Prospective, multicentre validation of appendicitis risk prediction models in children presenting with right iliac fossa pain

*RIFT Study Group on behalf of the West Midlands Research Collaborative**

*Collaborating authors are listed in S1 Appendix

Correspondence: Dmitri Nepogodiev, West Midlands Research Collaborative, Open Plan East, Second Floor, Institute for Translation Medicine, University of Birmingham, Edgbaston, Birmingham, B15 2TH, UK. Email: dnepogodiev@doctors.org.uk

Category: Original research

Abstract word count: 300

Main body word count: 3,999

Running title: Identifying children at low-risk of appendicitis

Keywords: Appendicitis; appendicectomy; paediatric surgery; emergency surgery; children; risk prediction models

Acknowledgements: We are grateful to the Birmingham Surgical Trials Consortium at the University of Birmingham for the use of their servers for secure online data collection.

Competing interests: The authors have no competing interests to declare.

Data availability: We will consider requests to make anonymised data available to interested researchers upon request and following a formal data use agreement. Data use requests should be forwarded to the West Midlands Research Collaborative Committee, who may be reached by email at. The mailing address for the committee is: West Midlands Research Collaborative Committee, Academic Department of Surgery, Room 29, Fourth Floor, Heritage Building, University of Birmingham, Mindelsohn Drive, Edgbaston, Birmingham, B15 2TT (phone number +44-121-371-8910).

Research in context

Evidence before this study: Acute appendicitis is the most common general surgical emergency in children. Its diagnosis remains challenging and children presenting with acute right iliac fossa (RIF) pain may be admitted for clinical observation or undergo normal appendicectomy (removal of a histologically normal appendix). A search for external validation studies of risk prediction models for acute appendicitis in children was performed on MEDLINE and Web of Science on 12 January 2017 using the search terms [“appendicitis” OR “appendectomy” OR “appendicectomy”] AND [“score” OR “model” OR “nomogram” OR “scoring”]. Studies validating prediction models aimed at differentiating acute appendicitis from all other causes of RIF pain were included. No date restrictions were applied. Validation studies were most commonly performed for the Alvarado, Appendicitis Inflammatory Response Score (AIRS), and Paediatric Appendicitis Score (PAS) models. Most validation studies were based on retrospective, single centre, or small cohorts, and findings regarding model performance were inconsistent. There was no high quality evidence to guide selection of the optimum model and threshold cut-off for identification of low-risk children in the UK and Ireland.

Added value of this study: Most children admitted to hospital with RIF pain do not undergo surgery. When children do undergo appendicectomy, removal of a normal appendix (normal appendicectomy) is common, occurring in around 1 in 6 children. The Shera score is able to identify a large low-risk group of children who present with acute RIF pain but do not have acute appendicitis (specificity 44%). This low-risk group has an overall 1 in 30 risk of acute appendicitis and a 1 in 270 risk of perforated appendicitis. The Shera score is unable to achieve a sufficiently high positive predictive value to select a high-risk group who should proceed directly to surgery. Current diagnostic performance of ultrasound is also too poor to select children for surgery.

Implications of all the available evidence: Routine pre-operative risk scoring could inform shared decision making by doctors, children, and parents by supporting safe selection of low-risk patients for ambulatory management, reducing unnecessary admissions and normal appendicectomy. Hospitals should ensure seven-day-a-week availability of ultrasound for medium and high-risk patients. Ultrasound should be performed by operators trained to assess for acute appendicitis in children. For children in whom diagnostic uncertainty remains following ultrasound, magnetic resonance imaging (MRI) or low-dose computed tomography (CT) are second-line investigations.

Abstract

Background: Acute appendicitis is the most common surgical emergency in children. Differentiating acute appendicitis from conditions that do not require operative management can be challenging in children. This study aimed to identify the optimum risk prediction model for stratifying acute appendicitis risk in children.

Methods: A rapid review was performed to identify acute appendicitis risk prediction models. A prospective, multicentre cohort study was undertaken to evaluate performance of these models. Children (age 5-15 years) presenting with acute right iliac fossa (RIF) pain in the UK and Ireland were included. For each model, score cut-off thresholds were systematically varied to identify the best achievable specificity whilst maintaining a failure rate (proportion of patients identified as low-risk who had acute appendicitis) under 5%. The normal appendicectomy rate (NAR) was the proportion of resected appendixes found to be normal on histopathological examination.

Findings: Fifteen risk prediction models were identified that could be assessed. The cohort study enrolled 1827 children from 139 centres, of whom 34.5% (630/1827) underwent appendicectomy. NAR was 15.9% (100/630). The Shera score (area under the curve 0.84) was the best performing model. Applying score cut-offs ≤3 for children aged 5-10 and females aged 11-15, and ≤2 for males aged 11-15, the failure rate was 3.3% (18/539), specificity 44.3% (521/1176), and positive predictive value (PPV) was 41.4% (463/1118). PPV for the Shera score with a cut-off of ≤6 was similar to that of ultrasound scan (72.6% versus 75.0%).

Interpretation: The Shera score (online calculator http://appy-risk.org) has the potential to identify a large group of children at low-risk of acute appendicitis, who could be considered for early discharge. Risk scoring does not identify children who should proceed directly to surgery. Medium and high-risk children should undergo routine pre-operative ultrasound imaging by operators trained to assess for acute appendicitis in children.

Funding: None

Introduction

Abdominal pain accounts for around 1 in 20 attendances by children to emergency departments[1]. Although most abdominal pain in children is self-limiting, around 1 in 10 have surgical conditions, the commonest being acute appendicitis[2]. Diagnosis of acute appendicitis is challenging, as clinical presentations vary across age groups, and clinical assessment may be more difficult in younger children[3]. Children presenting with right iliac fossa (RIF) pain are often initially reviewed either by paediatricians, who may lack confidence in assessment for acute appendicitis, or general surgeons in non-specialist units[4] who have limited paediatric training, and may lack confidence in the assessment of children.

Computed tomography (CT) imaging has a high sensitivity and specificity for the diagnosis of acute appendicitis[5]. Although reduced-dose protocols have been developed[6], in contemporary UK practice routine use of CT is limited by concerns about the risks of ionizing radiation in children[7]. This contrasts with the United States where as many as 40% of children with suspected appendicitis undergo CT imaging[8]. Ultrasound is operator dependent, rarely diagnostic, and specialist paediatric ultrasound may not be available outside of paediatric centres[9]. Magnetic resonance imaging (MRI) has limited availability out-of-hours and outside of specialist centres. Diagnostic uncertainty can result in children being admitted to hospital for observation and radiological investigations, increasing the burden on the healthcare system, carers, and children.

Routine clinical risk scoring is recommended by international guidelines[10,11] and has particular relevance to UK practice where use of cross-sectional imaging to investigate suspected appendicitis is limited. Despite a large number of appendicitis risk prediction models having been developed, few surgeons routinely use risk scores to assess children due to limited evidence to support selection of the optimum model and associated cut-off thresholds[12,13]. Robust validation of risk prediction models for the purpose of identifying children at low-risk of acute appendicitis could help to standardise clinical assessment, particularly in non-specialist settings. The aim of this study was to identify the optimum risk prediction model for stratifying acute appendicitis risk in children with right iliac fossa pain at the point of initial surgical assessment.

Methods

This is the primary analysis of the paediatric data captured by the Right Iliac Fossa Treatment (RIFT) Study. This was a prospective, multicentre, observational cohort study which collected data in accordance to a published, peer-reviewed protocol[14]. Hospitals providing acute general surgery in the United Kingdom (UK), Italy, Portugal, the Republic of Ireland, and Spain participated in RIFT, but paediatric data was only collected in the UK and Ireland.

The aim of the RIFT study was to use a multicentre cohort to validate the performance of appendicitis risk prediction models at the point of initial surgical assessment[14]. Prior to data analysis, collaborators agreed that model validation should be performed separately in the paediatric and adult populations. In addition, rather than only validating the most commonly used appendicitis risk prediction models, it was decided that a rapid review should be completed to identify all existing risk prediction models.

The rapid review is reported according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines[15]. The cohort study is reported according to Standards for Reporting Diagnostic Accuracy (STARD) guidelines[16] for diagnostic accuracy studies and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines[17].

Rapid review for risk prediction models

A search for existing risk prediction models for acute appendicitis was performed on MEDLINE and Web of Science. Search terms were [“appendicitis” OR “appendectomy” OR “appendicectomy”] AND [“score” OR “model” OR “nomogram” OR “scoring”]. Search results were supplemented by hand searching reference lists of relevant articles. The search was last updated on 25 July 2018. Studies that reported risk prediction models aimed at differentiating acute appendicitis from all other causes of RIF pain were included. Models that only attempted to differentiate simple from perforated appendicitis were excluded. Risk prediction models developed in either paediatric or adult populations were eligible. Models requiring rectal examination were excluded. Models that relied on standard blood tests such as full blood count, liver function tests or C-reactive protein were included, but those based on non-routine laboratory tests were excluded. Since a key application for risk prediction models could be to stratify patients at the point of initial hospital assessment to determine need for imaging, any models that included radiological parameters were excluded. No date restrictions were applied, but only English language studies were included.

Titles and abstracts were independently screened by two authors, followed by full-text review of selected relevant articles. Disagreements were resolved through discussion with a third author. Data extraction was independently completed by two authors.

Cohort study design and institutional approval

Hospitals providing acute general surgery were invited to contribute to this study. Participating hospitals were not required to standardise diagnostic or patient management pathways, and no changes were made to individual patients' care as part of this observational study. Eligible patients were identified over one of four pre-specified two-week blocks between 13 March 2017 and 18 June 2017.

All consecutive patients referred to the surgical team with acute RIF pain or suspected acute appendicitis were included. This paediatric analysis included children aged 5-15 years. Children who had had abdominal surgery in the preceding 90 days, or had had a previous appendicectomy, right hemicolectomy or total colectomy were excluded.

A standardised data collection proforma was completed at the patient bedside. To avoid placing an unrealistic workload on investigators, data collection was restricted to clinical and biochemical data points required by the most commonly used risk prediction models[10]. Patient-level variables collected included age, sex, clinical symptoms and examination findings, blood tests, and, if applicable, radiological investigations, operative details and histopathology results. Patients were followed up at 30 days following initial assessment using a combination of hospital electronic and paper records.

Institutional approval

In the UK the study was registered at each site as either clinical audit or service evaluation, as it was an observational study only collecting routine, anonymised data with no change to clinical care pathways. In the Republic of Ireland lead investigators at each site were responsible for securing research ethics committee approval, as required by local regulations.

Diagnosis of appendicitis

Patients were recorded as having acute appendicitis if they had a histopathological diagnosis of acute appendicitis within 30 days of index admission. Patients with histopathologically confirmed acute appendicitis were sub-classified as having either simple appendicitis or complex (gangrenous, perforated) appendicitis based on histopathology reports[7]. Patients were excluded if they had an appendicectomy but no histopathology report was available, as it was not possible to determine a final diagnosis (acute appendicitis versus normal appendicectomy).

