and Futuna, Yemen, Zambia, Zimbabwe (Centers for Disease Control and Prevention 2013).

Appendix 2. The Cochrane Library search strategy

#1 MeSH descriptor: [Appendicitis] explode all trees
#2 MeSH descriptor: [Appendectomy] explode all trees
#3 MeSH descriptor: [Appendix] explode all trees
#4 (appendectom* or appendice* or appendicit* or appendix):ti,ab,kw
#5 (#1 or #2 or #3 or #4)
#6 MeSH descriptor: [Biological Markers] explode all trees
#7 MeSH descriptor: [C-Reactive Protein] explode all trees
#8 MeSH descriptor: [Leukocyte Count] explode all trees
#9 MeSH descriptor: [Interleukins] explode all trees
#10 MeSH descriptor: [Bilirubin] explode all trees
#11 MeSH descriptor: [Leukocyte L1 Antigen Complex] explode all trees
#12 MeSH descriptor: [Lithostathine] explode all trees
#13 MeSH descriptor: [Fibrinogen] explode all trees
#14 MeSH descriptor: [Mean Platelet Volume] explode all trees
#15 MeSH descriptor: [Phospholipases A2] explode all trees
#16 MeSH descriptor: [Phospholipases A2] explode all trees
#17 MeSH descriptor: [Leukocyte Elastase] explode all trees
#18 MeSH descriptor: [Lactoferrin] explode all trees
#19 MeSH descriptor: [Adenosine Deaminase] explode all trees
#20 MeSH descriptor: [NF-kappa B] explode all trees 126
#21 (marker* or biomarker* or mutation* or C-reactive protein or CRP or white cell count or WCC or leucocyt count or neutrophil count or IL-1 or IL-2 or IL-3 or IL-4 or IL-5 or IL-6 or IL-7 or IL-8 or IL-9 or IL-10 or bilirubin or procalcitonin or calprotectin or lithostathine or D-lactate or D-dimer or fibrinogen or serum amyloid or mean platelet volume or phospholipase A2 or leucocyte elastase or lactoferrin or plasma total-oxidant capacity or adenosine deaminase or lipopolysaccharide binding protein or nuclear factor-kappaB):ti,ab,kw
#22 (#6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21)
#23 (#5 and #22)

Appendix 3. MEDLINE (Ovid) search strategy

1. Appendicitis/
2. Appendectomy/
3. Appendix/
4. (appendec* or appendic* or appendix).tw.
5. 1 or 2 or 3 or 4
6. Biological Markers/
7. exp C-Reactive Protein/
8. exp Leukocyte Count/
9. exp Interleukins/
10. exp Bilirubin/
11. exp Leukocyte L1 Antigen Complex/
12. exp Lithostathine/
13. exp Fibrinogen/
14. exp Mean Platelet Volume/
15. exp Phospholipases A2/
16. exp Leukocyte Elastase/
17. exp Lactoferrin/
18. exp Adenosine Deaminase/
19. exp NF-kappa B/
20. (marker* or biomarker* or mutation* or C-reactive protein or CRP or white cell count or WCC or leucocyt count or neutrophil count or IL-1 or IL-2 or IL-3 or IL-4 or IL-5 or IL-6 or IL-7 or IL-8 or IL-9 or IL-10 or bilirubin or procalcitonin or calprotectin or lithostathine or D-lactate or D-dimer or fibrinogen or serum amyloid or mean platelet volume or phospholipase A2 or leucocyte elastase or lactoferrin or plasma total-oxidant capacity or adenosine deaminase or lipopolysaccharide binding protein or nuclear factor-kappaB).tw.
21. 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22. 5 and 21
23. limit 22 to yr="1990 -Current"

Appendix 4. EMBASE (Ovid) search strategy

1. exp appendicitis/
2. exp acute appendicitis/
3. exp appendectomy/
4. exp appendix/
5. (appendectom* or appendice* or appendicit* or appendix).tw.
6. 1 or 2 or 3 or 4 or 5
7. *biological marker/
8. *C reactive protein/
9. *leukocyte count/
10. *cytokine/
11. *bilirubin/
12. *calgranulin/
13. *lithostathine/
14. *fibrinogen/
15. *thrombocyte volume/
16. *phospholipase A2/
17. *leukocyte elastase/
18. *Lactoferrin/
19. *adenosine deaminase/
20. (marker* or biomarker* or mutation* or C-reactive protein or CRP or white cell count or WCC or leucocyt count or neutrophil count or IL-1 or IL-2 or IL-3 or IL-4 or IL-5 or IL-6 or IL-7 or IL-8 or IL-9 or IL-10 or bilirubin or procalcitonin or calprotectin

or lithostathine or D-lactate or D-dimer or fibrinogen or serum amyloid or mean platelet volume or phospholipase A2 or leucocyte elastase or lactoferrin or plasma total-oxidant capacity or adenosine deaminase or lipopolysaccharide binding protein or nuclear factor-kappaB).tw.
21. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22. 6 and 21
23. limit 22 to yr="1990 -Current"