The normal appendicectomy rate (NAR) was calculated as the proportion of all patients whose appendix was excised who were found to have a histopathologically normal appendix.

Validation of risk prediction models

Risk prediction models were validated if the necessary constituent variables were available in the dataset. The reference standard against which all models were validated was histopathological diagnosis of acute appendicitis within 30 days of initial assessment.

Prior to commencing statistical analysis, collaborators agreed that the greatest clinical need was for a risk prediction model that could support identification of patients at low-risk of acute appendicitis. Therefore, the ideal appendicitis risk prediction model would maximise specificity (the proportion of patients who do not have acute appendicitis identified as being at low risk of acute appendicitis), whilst maintaining an acceptable failure rate (the proportion of patients stratified to the low-risk group who actually have acute appendicitis). The failure rate (false negatives / [true negatives + false negatives]) is the reciprocal of the negative predictive value (NPV), but is preferred as it is easier to interpret in this context.

For a parallel validation of appendicitis risk prediction models in adults, a modified Delphi process was completed to determine the maximum acceptable failure rate[18]. The maximum acceptable failure in adults was identified as 5%. Following discussion with paediatric surgery experts, the same benchmark was adopted *a priori* for this study. For each risk prediction model the score cut-off thresholds for defining the low-risk group were systematically varied in order to identify the maximum achievable specificity whilst maintaining the failure rate under 5%. Overall model discrimination was assessed by calculation of the area under the curve (AUC).

All risk prediction model scores were calculated post-hoc as part of the data analysis, using the raw values for constituent variables that had been prospectively collected at baseline. The primary analysis was performed on the whole dataset. There are distinct differential diagnoses for RIF pain pre-puberty, post-puberty males, and post-puberty females. Therefore, pre-planned sub-group (children aged 5-10 years, females aged 11-15 years, males aged 11-15 years) analyses were performed. The best performing model identified in the primary analysis was separately validated in the three sub-groups. If the failure rate in any of the sub-groups exceeded the pre-defined maximum acceptable rate of 5%, the cut-off threshold for that sub-group would be reduced to decrease the failure rate below 5%.

Having identified the best performing appendicitis risk prediction model and score cut-off for identifying low-risk patients, sensitivity and positive predictive value (PPV) were calculated to determine the model's ability to identify high-risk patients.

Validation of the accuracy of submitted data and sensitivity analyses for missing data are described in S2 Appendix. Additional analyses for baseline characteristics were conducted by presenting simple counts and percentages. The performance of ultrasound by age-sex subgroup was assessed by calculation of AUC, sensitivity, specificity, NPV and PPV.

Analyses were carried out in Stata (Version 15, Stata Corp., College Station, Texas).

Role of the funding source

This study did not receive funding. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Rapid review

The rapid review identified 15 eligible risk prediction models[19-33] that could be validated in the cohort study (Figure 1). Of these, 14 required routine blood tests, whilst one was entirely based on clinical history and examination (Supplemental Table 1). Twelve models were excluded as specific variables required to calculate scores were not available in the study dataset (Supplemental Table 2).

Cohort study

A total of 130 UK hospitals and 9 Irish hospitals contributed data on 1827 children aged 5-15 years (Figure 2). Participating hospitals contributed a median of 11 (interquartile range 5-18) patients each (Supplementary Table 3). A risk score was used for 2.0% (36/1827) of patients at the time of their initial surgical assessment. Across the 19 hospitals where at least one patient was risk scored, a median of 9.4% (interquartile range 3.9%-16.7%) of patients were scored.

Within 30 days of index admission with acute RIF pain, fewer than a third of children aged 5-10 years (31.2%, 259/829) or females aged 11-15 (31.6%, 170/538) underwent surgery. Only half of males aged 11-15 (49.6%, 228/460) underwent surgery. Amongst the patients who had surgery, almost all underwent appendicectomy (95.9%, 630/657). The overall NAR was 15.9% (100/630). The NAR was highest in females aged 11-15 (22.4%, 34/152, Table 1).

Validation of risk prediction models

Ten of the 15 risk prediction models demonstrated very good discrimination for acute appendicitis with AUC>0.8 (Figure 3). The Shera score[30] (Supplementary Table 4) achieved the highest specificity whilst maintaining a failure rate under 5% (Table 2). Across the whole cohort, at a score cut-off value of ≤3, the Shera score had a specificity of 49.2% (579/1176) with a failure rate of 4.8% (29/579).

In the sub-group analyses (Table 3), application of a cut-off ≤3 to the Shera score maintained the failure rate <5% in children aged 5-10 years (failure rate 2.7% [7/257], specificity 47.3% [250/528]) and in females aged 11-15 years (failure rate 3.9% [8/207], specificity 50.3% [199/207]), but in males aged 11-15 years the failure rate was unacceptably high (9.7% [14/144]). Reducing the score cut-off for males aged 11-15 to ≤2, decreased the failure rate in that sub-group to 4.0%.

Therefore, the optimum model for identifying patients at low-risk of acute appendicitis was the Shera score, with score cut-off of ≤3 for children aged 5-10 years and females aged 11-15, and a score cut-off of ≤2 in males aged 11-15. Applying these cut-offs, overall the Shera score

identified 44.3% (521/1176) of patients who did not have acute appendicitis as low-risk, with a failure rate of 3.3% (18/539).

Children stratified as low-risk

Overall, 0.4% (2/539) of children in the low-risk group had complex appendicitis. The most common diagnosis in the low-risk group was non-specific abdominal pain (53.5%, 255/477). Few patients had diagnoses requiring acute management (Supplemental Table 5). Only 11.0% (59/539) of patients in the low-risk group underwent surgery (Table 4), with 50 appendicectomies (84.7%), 7 diagnostic laparoscopies (11.9%), and 2 gynaecological procedures (3.3%) performed. Amongst the patients who underwent appendicectomy, the NAR was 40% (20/50).

Median hospital length of stay (LOS) in low-risk patients who did not have appendicitis was 1 day (interquartile range 1-2 days). Amongst these patients 38.4% (197/513) had LOS of ≥2 days, and 13.8% (71/5513) had LOS of ≥3 days. The most common diagnosis in both patients with LOS of ≥2 days (46.7%, 92/197) and LOS of ≥3 days (33.8%, 24/71) was non-specific abdominal pain.

Children stratified as medium and high-risk

Fewer than half of all children stratified as medium or high-risk (Shera score >3 in children aged 5-10 years and females aged 11-15, and >2 in males aged 11-15) had acute appendicitis (41.4%, 463/1118). In the patients who did not undergo appendicectomy (Supplementary Table 6) the most common diagnoses were non-specific abdominal pain (45.4%, 252/554) and mesenteric adenitis (23.6%, 131/554). A total of 17 patients underwent procedures other than appendicectomy (Table 4).

Amongst the medium and high-risk patients who underwent appendicectomy, 75.7% (426/563) were operated within 24 hours of admission to hospital, 19.4% (109/563) at 24-48 hours, and 5.0% (28/563) beyond 48 hours (time to surgery missing for one child). The overall NAR was 14.2% (80/564). NAR increased from 11.0% (47/426) in children operated within 24 hours of hospital admission, to 24.1% (33/137) in children operated beyond 24 hours.

Shera[30] proposed a cut-off of ≥7 to identify high-risk patients; this cut-off was associated with a PPV of 72.6% (233/321) and a sensitivity of 48.4% (233/481). The Shera score is an adaptation of the Alvarado score[19], which proposed a cut-off ≥9 to identify 'very probable appendicitis'. At a cut-off of ≥9, the PPV was 79.1% (68/86) and the sensitivity was 14.1% (68/481). Risk score performance by age-sex sub-group is summarised in Table 3.

Imaging

Few children underwent either pre-operative CT (0.6%, 10/1827) or MRI (0.7%, 13/1827). Ultrasound was performed for 40.1% (733/1827) of patients. Ultrasound had a high specificity (95.9%, 95% confidence interval 94.0%-97.3%) and NPV (92.7%, 95% CI 90.4%-94.6%). However, the sensitivity (62.0%, 95% CI 52.7%-70.7%) and PPV (75.0%, 95% CI 65.3%-83.1%) for ultrasound were low. In total, 15.4% (97/630) of patients with an ultrasound scan negative for acute appendicitis went on to have an appendicectomy, and 40.2% (39/97) of these procedures resulted in normal appendicectomy. Amongst the patients with abnormal ultrasound findings other than appendicitis, 6.1% (8/131) of patients subsequently underwent appendicectomy with histopathological findings of appendicitis. The diagnostic performance of ultrasound was similar across age-sex sub-groups (Table 5).

Discussion

Main findings

Fewer than half of all children aged 5-15 years in the UK and Ireland assessed by surgeons for acute RIF pain underwent surgery. Amongst the children who underwent appendicectomy the overall NAR was 16%, rising to 22% in females aged 11-15 years. The Shera score[30] was the best performing risk prediction model for acute appendicitis. It identified 44% of children as having a low-risk (1 in 30 chance) of acute appendicitis. Patients in the low-risk group had a very low (1 in 270) incidence of complex (perforated) appendicitis.

The only previous validation of the Shera score was a single centre cohort of around 100 Dutch children, that found the score to have a high NPV (98.5%) and sensitivity (98.4%) at a cut-off score of ≤4[34]. A larger Dutch study evaluated the Alvarado, AIRS (Appendicitis Inflammatory Response Score) and PAS (Paediatric Appendicitis Score) models in children, finding high AUC values (0.82-0.90) but high failure rates of 7-18%[35]. Overall, previous validations of appendicitis risk prediction models in children have been based on retrospective, single centre, or small cohorts, and findings regarding model performance have been inconsistent[12,36]. Model performance is likely to be context-specific, being influenced by baseline prevalence of acute appendicitis in patients presenting with RIF pain. This study represents the most comprehensive evaluation to date of appendicitis risk prediction models in children in the laparoscopic era and its findings are broadly generalisable across the UK and Ireland. The study's findings may also apply in other settings with high baseline NAR and high levels of admissions for non-specific RIF pain, but prior to adoption each health system should separately validate appendicitis risk scores.

Normal appendicectomy and imaging

Although this study suggests that the overall NAR in children in the UK and Ireland has reduced since 2012[4] (16% in 2017 versus 19% in 2012), this remains one of the highest normal appendicectomy rates in the world. Normal appendicectomy rates of 3-5% are typical in the United States and the Netherlands, where a greater proportion of patients undergo CT imaging[8,35,37,38]. In adults CT is the gold standard radiological investigation for acute appendicitis as it offers both high sensitivity and specificity[5,6]. Concern about radiation exposure and associated lifetime risk of malignancy restricts the use of CT imaging in the UK[7,39]. Whereas a multicentre study in the United States identified an overall 42% rate of CT imaging for children[8], in this cohort fewer than 1% of patients underwent CT imaging. MRI imaging has excellent diagnostic performance for assessment of acute appendicitis in children[40,41], but its availability is limited.

Ultrasound was the most frequently used imaging modality. In some settings routine ultrasound may be the most cost-effective approach to diagnosing appendicitis in children[42]. However, it can be difficult to visualise the appendix with ultrasound and it is operator dependent, so ultrasound results are frequently equivocal[9,43]. In this study ultrasound had high specificity, but low sensitivity, suggesting that it is not possible to rule out acute appendicitis based on ultrasound alone. Amongst patients with findings of pathology other than appendicitis on ultrasound, a small proportion were later confirmed to have appendicitis, so a finding of alternative pathology on ultrasound does not exclude the possibility of appendicitis.

Strengths and weaknesses

The use of a simple bedside pro-forma which captured common symptoms and signs, and routine investigations ensured high levels of data completeness, accuracy, and case ascertainment. However, some existing risk prediction models require variables that could only be captured if changes to clinical pathways and patient care were made; this was outside the scope of this study. It is unknown therefore if the models excluded from this study would have superior performance over the Shera score. Several models, such as Alvarado, AIRS and PAS were found to have equivalent AUC values to the Shera score. Based on this study's pre-defined criteria, the Shera score was identified as the single best performing model, but some clinicians may prefer to use Alvarado, AIRS or PAS if they are more familiar with them.

The study only captured patients referred to surgical teams, as its aim was to validate the use of risk prediction models to stratify acute appendicitis risk at point of initial surgical assessment. In the UK most children presenting in general practice or the emergency department for whom appendicitis is a differential are referred for surgical review. However, a small number of children with low-risk symptoms relating to acute RIF pain would have been missed from this study if they were not referred for surgical assessment. Had they been included it is likely that there would have been a greater proportion of true negatives, resulting in lower failure rates and higher specificities. As follow-up was limited to the index hospital where patients first presented, it is possible that patients were readmitted to a different hospital where they were subsequently found to have acute appendicitis. However, this would be the case for a very small number of patients, so this is unlikely to have had a substantial impact on the study's results.

There is likely to be significant variation in diagnostic and management strategies across participating hospitals, but due to the relatively low volume of patients contributed by each hospital it was not possible to explore this. Data were captured during brief data collection windows so it was not possible to establish whether there is seasonal variation in the outcomes we studied.

Implications for policy and clinical practice

The Shera score can be implemented by clinicians, including paediatricians, at the point of initial assessment, without the need for invasive examinations or expensive tests. It could serve as an adjunct to support clinical assessment and shared decision making with carers and children, by identifying a large group of patients who are at low risk of either acute appendicitis or other serious pathology that requires inpatient admission. Systemically well children and their carers may prefer the option of early discharge, rather than inpatient admission or transfer to a specialist centre. There is a risk that a small number of children with patients with appendicitis may be missed, so there should be a low threshold to offer a safety net of ambulatory clinic follow-up. Since short delays to appendicectomy do not increase the risk of perforation[44], serial examination within 24 hours should safely identify any cases of appendicitis missed at initial presentation, without increasing the risk of complications.

Following clinical assessment, some children in the low-risk group will require admission, for example for rehydration, analgesia, or observation and further investigation in the case of unclear diagnoses. Since low-risk patients who undergo appendicectomy are at high risk of normal appendicectomy (40%), ultrasound should be performed as first-line imaging to confirm the diagnosis prior to making the decision to operate (Figure 4). If diagnostic uncertainty persists following ultrasound imaging, cross-sectional imaging should be considered. Preferentially MRI should be performed to avoid exposure to ioinising radiation, but if this is not available low-dose CT protocols should be used.

The Shera score does not achieve a sufficient PPV to select a high-risk group who should proceed directly to surgery. Medium and high-risk patients are frequently admitted for clinical observation, but observation for over 24 hours was found to be associated with increased risk of normal appendicectomy. Routine radiological imaging at the point of initial assessment may be the most effective means of reducing unnecessary admissions and operations in medium and high-risk patients, whilst ensuring children who do have acute appendicitis receive timely treatment (Figure 4). Hospitals should provide a seven-day-a-week ultrasound service for medium and high-risk patients. In order to inform clinical decision making, the diagnostic accuracy of ultrasound should be increased by ensuring that this service is provided by operators specially trained to assess for acute appendicitis in children. This may be achieved either by centralising services or rolling out a national training programme aimed at improving sonographic diagnosis of appendicitis. If the availability of MRI imaging increases in the future, this could become the standard for routine assessment for appendicitis in children, but in the meantime low-dose CT imaging protocols remain the second-line investigation for patients in whom diagnostic uncertainty remains following ultrasound.

Internationally, routine use of appendicitis risk prediction models has been found to reduce imaging, inpatient admission, and normal appendicectomy in both low-risk adults[45,46] and children[47], but there is little evidence for its benefit and safety in the UK and Ireland. Since paediatric appendicectomy in the UK costs around £3,700, increasing imaging rates is likely to be a cost-effective means of reducing potentially unnecessary surgery (Supplementary Table 11). A large-scale multicentre prospective trial is required to evaluate the impact and cost-effectiveness of incorporating routine clinical risk scoring in to clinical pathways for children presenting with acute RIF pain. In this study it was not possible to measure patient and carer satisfaction and this should be a key outcome in future studies.

To support calculation and application of appropriate Shera score cut-offs at the patient bedside, we have developed a mobile, tablet, and desktop compatible web application (http://appy-risk.org).

References

1. Armon K, Stephenson T, Gabriel V, et al. Determining the common medical presenting problems to an accident and emergency department. *Arch Dis Child* 2001; **84**(5): 390-2.

2. Rothrock SG, Pagane J. Acute appendicitis in children: emergency department diagnosis and management. *Ann Emerg Med* 2000; **36**(1): 39-51.

3. Nance ML, Adamson WT, Hedrick HL. Appendicitis in the young child: a continuing diagnostic challenge. *Pediatr Emerg Care* 2000; **16**(3): 160-2.

4. Tiboni S, Bhangu A, Hall NJ, Paediatric Surgery Trainees Research Network, the National Surgical Research Collaborative. Outcome of appendicectomy in children performed in paediatric surgery units compared with general surgery units. *Br J Surg* 2014; **101**(6): 707-14.

5. Kim DW, Yoon HM, Lee JY, et al. Diagnostic performance of CT for pediatric patients with suspected appendicitis in various clinical settings: a systematic review and meta-analysis. *Emerg Radiol* 2018; **25**(6): 627-37.

6. Yoon HM, Suh CH, Cho YA, et al. The diagnostic performance of reduced-dose CT for suspected appendicitis in paediatric and adult patients: A systematic review and diagnostic meta-analysis. *Eur Radiol* 2018; **28**(6): 2537-48.

7. Bhangu A, Soreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet* 2015; **386**(10000): 1278-87.

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

9. Sola R, Jr., Theut SB, Sinclair KA, et al. Standardized reporting of appendicitis-related findings improves reliability of ultrasound in diagnosing appendicitis in children. *J Pediatr Surg* 2018; **53**(5): 984-7.

10. Di Saverio S, Birindelli A, Kelly MD, et al. WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. *World J Emerg Surg* 2016; **11**: 34.

11. Gorter RR, Eker HH, Gorter-Stam MA, et al. Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. *Surg Endosc* 2016; **30**(11): 4668-90.

12. Kularatna M, Lauti M, Haran C, et al. Clinical Prediction Rules for Appendicitis in Adults: Which Is Best? *World J Surg* 2017; **41**(7): 1769-81.

13. Royal College of Surgeons of England (2014). Commissioning guide: Emergency general surgery (acute abdominal pain). Available online at: https://www.rcseng.ac.uk/-/media/files/rcs/standards-and-research/commissioning/commissioning-guide--egs-published-v3.pdf (Accessed 5 February 2018).

14. RIFT Study Group on behalf of the West Midlands Research Collaborative. Right Iliac Fossa Pain Treatment (RIFT) Study: protocol for an international, multicentre, prospective observational study. *BMJ Open* 2018; **8**(1): e017574.

15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

16. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016; **6**(11): e012799.

17. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; **370**(9596): 1453-7.

18. RIFT Study Group on behalf of the West Midlands Research Collaborative. Evaluation of appendicitis risk prediction models in adults with suspected appendicitis. *Br J Surg* 2019.

19. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; **15**(5): 557-64.

20. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg* 2008; **32**(8): 1843-9.

21. Birkhahn RH, Briggs M, Datillo PA, Van Deusen SK, Gaeta TJ. Classifying patients suspected of appendicitis with regard to likelihood. *Am J Surg* 2006; **191**(4): 497-502.

22. Christian F, Christian GP. A simple scoring system to reduce the negative appendicectomy rate. *Ann R Coll Surg Engl* 1992; **74**(4): 281-5.

23. Eskelinen M, Ikonen J, Lipponen P. A computer-based diagnostic score to aid in diagnosis of acute appendicitis. A prospective study of 1333 patients with acute abdominal pain. *Theoretical Surgery* 1992; **7**(2): 86-90.

24. Eskelinen M, Ikonen J, Lipponen P. Sex-specific diagnostic scores for acute appendicitis. *Scand J Gastroenterol* 1994; **29**(1): 59-66.

25. Goh PL. A Simplified Appendicitis Score in the Diagnosis of Acute Appendicitis. *Hong Kong J Emerg Med* 2017; **17**: 230–5.

26. Izbicki JR, Knoefel WT, Wilker DK, et al. Accurate diagnosis of acute appendicitis: a retrospective and prospective analysis of 686 patients. *Eur J Surg* 1992; **158**(4): 227-31.

27. Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. *Ann R Coll Surg Engl* 1994; **76**(6): 418-9.

28. Mikaere H, Zeng I, Lauti M, Kularatna M, MacCormick AD. Derivation and validation of the APPEND score: an acute appendicitis clinical prediction rule. *ANZ J Surg* 2018; **88**(4): E303-E7.

29. Samuel M. Pediatric appendicitis score. *J Pediatr Surg* 2002; **37**(6): 877-81.

30. Shera AH, Nizami FA, Malik AA, Naikoo ZA, Wani MA. Clinical scoring system for diagnosis of acute appendicitis in children. *Indian J Pediatr* 2011; **78**(3): 287-90.

31. Ting HW, Wu JT, Chan CL, Lin SL, Chen MH. Decision model for acute appendicitis treatment with decision tree technology--a modification of the Alvarado scoring system. *J Chin Med Assoc* 2010; **73**(8): 401-6.

32. van den Broek WT, Bijnen BB, Rijbroek B, Gouma DJ. Scoring and diagnostic laparoscopy for suspected appendicitis. *Eur J Surg* 2002; **168**(6): 349-54.