Appendix 5. Science Citation Index-Expanded/Conference Proceedings Citation Index-Science search strategy

#1 TOPIC: ((appendectom* or appendice* or appendicit* or appendix))
#2 TOPIC: ((marker* or biomarker* or mutation* or C-reactive protein or CRP or white cell count or WCC or leucocyt count or neutrophil count or IL-1 or IL-2 or IL-3 or IL-4 or IL-5 or IL-6 or IL-7 or IL-8 or IL-9 or IL-10 or bilirubin or procalcitonin or calprotectin or lithostathine or D-lactate or D-dimer or fibrinogen or serum amyloid or mean platelet volume or phospholipase A2 or leucocyte elastase or lactoferrin or plasma total-oxidant capacity or adenosine deaminase or lipopolysaccharide binding protein or nuclear factor-kappaB))
#3 (#1 AND #2)

Appendix 6. Review-specific tailoring of QUADAS2

Domain 1: patient selection

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

Signalling questions and answering guidelines

1) Was a consecutive or a random sample of patients enrolled?
Answer 'yes' if the study explicitly reports that a consecutive or random sample of patients was recruited. To fulfil this criterion, recruitment must have taken place at all hours and all days of the enrolment period.
Answer 'no' if the above statement was not fulfilled.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.
2) Was a case-control design avoided?
This question is irrelevant because case-control studies are excluded from the review. The default answer would always be 'yes' raising no concerns of bias.
3) Did the study avoid inappropriate exclusions?
Answer 'yes' if the exclusion criteria would not be expected to alter the estimated diagnostic performance of the index test. Appropriate exclusions include patients with anxiety, needle phobia, limited understanding or communication, and patients lacking capacity to consent.
Answer 'no' if the exclusion criteria would be expected to alter estimated diagnostic performance. For example excluding 'difficult to diagnose' patients, such as women of reproductive age, or the elderly, would overestimate diagnostic accuracy.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

Risk of bias from patient selection will be assessed as 'low' when the answer to signalling question 1 and 3 is 'yes'.
Risk will be assessed as 'high' when the answer to either signalling question is 'no'.
Risk will be assessed as 'unclear' when there is insufficient information available to answer 'yes' to signalling question 1 and 3.

Applicability: are there concerns that the included patients and setting do not match the review question?

Guidelines for assessing concern regarding applicability

Concerns regarding applicability relating to patient selection will be assessed as 'low' when the study population represents an unselected sample of adults with clinically suspected appendicitis. Appropriate exclusions are patients that do not match the review question, such as immunocompromised patients where it would be reasonable to expect biomarker performance to be altered to the extent that estimated performance characteristics from a general population would not be applied to these patients in clinical practice. Pregnancy is likely to be excluded in a number of studies due to concerns regarding the influence of altered physiology on biomarker normal ranges. This would be considered as an acceptable exclusion, as although we intend to perform subgroup analysis on this group and have therefore not excluded them from the review, the results of biomarker performance in the general population are not likely to be applied to pregnant women in clinical practice. If inappropriate exclusions account for five per cent or less of the included patients, the potential impact will be considered negligible.
Concern will be assessed as 'high' when the study population does not represent an unselected sample of adults with suspected appendicitis.
Concern will be assessed as 'unclear' when insufficient information is reported.

Domain 2: index test

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

Signalling questions and answering guidelines

1) Were the index test results interpreted without knowledge of the results of the reference standard?
In the majority of cases the index test is an objective measurement carried out before the reference standard and is therefore not subject to bias. However, some experimental biomarker assays maybe carried out post-hoc and may therefore be interpreted with knowledge of the reference standard.
Answer 'yes' if one of the following conditions is met:
a) the index test was carried out before surgery or patient discharge from hospital.
b) the biomarker assay is performed after the patient is discharged by a technician who is blinded to the reference standard, ie histology results or discharge diagnosis.
Answer 'no' if neither of the conditions are met.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.
2) If a threshold was used, was it pre-specified?
Answer 'yes' if a pre-specified biomarker positivity threshold is stated.
Answer 'no' if it is stated that a threshold was not pre-specified or that information in the article suggests it was not pre-specified (e.g. non-round numbers such as threshold of 12.3ng/ml).
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

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

Applicability: are there concerns that the index test, its conduct, or its interpretation differ from the review question?

Guidelines for assessing concern regarding applicability

Biomarker assays must be standardised across eligible studies for each index test. Assays for the conventional biomarkers such as CRP and WCC count are well described. Experimental biomarkers are likely to have greater variation in assay technique, and so the most widely performed assay technique for each index test will be identified from included papers and considered as the standard.
Concern regarding applicability relating to index test will be assessed as 'low' if one of the following conditions is met:
a) a standard assay is employed for conventional biomarkers.
b) the assays for experimental biomarkers are consistent with the most widely performed assay identified from included papers for each specific index test.
Concern will be assessed as 'high' if the biomarker assay does not fulfil either of the above conditions.
Concern will be assessed as 'unclear' if there is insufficient information available.