33. Van Way CW, 3rd, Murphy JR, Dunn EL, Elerding SC. A feasibility study of computer aided diagnosis in appendicitis. *Surg Gynecol Obstet* 1982; **155**(5): 685-8.

34. van Amstel P, Gorter RR, van der Lee JH, Cense HA, Bakx R, Heij HA. Ruling out Appendicitis in Children: Can We Use Clinical Prediction Rules? *J Gastrointest Surg* 2019; **23**(10): 2027-48.

35. Macco S, Vrouenraets BC, de Castro SM. Evaluation of scoring systems in predicting acute appendicitis in children. *Surgery* 2016; **160**(6): 1599-604.

36. Frountzas M, Stergios K, Kopsini D, Schizas D, Kontzoglou K, Toutouzas K. Alvarado or RIPASA score for diagnosis of acute appendicitis? A meta-analysis of randomized trials. *Int J Surg* 2018; **56**: 307-14.

37. Saucier A, Huang EY, Emeremni CA, Pershad J. Prospective evaluation of a clinical pathway for suspected appendicitis. *Pediatrics* 2014; **133**(1): e88-95.

38. van Rossem CC, Bolmers MD, Schreinemacher MH, et al. Diagnosing acute appendicitis: surgery or imaging? *Colorectal Dis* 2016; **18**(12): 1129-32.

39. Rogers W, Hoffman J, Noori N. Harms of CT scanning prior to surgery for suspected appendicitis. *Evid Based Med* 2015; **20**(1): 3-4.

40. Duke E, Kalb B, Arif-Tiwari H, et al. A Systematic Review and Meta-Analysis of Diagnostic Performance of MRI for Evaluation of Acute Appendicitis. *AJR Am J Roentgenol* 2016; **206**(3): 508-17.

41. Mushtaq R, Desoky SM, Morello F, et al. First-Line Diagnostic Evaluation with MRI of Children Suspected of Having Acute Appendicitis. *Radiology* 2019; **291**(1): 170-7.

42. Pershad J, Waters TM, Langham MR, Jr., Li T, Huang EY. Cost-effectiveness of diagnostic approaches to suspected appendicitis in children. *J Am Coll Surg* 2015; **220**(4): 738-46.

43. Mittal MK, Dayan PS, Macias CG, et al. Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. *Acad Emerg Med* 2013; **20**(7): 697-702.

44. United Kingdom National Surgical Research Collaborative. Safety of short, in-hospital delays before surgery for acute appendicitis: multicentre cohort study, systematic review, and meta-analysis. *Ann Surg* 2014; **259**(5): 894-903.

45. Andersson M, Kolodziej B, Andersson RE. Randomized clinical trial of Appendicitis Inflammatory Response score-based management of patients with suspected appendicitis. *Br J Surg* 2017; **104**(11): 1451-61.

46. Sammalkorpi HE, Mentula P, Savolainen H, Leppaniemi A. The Introduction of Adult Appendicitis Score Reduced Negative Appendectomy Rate. *Scand J Surg* 2017; **106**(3): 196-201.

47. Shah SR, Sinclair KA, Theut SB, Johnson KM, Holcomb GW, 3rd, St Peter SD. Computed Tomography Utilization for the Diagnosis of Acute Appendicitis in Children Decreases With a Diagnostic Algorithm. *Ann Surg* 2016; **264**(3): 474-81.

Table 1: Patient management stratified by age-sex sub-group

	Both sexes, age 5-10 years	Females, age 11-15 years	Males, age 11-15 years
Total patients	**829**	**538**	**460**
Total patients operated	31.2% (259/829)	31.6% (170/538)	49.6% (228/460)
Appendicectomy performed	97.7% (253/259)	89.4% (152/170)	98.7% (225/228)
Confirmed acute appendicitis	81.0% (205/253)	70.4% (107/152)	81.8% (184/225)
Other appendix pathology	7.1% (18/253)	7.2% (11/152)	2.2% (5/225)
Histologically normal appendix	11.9% (30/253)	22.4% (34/152)	16.0% (36/225)
Total patients not operated	68.8% (570/829)	68.4% (368/538)	50.4% (232/460)

Table 2: Validation and identification of optimal thresholds for risk prediction models

	AUC	Optimal threshold	Failure rate	Specificity
AIRS[20]	0.85 (0.83-0.87	≤3	5.1% (3.6%-7.0%)	57.8% (54.9%- 60.7%)
Alvarado[19]	0.84 (0.83-0.86)	≤3	4.0% (2.5%-6.0%)	44.5% (41.6%-47.4%)
Birkhahn[21]	0.75 (0.73-0.77)	=1	5.2% (3.4%-7.6%)	38.1% (35.3%-40.9%)
Christian[22]	0.79 (0.77-0.81)	=0	0% (0.0%-4.4%)	7.0% (5.6%-8.6%)
Eskelinen, 1992[23]	0.82 (0.80-0.84)	≤47.34	3.9% (2.0%-6.9%)	22.8% (20.5%-25.3%)
Eskelinen, 1994[24]	0.79 (0.76-0.81)	≤-4.76	3.4% (1.7%-5.8%)	26.2% (23.8%-28.7%)
Goh[25]	0.83 (0.81-0.85)	≤1	1.1% (0.1%-3.8%)	15.7% (13.7%-17.9%)
Izbicki[26]	0.80 (0.79-0.82)	≤1	4.1% (2.0%-7.5%)	19.7% (17.5%-22.1%)
Mikaere[28]	0.82 (0.80-0.84)	=0	4.5% (1.9%-8.6%)	14.8% 12.8%-17.0%)
Modified Alvarado[27]	0.84 (0.82-0.86)	≤3	4.8% (3.2%-6.8%)	47.6% (44.7%-50.5%)
PAS[29]	0.84 (0.82-0.86)	≤2	3.1% (1.5%-5.4%)	29.7% (27.1%-32.4%)
Shera[30]	0.84 (0.82-0.86)	≤3	4.8% (3.2%-6.8%)	49.2% (46.3%-52.1%)
Ting[31]	0.68 (0.66-0.70)	=0	10.6% (8.3%-13.2%)	50.3% (47.4%-53.2%)
van der Broek[32]	0.80 (0.78-0.82)	≤1	4.1% (2.4%-6.6%)	31.9% (29.2%-34.6%)
Van Way[33]	0.61 (0.59-0.64)	=36	13.1% (8.0%-20.0%)	9.0% (7.5%-10.6%)

AUC: Area under the curve; AIRS: Appendicitis Inflammatory Response Score; PAS: Pediatric Appendicitis Score

Data presented with 95% confidence intervals

Table 3: Validation of the Shera score in age-sex sub-groups

	AUC	Cut-off	Failure rate	Specificity	Sensitivity	PPV
Children aged 5-10 years	0.85 (0.82-0.88)	≤3	2.7% (1.5%-5.5%)	47.3% (43.0%-51.7%)	96.4% (92.8%-98.6%)	40.5% (36.0%-45.1%)
Females aged 11-15 years	0.84 (0.80-0.88)	≤3	3.9% (1.7%-7.5%)	50.3% (45.2%-55.3%)	92.5% (85.7%-96.7%)	33.2% (27.9%-38.9%)
Males aged 11-15 years	0.83 (0.80-0.87)	≤3	9.7% (5.4%-15.8%)	51.6% (45.2%-57.9%)	92.2% (87.2%-95.7%)	57.5% (51.5%-63.3%)
		≤2	4.0% (0.8%-11.2%)	28.6% (23.1%-34.6%)	98.3% (95.2%-99.7%)	49.4% (44.1%-54.8%)
All patients	0.84 (0.82-0.86)	*	3.3% (2.0%-5.2%)	44.3% (41.4%-47.2%)	96.3% (94.2-97.8%_	41.4% (38.5%-44.4%)
		≤6	18.6% (16.5%-20.8%)	92.5% (90.9%-94.0%)	48.4% (43.9%-53.0%)	72.6% (67.4%-77.4%)
		≤8	26.3% (25.1%-28.5%)	98.5% (97.6%-99.1%)	14.1% (11.1%-17.6%)	79.1% (69.0%-87.1%)

AUC: Area under the curve; PPV: positive predictive value

Data presented with 95% confidence intervals

*Cut-off ≤3 in children aged 5-10 and females aged 11-15; and cut-off ≤2 in males aged 11-15

Table 4: Management by Shera score risk category and age-sex sub-group*

	Both sexes, age 5-10 years	Females, age 11-15 years	Males, age 11-15 years
Low-risk patients*	**n=257**	**n=207**	**n=75**
Total patients operated	8.9% (23/257)	14.0% (29/207)	9.3% (7/75)
Appendicectomy performed†	95.7% (22/23)	75.9% (22/29)	85.7% (6/7)
Simple appendicitis	22.7% (5/22)	36.4% (8/22)	50.0% (3/6)
Perforated appendicitis	9.1% (2/22)	0	0
Other appendix pathology§	27.3% (6/22)	18.2% (4/22)	33.3% (2/6)
Histologically normal appendix	40.9% (9/22)	45.5% (10/22)	16.7% (1/6)
Total patients not operated	91.1% (234/257)	86.0% (178/207)	90.7% (68/75)
Medium and high-risk patients**	**n=467**	**n=295**	**n=356**
Total patients operated	48.4% (226/467)	47.1% (139/295)	60.7% (216/356)
Appendicectomy performed‡	97.8% (221/226)	92,8% (129/139)	99.1% (214/216)
Simple appendicitis	40.3% (89/221)	42.6% (55/129)	51.4% (110/214)
Perforated appendicitis	45.3% (100/221)	33.3% (43/129)	30.8% (66/214)
Other appendix pathology§	5.0% (11/221)	5.4% (7/129)	1.4% (3/214)
Histologically normal appendix	9.5% (21/221)	18.6% (24/129)	16.4% (35/214)
Total patients not operated	51.6% (241/467)	52.9% (156/295)	39.3% (140/356)

*Low-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score ≤3, and males aged 11-15 years with Shera score ≤2

**Medium and high-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score >3, and males aged 11-15 years with Shera score >2

† Procedures performed other than appendicectomy included: 7 diagnostic laparoscopies and 2 gynaecological procedures.

‡ Procedures performed other than appendicectomy included: 8 diagnostic laparoscopies, 3 gynaecological procedures, 2 resections of Meckels diverticulum, 2 urological procedures, 1 drain insertion, and 1 procedure for which details were missing.

§ Amongst the 33 patients with other pathology, there was one carcinoid case, zero adenocarcinoma cases, and zero Crohns cases, with no further detail available for the remaining 32 cases.

Table 5: Ultrasound utilisation and performance by age-sex sub-group

	Proportion of patients scanned*	AUC	Sensitivity	Specificity	NPV	PPV
Children aged 5-10	34.1% (283/829)	0.79 (0.72-0.87)	64.4% (48.8%-78.1%)	94.1% (90.2%-96.7%)	93.3% (89.3%-96.1%)	67.4% (51.5%-80.9%)
Females aged 11-15	59.9% (322/538)	0.74 (0.67-0.82)	50.0% (34.9%-65.1%)	98.5% (96.3%-99.6%)	92.2% (88.5%-95.0%)	85.2% (66.3%-95.8%)
Males aged 11-15	27.8% (128/460)	0.85 (0.77-0.93)	76.7% (57.7%-90.1%)	92.9% (85.8%-97.1%)	92.9% (85.8%-97.1%)	76.7% (57.7%-90.1%)

AUC: Area under the curve; CI: Confidence interval; NPV: Negative predictive value; PPV: Positive predictive value.

Data presented with 95% confidence intervals.

*Proportion of patients scanned by risk score: children aged 5-10 years – 37.0% low risk patients, 38.1% medium/high risk; females aged 11-15 years– 60.4% low risk patients, 62.7% medium/high risk; males aged 11-15 years – 38.7% low risk patients, 27.2% medium/high risk.

Figure 1: Flowchart of study inclusion in the rapid review

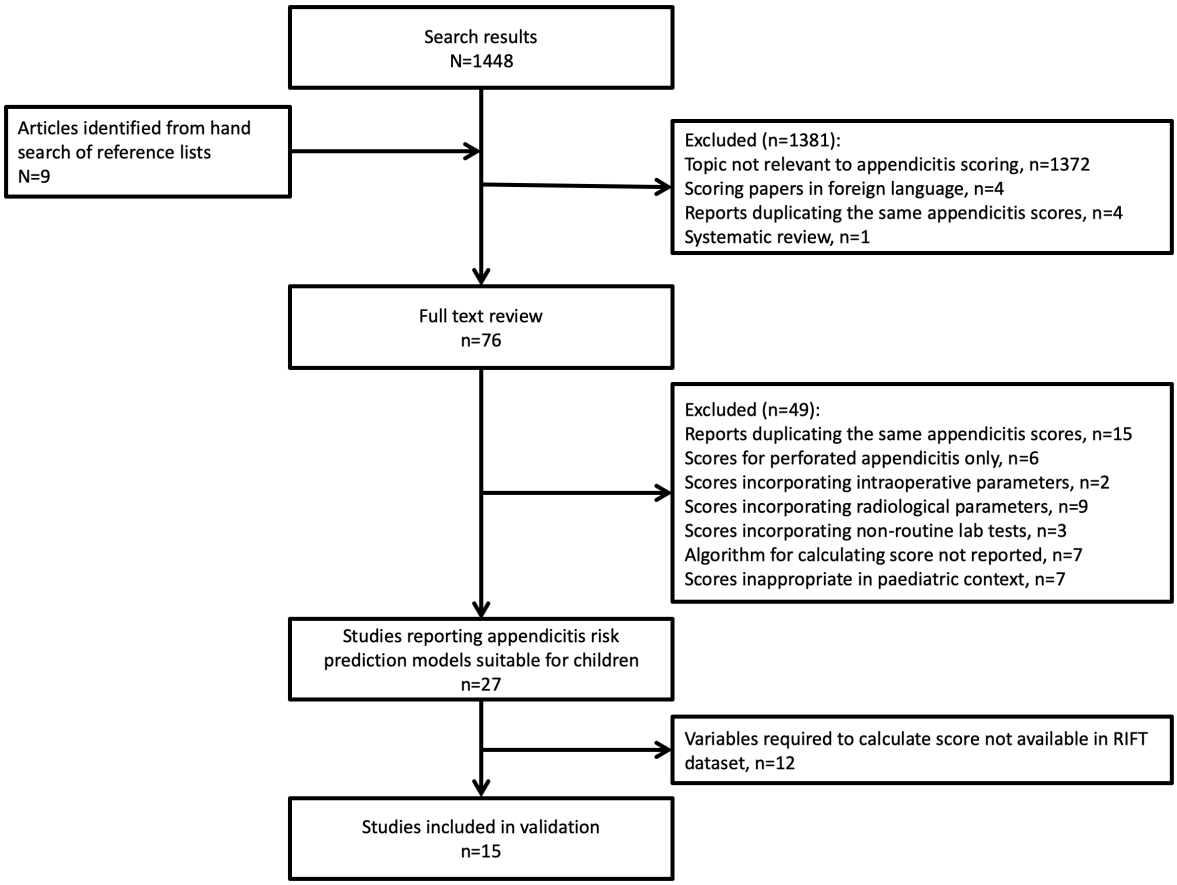
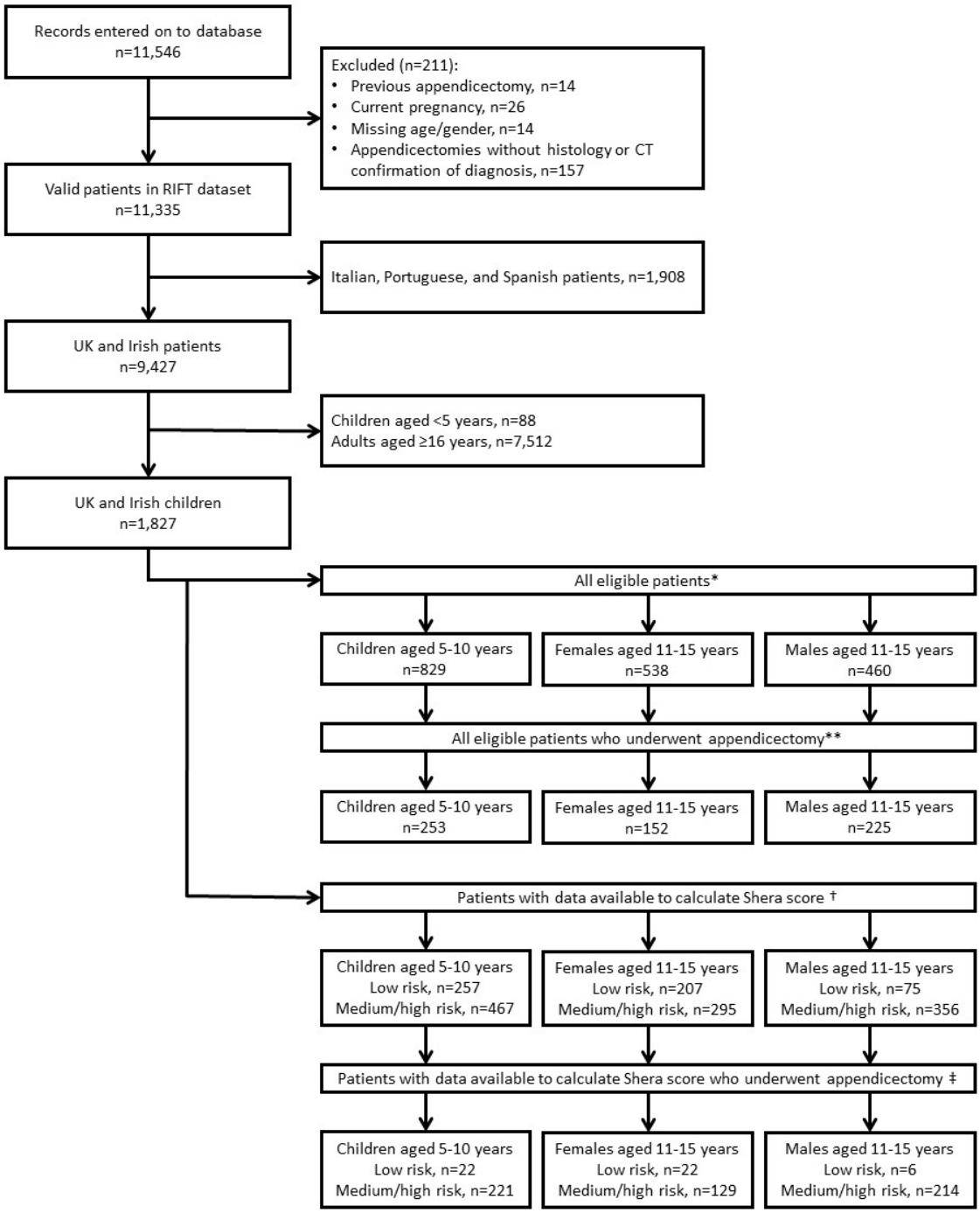


Figure 2: Flowchart of patient inclusion in cohort study



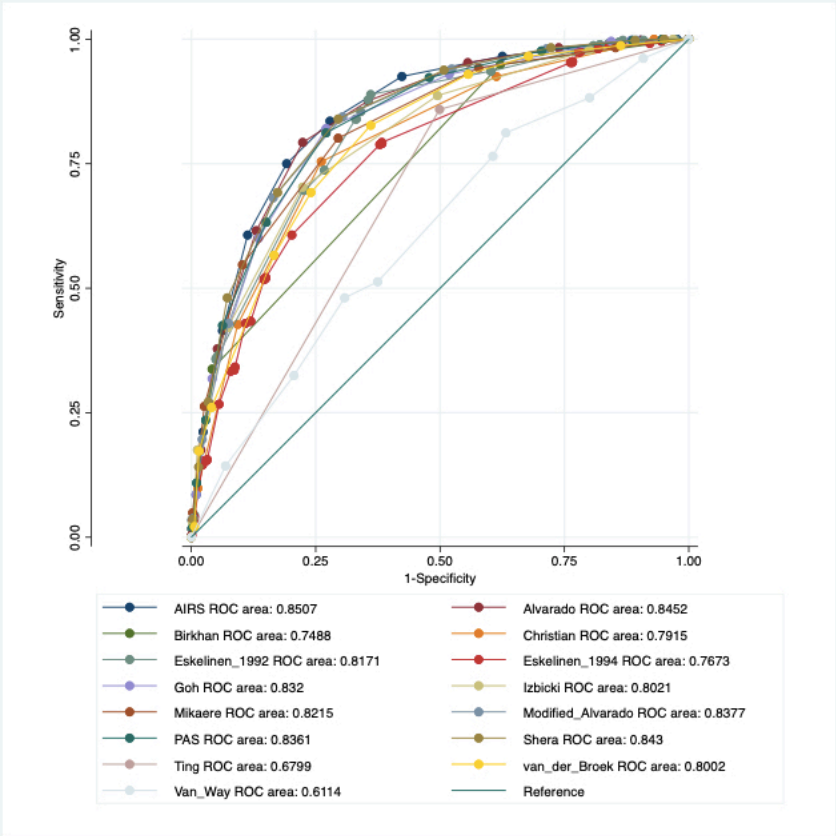
* Denominators for overall description of patient management (Table 1) and evaluation of ultrasound performance (Table 5)

** Denominators for overall NARs (Table 1)

† Denominators for assess of the performance of the Shera score (Table 3) and description of patient management by Shera risk category (Table 4)

‡ Denominators for NAR by Shera risk category (Table 4)