Domain 3: reference standard

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

Signalling questions and answering guidelines

1) Is the reference standard likely to correctly classify the target condition?
Answer 'yes' if the diagnosis of appendicitis (or its absence) is based on one of the following criteria:
a) histological assessment of the excised appendix or surgical assessment of the appendix if it is deemed non-inflamed and is therefore left in situ.
b) CT diagnosis of non-operatively managed patients who underwent CT imaging.
c) discharge diagnosis, readmission data, or clinical or telephone follow-up in non-operatively managed patients. The clinical reference standard must document resolution of symptoms at 7 to 31 days after the initial admission.
Answer 'no' if none of the above criteria are met.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.
2) Were the reference standard results interpreted without knowledge of the results of the index test?
Answer 'yes' if all of the following conditions are met:
a) the pathologist performing histological assessment of the excised appendix is blinded to the index test result.
b) the surgeon performing the intra-operative assessment of an appendix that is to remain in situ is blinded to the index test result.
c) the radiologist interpreting the CT scan of a non-operatively managed patient is blinded to the index test result.
d) the clinician or nurse assessing a patient on readmission or conducting clinical or telephone follow-up is blinded to the index test result.
Answer 'no' one of the above conditions is not met.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

Risk of bias relating to the reference standard will be assessed as 'low' when the answer to signalling questions 1 and 2 is 'yes'.
Risk will be assessed as 'high' when the answer to signalling questions 1 or 2 is 'no'.
Risk will be assessed as 'unclear' when there is insufficient information available to answer signalling questions 1 and 2.

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

Variations between the reference standard and the research question may arise from differing interpretations of histology or CT imaging. The accepted diagnosis of acute appendicitis on histology requires transmural inflammation with polymorphic granulocytes. The clinical importance of inflammation confined to the mucosa is dubious and would therefore not be applicable. The clinical reference standard is subject to the greatest degree of variations with the option for readmission data, clinic or telephone follow-up which may all be carried out at differing time points over a period of 7 to 31 days following initial presentation. The details of the clinical reference standard will be extracted for descriptive purposes.

Guidelines for assessing concern regarding applicability

Domain 4: flow and timing

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

Signalling questions and answering guidelines

1) Did all patients receive a reference standard?
Answer 'yes' if at least 95% of the included patients underwent a diagnosis by intra-operative, histological, or CT assessment or clinical follow-up (30 day readmission data or clinical or telephone follow-up).
Answer 'no' if less than 95% of the included patients underwent a diagnosis by intra-operative, histological, or CT assessment or clinical follow-up (30 day readmission data or clinical or telephone follow-up).
2) Did all patients receive the same reference standard?
We would not expect all patients to undergo surgery and it is therefore unlikely that all patients will have the same reference standard.
Answer 'yes' if one of the following conditions is met:
a) 90% of the included patients had surgery with histological examination of the excised appendix or intra-operative assessment of macroscopically normal non-excised appendix.
b) 90% of the included patients with no surgery underwent clinical follow-up (30 day readmission or clinic or telephone follow-up within the 7 to 31 day period following initial presentation).
Answer 'no' if neither of the above conditions is met.
Answer 'unclear' if there is insufficient information available to answer 'yes' or 'no'.
3) Were all patients included in the analysis?
Answer 'yes' if the analyses incorporated all patients. Also, answer 'yes' if 5% or less are excluded from the analysis because no reference standard assessment was available and/or if 5% or less were excluded due to missing or un-interpretable index test results.
Answer 'no' if the above statement is not met.
Answer 'unclear' if insufficient information is available to answer 'yes' or 'no'.

Guidelines for assessing risk of bias

The risk of bias relating to patient flow and timing will be assessed as 'low' when the answer to signalling question 1, 2 and 3 is 'yes'.
The risk will be assessed as 'high' when the answer to signalling question 1, 2 or 3 is 'no'.
The risk will be assessed as 'unclear' when there is insufficient information available to answer signalling question 1, 2 or 3.

CONTRIBUTIONS OF AUTHORS

Task	Contributing author
Draft the protocol	MJM, NJS, IRD, CH, PW
Develop a search strategy	MJM, NJS, IRD
Search for studies (usually 2 people)	MJM, NJS
Obtain copies of studies	MJM, NJS
Select which studies to include (2 + 1 arbiter)	MJM, NJS, IRD
Extract data from studies (2 people)	MJM, NJS
Enter data into RevMan	MJM
Carry out the analysis	MJM, AMS, CH
Interpret the analysis	MJM, NJS, IRD, AMS, CH, PW
Draft the final review	MJM, NJS, IRD, AMS, PW
Update the review	MJM, NJS, IRD, AMS, CH

DECLARATIONS OF INTEREST

The authors have no conflicts of interest to declare.

SOURCES OF SUPPORT

Internal sources

- Royal Devon and Exeter NHS Foundation Trust, UK.

Author salary, library and research facilities

- University of Exeter, UK.

Author salary, library and research facilities

- University of Exeter Medical School, UK.

Author salary and statistical support

External sources

- No sources of support supplied