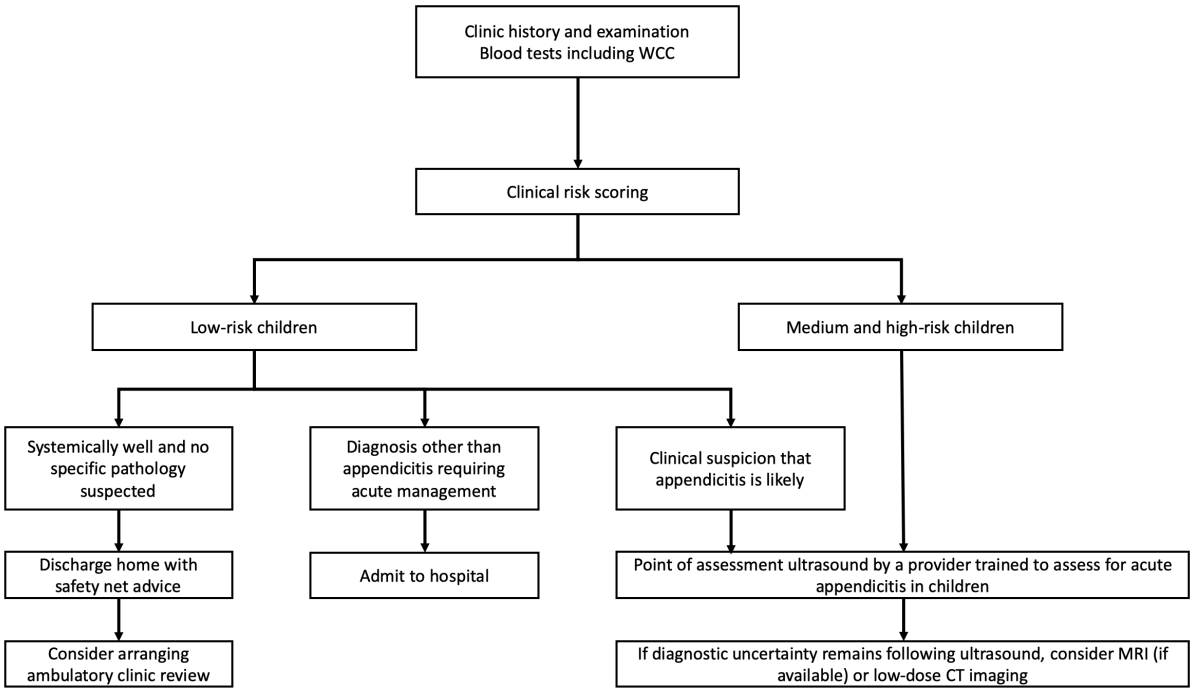
Figure 3: Receiver operating characteristic curves for 15 appendicitis risk prediction models validated in children aged 5-15 years*



AIRS: Appendicitis Inflammatory Response Score; PAS: Pediatric Appendicitis Score; ROC: Receiver operator characteristic

*To produce these ROC curves, only patients with complete data required for all 15 risk prediction models were included. The ROC curve data presented in Table 2 was based on patients who had complete data required for each individual model being validated (so some patients eligible to be included in the analyses in Table 2 were not used when producing the ROC curves in this figure). This accounts for any minor discrepancies between data presented here and Table 2.

Figure 4: Proposed clinical algorithm for children (age 5-15 years) presenting with right iliac fossa pain



CT: computed tomography; MRI: magnetic resonance imaging; WCC: white cell count

Low-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score ≤3, and males aged 11-15 years with Shera score ≤2

Medium and high-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score >3, and males aged 11-15 years with Shera score >2

Supplementary Table 1: Risk prediction models included in validation

Model	Year	Country	Sex	Clinical symptoms							Examination findings			Laboratory tests		
				Symptom duration	Recurrent pain	Anorexia	Nausea/ vomiting	Migration of pain	RIF pain	Fever	RIF tenderness	RIF guarding	Peritonism†	WCC	neutrophilia	CRP
AIRS	2008	Sweden					x		x	x		x	x	x	x	x
Alvarado	1986	USA				x	x	x		x	x		x	x	x	
Birkhahn	2005	USA									x	x	x	x	x	
Christian	1992	India					x		x	x	x			x	x	
Eskelinen, 1992	1992	Finland		x					x		x	x	x	x		
Eskelinen, 1994	1994	Finland	x						x	x	x	x	x			
Goh *	2017	China						x		x	x		x	x		
Izbicki	1992	Germany	x		x			x				x	x	x		
Mikaere	2018	NZ	x			x		x				x			x	x
Modified Alvarado *	1994	UK				x	x	x		x	x		x	x		
PAS	2002	UK				x	x	x		x	x		x	x	x	
Shera *	2010	India				x	x	x		x	x		x	x	x	
Ting *	2010	Taiwan				x	x	x		x	x			x	x	
van der Broek	2002	Holland	x	x						x			x	x		
Van Way	1982	USA	x	x		x	x									

*modifications of the Alvarado Score

AIRS: Appendicitis Inflammatory Response Score, PAS: Paediatric Appendicitis Score, NZ: New Zealand, UK: United Kingdom, USA: United States of America, CRP: C-reactive protein

† Signs of peritonism include rebound tenderness, percussion tenderness, and pain on coughing

Supplementary Table 2: Risk prediction models excluded from validation

Reference	First author	Year	Reason for exclusion
1	Arnbjornsson	1985	Score inappropriate in paediatric context (requires rectal examination)
2	Fenyo	1987	Score inappropriate in paediatric context (requires rectal examination)
3	Teicher	1983	Score inappropriate in paediatric context (includes adult age as variable)
4	Fente	2009	Score inappropriate in paediatric context (requires rectal examination)
5	Jearwattanakanok	2013	Score inappropriate in paediatric context (includes pregnancy as variable)
6	Chong	2010	Score inappropriate in paediatric context (includes adult age as variable)
7	Sammalkorpi	2014	Score inappropriate in paediatric context (includes adult age as variable)
8	Garst	2013	Intra-operative score
9	Reid	2017	Intra-operative score
10	Impellizzeri	2002	Requires a non-routine laboratory test (fibrinogen)
11	Son	2012	Requires a non-routine laboratory test (lipase)
12	Andersson	2014	Requires a non-routine laboratory test (monocyte chemotactic protein [MCP]-1)
13	Atema	2015	Score for distinguishing perforated from non-perforated appendicitis
14	Avanesov	2018	Score for distinguishing perforated from non-perforated appendicitis
15	Blumfield	2017	Score for distinguishing perforated from non-perforated appendicitis
16	Bonadio	2017	Score for distinguishing perforated from non-perforated appendicitis
17	Peng	2006	Score for distinguishing perforated from non-perforated appendicitis
18	Williams	2009	Score for distinguishing perforated from non-perforated appendicitis
19	Leeuwenburgh	2014	Requires radiological investigation
20	Mannil	2018	Requires radiological investigation
21	Tzanakis	2005	Requires radiological investigation
22	Reddy	2017	Requires radiological investigation
23	van den Bogaard	2016	Requires radiological investigation
24	Boettcher	2016	Requires radiological investigation
25	Gorter	2016	Requires radiological investigation
26	Zakaria	2011	Requires radiological investigation
27	Athans	2016	Requires radiological investigation
28	Anupam	2018	Details of how to calculate score not provided

29	Sakai	2007	Details of how to calculate score not provided
30	Chattopadhyay	2012	Details of how to calculate score not provided
31	Hsieh	2011	Details of how to calculate score not provided
32	Sakai	2007	Details of how to calculate score not provided
33	de Dombal	1991	Details of how to calculate score not provided
34	Prabhudesai	2008	Details of how to calculate score not provided
35	Jahn	1997	Unable to validate using RIFT dataset (missing nature of onset of pain, duration of pain <12hr, exacerbation of pain on movement/ coughing, pain intensity)
36	Jawaid	1999	Unable to validate using RIFT dataset (missing initial location of pain)
37	Ahn	2016	Unable to validate using RIFT dataset (missing heel drop test)
38	Wilasrusmee	2017	Unable to validate using RIFT dataset (missing pain intensity, exacerbation of pain on movement/ coughing)
39	Ramirez	1994	Unable to validate using RIFT dataset (missing diarrhoea, location of initial pain)
40	Lindberg	1988	Unable to validate using RIFT dataset (missing duration of pain <12hr, pain outside RIF, pain intensity, exacerbation of pain on coughing)
41	Ohmann	1999	Unable to validate using RIFT dataset (missing genitourinary symptoms, pain intensity)
42	Lintula	2005	Unable to validate using RIFT dataset (missing bowel sounds, pain intensity)
43	Yap	2015	Unable to validate using RIFT dataset (missing constant versus intermittent pain, exacerbation of pain on coughing)
44	Kharbanda	2005	Unable to validate using RIFT dataset (missing location of maximal pain)
45	Kharbanda	2012	Unable to validate using RIFT dataset (missing location of maximal pain, exacerbation of pain on coughing/walking)
46	Kharbanda	2017	Unable to validate using RIFT dataset (missing location of maximal pain, exacerbation of pain on coughing/walking)

1 Arnbjornsson E. Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective versus retrospective studies. Ann. Chir. Gynaecol. 1985;74:159–166
2 Fenyo G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. Acta chirurgica Scandinavica. 1987;153(9):545-51.
3 Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. Ann Surg. 1983;198(6):753-9.
4 Fente BG, Echem RC. Prospective evaluation of the Bengezi and Al-Fallouji modified Alvarado score for presumptive accurate diagnosis of acute appendicitis in University of Port Harcourt Teaching Hospital, Port Harcourt. Nigerian J Med. 2009;18(4):398-401.
5 Jearwattanakanok K, Yamada S, Suntornlimsiri W, Smuthtai W, Patumanond J. Clinical Scoring for Diagnosis of Acute Lower Abdominal Pain in Female of Reproductive Age. Emerg Med Int. 2013;2013:730167
6 Chong CF, Adi MI, Thien A, Suyoi A, Mackie AJ, Tin AS, et al. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. Singapore medical journal. 2010;51(3):220-5.

7 Sammalkorpi HE, Mentula P, Leppäniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis - a prospective study. BMC Gastroenterology. 2014;14(1):114.
8 Garst GC, Moore EE, Banerjee MN, Leopold DK, Burlew CC, Bensard DD, et al. Acute appendicitis: a disease severity score for the acute care surgeon. J Trauma Acute Care Surg. 2013;74(1):32-6.
9 Reid F, Choi J, Williams M, Chan S. Prospective evaluation of the Sunshine Appendicitis Grading System score. ANZ J Surg. 2017;87(5):368-71.
10 Impellizzeri P, Centonze A, Antonuccio P, Turiaco N, Cifala S, Basile M, et al. Utility of a scoring system in the diagnosis of acute appendicitis in pediatric age. A retrospective study. Minerva chirurgica. 2002;57(3):341-6.
11 Son CS, Jang BK, Seo ST, Kim MS, Kim YN. A hybrid decision support model to discover informative knowledge in diagnosing acute appendicitis. BMC medical informatics and decision making. 2012;12:17.
12 Andersson WJS 38:2777-2783 Can new inflammatory markers improve the diagnosis of acute appendicitis?
13 Atema JJ, van Rossem CC, Leeuwenburgh MM, Stoker J, Boermeester MA. Scoring system to distinguish uncomplicated from complicated acute appendicitis. Br J Surg. 2015;102(8):979-90.
14 Diagnostic prediction of complicated appendicitis by combined clinical and radiological appendicitis severity index (APSI).
15 Blumfield E, Yang D, Grossman J. Scoring system for differentiating perforated and non-perforated pediatric appendicitis. Emergency radiology. 2017;24(5):547-54.
16 Bonadio W, Shahid S, Vardi L, Buckingham C, Kornblatt A, Free C, et al. A pre-operative clinical scoring system to distinguish perforation risk with pediatric appendicitis. J Pediatr Surg. 2017.
17 Peng YS, Lee HC, Yeung CY, Sheu JC, Wang NL, Tsai YH. Clinical criteria for diagnosing perforated appendix in pediatric patients. Pediatr Emerg Care. 2006;22(7):475-9.
18 Williams RF, Blakely ML, Fischer PE, Streck CJ, Dassinger MS, Gupta H, et al. Diagnosing ruptured appendicitis preoperatively in pediatric patients. J Am Coll Surg. 2009;208(5):819-25; discussion 26-8.
19 Leeuwenburgh MM, Stockmann HB, Bouma WH, Houdijk AP, Verhagen MF, Vrouenraets B, et al. A simple clinical decision rule to rule out appendicitis in patients with nondiagnostic ultrasound results. Academic Emerg Med. 2014;21(5):488-96.
20 Mannil M, Polysopoulos C, Weishaupt D, Hansmann A. Clinical-radiological scoring system for enhanced diagnosis of acute appendicitis. European journal of radiology. 2018;98:174-8.
21 Tzanakis NE, Efstathiou SP, Danulidis K, Rallis GE, Tsioulos DI, Chatzivasiliou A, Peros G, Nikiteas NI. A New Approach to Accurate Diagnosis of Acute Appendicitis. World J Surg. 2005 Sep;29(9):1151-6
22 Reddy SB, Kelleher M, Bokhari SAJ, Davis KA, Schuster KM. A highly sensitive and specific combined clinical and sonographic score to diagnose appendicitis. J Trauma Acute Care Surg. 2017;83(4):643-9.
23 van den Bogaard VA, Euser SM, van der Ploeg T, de Korte N, Sanders DG, de Winter D, et al. Diagnosing perforated appendicitis in pediatric patients: a new model. J Pediatr Surg. 2016;51(3):444-8.
24 Boettcher M, Breil T, Gunther P. The Heidelberg Appendicitis Score Simplifies Identification of Pediatric Appendicitis. Indian J Pediatr. 2016;83(10):1093-7.
25 Gorter RR, van den Boom AL, Heij HA, Kneepkens CM, Hulsker CC, Tenhagen M, et al. A scoring system to predict the severity of appendicitis in children. J Surg Res. 2016;200(2):452-9.
26 Zakaria O, Sultan TA, Khalil TH, Wahba T. Role of clinical judgment and tissue harmonic imaging ultrasonography in diagnosis of paediatric acute appendicitis. World J Emerg Surg. 2011;6(1):39.
27 Athans BS, Depinet HE, Towbin AJ, Zhang Y, Zhang B, Trout AT. Use of Clinical Data to Predict Appendicitis in Patients with Equivocal US Findings. Radiology. 2016;280(2):557-67.
28 Development and Validation of a Novel Pediatric Appendicitis Risk Calculator (pARC).
29 Sakai S, Kobayashi K, Toyabe S, Mandai N, Kanda T, Akazawa K. Comparison of the levels of accuracy of an artificial neural network model and a logistic regression model for the diagnosis of acute appendicitis. Journal of medical systems. 2007;31(5):357-64.
30 Chattopadhyay S, Rabhi F, Acharya UR, Joshi R, Gajendran R. An approach to model Right Iliac Fossa pain using pain-only-parameters for screening acute appendicitis. Journal of medical systems. 2012;36(3):1491-502.
31 Hsieh CH, Lu RH, Lee NH, Chiu WT, Hsu MH, Li YC. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery. 2011;149(1):87-93.
32 Sakai S, Kobayashi K, Nakamura J, Toyabe S, Akazawa K. Accuracy in the diagnostic prediction of acute appendicitis based on the Bayesian network model. Methods of information in medicine. 2007;46(6):723-6.
33 de Dombal FT. Diagnosis of Acute Abdominal Pain. Churchill Livingstone: Edinburgh, l991, pp 105–106
34 Artificial neural networks: Useful aid in diagnosing acute appendicitis
35 Jahn H, Mathiesen FK, Neckelmann K, Hovendal CP, Bellstrom T, Gottrup F. Comparison of clinical judgment and diagnostic ultrasonography in the diagnosis of acute appendicitis: experience with a score-aided diagnosis. The European journal of surgery = Acta chirurgica. 1997;163(6):433-43.
36 Jawaid A, Asad A, Motiei A, Munir A, Bhutto E, Choudry H, Idrees K, Durrani K, Rahman M, Ahuja M, Nawab Q, Ahmed R, Ali S, Aslam S, Abbasi S, Feerasta S, Alam S, Ahmed U, Jehan I. Clinical scoring system: a valuable tool for decision making in cases of acute appendicitis. J Pak Med Assoc. 1999 Oct;49(10):254-9.
37 Ahn S, Lee H, Choi W, Ahn R, Hong JS, Sohn CH, et al. Clinical Importance of the Heel Drop Test and a New Clinical Score for Adult Appendicitis. PLoS One. 2016;11(10):e0164574.
38 Wilasrusmee C, Siribumrungwong B, Phuwapraisirisan S, Poprom N, Woratanarat P, Lertsithichai P, et al. Developing and validating of Ramathibodi Appendicitis Score (RAMA-AS) for diagnosis of appendicitis in suspected appendicitis patients. World J Emerg Surg. 2017;12:49.

39 Ramirez JM, Deus J. Practical score to aid decision making in doubtful cases of appendicitis. Br J Surg. 1994;81(5):680-3.
40 Lindberg G, Fenyo G. Algorithmic diagnosis of appendicitis using Bayes' theorem and logistic regression. Bayesian statistics 1988;3:665–668
41 Ohmann Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. German Study Group of Acute Abdominal Pain. Arch Surg. 1999 Sep;134(9):993-6.
42 Lintula H, Pesonen E, Kokki H, Vanamo K, Eskelinen M. A diagnostic score for children with suspected appendicitis. Langenbecks Arch Surg. 2005;390(2):164-70.
43 Yap TL, Chen Y, Low WW, Ong CC, Nah SA, Jacobsen AS, et al. A new 2-step risk-stratification clinical score for suspected appendicitis in children. J Pediatr Surg. 2015;50(12):2051-5.
44 Kharbanda AB, Taylor GA, Fishman SJ, Bachur RG. A clinical decision rule to identify children at low risk for appendicitis. Pediatrics. 2005;116(3):709-16.
45 Kharbanda AB, Dudley NC, Bajaj L, Stevenson MD, Macias CG, Mittal MK, et al. Validation and refinement of a prediction rule to identify children at low risk for acute appendicitis. Archives of pediatrics & adolescent medicine. 2012;166(8):738-44.
46 Kharbanda AB, Monuteaux MC, Bachur RG, Dudley NC, Bajaj L, Stevenson MD, et al. A Clinical Score to Predict Appendicitis in Older Male Children. Academic pediatrics. 2017;17(3):261-6.

Supplementary Table 3: Distribution of patient totals contributed per hospital*

Number of patients aged 5-15 years contributed	Number of hospitals
1-4	29
5-9	33
10-19	50
20-29	15
30-39	8
40-49	2
50-54	2

*Participating hospitals contributed a median of 11 (interquartile range 5-18) patients. The number of patients contributed ranged 1-54.

Supplementary Table 4: The Shera score[†]

Variable	Score
Migratory right iliac fossa pain	1
Nausea and/or vomiting	1
Anorexia	1
Right iliac fossa tenderness	2
Pyrexia (temperature ≥37.3^{o}C)	1
Right iliac fossa tenderness on cough/ percussion/ hopping*	2
Leucocytosis (white cell count ≥10 × 10^{9}/L)	1
Neutrophilia (neutrophils as proportion of total white cell count ≥75%)	1

*In this validation this was substituted for right iliac fossa rebound tenderness, which is an equivalent clinical sign

[†]Reference: Shera AH, Nizami FA, Malik AA, Naikoo ZA, Wani MA. Clinical scoring system for diagnosis of acute appendicitis in children. Indian J Pediatr. 2011 Mar;78(3):287-90.

Supplementary Table 5: Diagnoses in low-risk patients who did not have an appendicectomy*

	Total	Both sexes, age 5-10	Females, age 11-15	Males, age 11-15
Gastrointestinal	**144**	**90**	**28**	**26**
Gastritis/ gastroenteritis	39	26	5	8
Meckel's diverticulum	1	1	0	0
Mesenteric adenitis	83	55	18	10
Pancreatitis	1	1	0	0
Colitis	1	0	0	1
Inflammatory bowel syndrome	2	0	0	2
Constipation	14	5	4	5
Adhesional symptoms	1	1	0	0
Intra-abdominal abscess	0	0	0	0
Other gastrointestinal pathology	2	1	1	0
Gynaecological	**35**	**1**	**34**	**-**
Benign ovarian cyst	19	0	19	-
Other ovarian pathology	1	0	1	-
Menstrual pain	13	0	13	-
Other gynaecological pathology	2	1	1	-
Urological	**22**	**10**	**10**	**2**
Testicular/ epididymal pathology	1	1	-	0
Urinary tract infection	19	8	10	1
Other urological pathology	2	1	0	1
Other	**276**	**128**	**109**	**39**
Non-specific abdominal pain	255	116	105	34
Musculoskeletal pain	3	0	2	1
Hernia	2	1	0	1
Lower respiratory tract infection	2	2	0	0
Other miscellaneous pathology	14	9	2	3
Data missing	**12**	**6**	**4**	**2**

*Low-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score ≤3, and males aged 11-15 years with Shera score ≤2

Supplementary Table 6: Diagnoses in medium and high-risk patients who did not have an appendicectomy*

	Total	Both sexes, age 5-10	Females, age 11-15	Males, age 11-15
Gastrointestinal	**204**	**107**	**42**	**55**
Gastritis/ gastroenteritis	52	23	10	19
Meckel's diverticulum	2	2	0	0
Mesenteric adenitis	131	78	24	29
Pancreatitis	0	0	0	0
Colitis	2	0	0	2
Inflammatory bowel syndrome	0	0	0	0
Constipation	14	2	8	4
Adhesional symptoms	0	0	0	0
Intra-abdominal abscess	2	1	0	1
Other gastrointestinal pathology	1	1	0	0
Gynaecological	**24**	**0**	**24**	**-**
Benign ovarian cyst	16	0	16	-
Other ovarian pathology	1	0	1	-
Menstrual pain	7	0	7	-
Other gynaecological pathology	0	0	0	-
Urological	**25**	**13**	**9**	**3**
Testicular/ epididymal pathology	2	0	-	2
Urinary tract infection	21	13	7	1
Other urological pathology	2	0	2	0
Other	**266**	**108**	**84**	**74**
Non-specific abdominal pain	252	100	81	71
Musculoskeletal pain	2	1	0	1
Hernia	0	0	0	0
Lower respiratory tract infection	2	2	0	0
Other miscellaneous pathology	10	5	3	2
Data missing	**35**	**18**	**7**	**10**

*Medium and high-risk defined as children aged 5-10 years and females aged 11-15 years with Shera score >3, and males aged 11-15 years with Shera score >2

Supplementary Table 7: Number of patients missing data for each variable

Data field on case report form	Number of patients missing this variable
Clinical variables	
Previous admissions	0
Duration of pain	3
Presence of RIF pain	2
Migration of pain to RIF	5
Anorexia	4
Nausea	2
Vomiting	2
RIF examination findings	4
Rebound tenderness	4
Temperature	17
Laboratory test variables	
C-Reactive Protein	161
Neutrophil count	173
White cell count	153

RIF: Right iliac fossa

Supplementary Table 8: Association between clinical signs and symptoms and missing laboratory data

Clinical Shera score*	All lab data available†	Odds ratio (95% CI)	p-value
0-1	73.6% (176/239)	-	-
2-3	88.5% (646/730)	2.75 (1.91-3.97)	<0.001
4-8	94.8% (794/838)	6.46 (4.25-9.81)	<0.001

CI: confidence interval

* Calculation of the Shera score based on clinical signs and symptoms (anorexia, nausea, vomiting, fever, migration of right iliac fossa pain, right iliac fossa examination findings) alone, i.e. excluding the laboratory parameters that are usually included in the Shera score. Patients missing clinical variables required to calculate Shera score excluded from this analysis.

† Data available for all three laboratory tests required to validate all risk prediction models (R-Reactive Protein, white cell count, neutrophil count. Amongst patients with one or more laboratory test results not available, 33.0% (63/191) had a clinical Shera score of 0-1, 44.0% (84/191) had a score of 2-3, and 23.0% (44/191) had a score of 4-8.

Supplementary Table 9: Sensitivity analysis with head-to-head comparison of risk prediction models using a constant sample of patients

	Main analysis				Sensitivity analysis			
Model	**AUC**	**Optimal threshold**	**Failure rate**	**Specificity**	**AUC**	**Optimal threshold**	**Failure rate**	**Specificity**
AIRS	0.85 (0.83-0.87)	≤2	3.8% (2.2%-5.9%)	37.6% (34.8%-40.5%)	0.85 (0.83-0.87)	≤2	3.6% (2.1%-5.8%)	37.5% (34.6%-40.3%)
Alvarado	0.84 (0.83-0.86)	≤3	4.0% (2.5%-6.0%)	44.5% (41.6%-47.4%)	0.85 (0.83-0.87)	≤3	4.1% (2.6%-6.2%)	44.4% (41.5%-47.4%)
Birkhahn	0.75 (0.73-0.77)	1	5.2% (3.4%-7.6%)	38.1% (35.3%-40.9%)	0.75 (0.73-0.77)	1	5.2% (3.4%-7.7%)	37.9% (35.1%-40.8%)
Christian	0.79 (0.77-0.81)	0	0% (0.0%-4.4%)	7.0% (5.6%-8.6%)	0.79 (0.77-0.81)	0	0% (0%-4.5%)	7% (5.6%-8.6%)
Eskelinen, 1992	0.82 (0.80-0.84)	≤47.34	3.9% (2.0%-6.9%)	22.8% (20.5%-25.3%)	0.82 (0.80-0.84)	≤47.34	4.0% (2.0%-7.1%)	22.8% (20.4%-25.4%)
Eskelinen, 1994	0.79 (0.76-0.81)	≤-4.76	3.4% (1.7%-5.8%)	26.2% (23.8%-28.7%)	0.77 (0.74-0.79)	≤-4.76	4.5% (2.4%-7.7%)	22.1% (19.8%-24.6%)
Goh	0.83 (0.81-0.85)	≤1	1.1% (0.1%-3.8%)	15.7% (13.7%-17.9%)	0.83 (0.81-0.85)	≤1	1.1% (0.1%-3.9%)	15.6% (13.5%-17.8%)
Izbicki	0.80 (0.79-0.82)	≤1	4.1% (2.0%-7.5%)	19.7% (17.5%-22.1%)	0.80 (0.78-0.82)	≤1	4.2% (2.0%-7.6%)	19.9% (17.7%-22.4%)
Mikaere	0.82 (0.80-0.84)	0	4.5% (1.9%-8.6%)	14.8% 12.8%-17.0%)	0.82 (0.80-0.84)	0	4.5% (2.0%-8.7%)	14.7% (12.7%-16.9%)
Modified Alvarado	0.84 (0.82-0.86)	≤3	4.8% (3.2%-6.8%)	47.6% (44.7%-50.5%)	0.84 (0.82-0.86)	≤3	4.9% (3.3%-7.0%)	47.6% (44.7%-50.6%)
PAS	0.84 (0.82-0.86)	≤2	3.1% (1.5%-5.4%)	29.7% (27.1%-32.4%)	0.84 (0.82-0.86)	≤2	3.1% (1.6%-5.5%)	29.6% (27.0%-32.4%)
Shera	0.84 (0.82-0.86)	≤3	4.8% (3.2%-6.8%)	49.2% (46.3%-52.1%)	0.84 (0.82-0.86)	≤3	4.9% (3.3%-6.9%)	49.2% (46.3%-52.2%)
Ting	0.68 (0.66-0.70)	0	10.6% (8.3%-13.2%)	50.3% (47.4%-53.2%)	0.68 (0.66-0.70)	0	10.3% (8.1%-12.9%)	50.1% (47.2%-53.0%)
van der Broek	0.80 (0.78-0.82)	≤1	4.1% (2.4%-6.6%)	31.9% (29.2%-34.6%)	0.80 (0.78-0.82)	≤1	4.1% (2.4%-6.6%)	32.2% (29.5%-35.0%)
Van Way	0.61 (0.59-0.64)	36	13.1% (8.0%-20.0%)	9.0% (7.5%-10.6%)	0.61 (0.58-0.64)	36	14.6% (8.9%-22.1%)	9.2% (7.5%-11.0%)

AUC: Area under the curve; AIRS: Appendicitis Inflammatory Response Score; PAS: Pediatric Appendicitis Score
Data presented with 95% confidence intervals

Supplementary Table 10: Sensitivity analysis for missing operation field data†

	Cut-off	Failure rate	
		Main analysis	**Sensitivity analysis**
Children aged 5-10 years	≤3	2.7% (7/257)	2.7% (7/256)
Females aged 11-15 years	≤3	3.9% (8/207)	3.9% (8/206)
Males aged 11-15 years	≤2	4.0% (3/75)	4.1% (3/73)
All patients	*	3.3% (18/539)	3.4% (18/535)

† A detailed explanation is provided in Appendix 2. This table presents a sensitivity analysis excluding six patients with missing data for the operation field from calculation of failure rates based on the Shera score. To allow an easier comparison, results of the sensitivity analysis are presented against the main analysis (as per Table 3).

*Cut-off ≤3 in children aged 5-10 and females aged 11-15; and cut-off ≤2 in males aged 11-15

Supplementary Table 11: Estimated cost savings per 1,000 children presenting with suspected appendicitis ¶

	Low-risk patients	Medium/high-risk patients
Total patients	325	675
Ultrasound imaging*		
Additional scans required	13	396
Cost for additional scans	£702	£21,384
Computed tomography imaging**		
Additional scans required	3	99
Additional CT scan cost	£165	£5,445
Normal appendicectomies†		
Total normal appendicectomies	12	48
Potentially avoidable procedures	10	38
Estimated cost savings	£37,370	£142,006
Balance for cost savings‡	£36,503	£115,177

¶ Core recommendations from this study are that all low-risk patients should undergo imaging prior to a decision being made to undertake appendicectomy, and all medium/high-risk patients should undergo imaging at the point of initial assessment. First-line imaging should be with ultrasound. Computerised tomography should be performed preoperatively if diagnostic uncertainty persists. Tabulated data present cost savings associated with implementing these recommendations. Data are presented for 1,000 children aged 5-15 years, based on the RIFT cohort where 32.5% (539/1657) children were stratified as low-risk and 67.5% (1118/1657) were stratified as medium/high risk.

* Additional cost for all patients who did not have pre-operative imaging to have an ultrasound scan. Costs are based on the 2017-18 UK National Health Service National Schedule of Reference Costs. An “ultrasound scan with duration of less than 20 minutes, without contrast” is costed for £54.

** Additional cost for 25% of patients who did not have pre-operative imaging to undergo a computerised tomography scan in addition to ultrasound. Costs are based on the 2017-18 UK National Health Service National Schedule of Reference Costs. A “computerised tomography scan of one area, without contrast, between 6 and 18 years” is costed for £55.

† Assumption that at least 80% of normal appendicectomies are avoidable if preoperative diagnosis is improved. Figure based on the total cost of surgery for 80% of normal appendicectomies. Cost for appendicectomy is based on the 2017-18 UK National Health Service National Schedule of Reference Costs. An “appendicectomy procedures, 18 years and under, with no comorbidity” is costed for £3,737.

‡ Difference between the potential savings on avoidable appendicectomies and the cost for increasing preoperative imaging. The potential benefits of reducing length of hospital stay and avoiding some patients being admitted altogether have not been considered in these calculations, but these would considerably increase overall costs savings identified here.

S1 Appendix: Authorship (all co-authors PubMed citable)

Writing group (investigation; writing – original draft preparation; writing – review & editing): D Nepogodiev*, RJW Wilkin*, C Bradshaw*, C Skerritt, A Ball, W Moni-Nwinia, R Blanco-Colino, P Chauhan, TM Drake, M Frasson, O Gee, JC Glasbey, JH Matthews, GL Morley, DN Naumann, F Pata, AS Soares, A Bhangu (senior author & overall guarantor)
**Joint first authors who contributed equally*

Statistical analysis (data curation; formal analysis): D Nepogodiev

Steering group (conceptualization; methodology; investigation; project administration; writing – review & editing): JH Matthews, GL Morley, S Bhanderi, A Ball, P Chauhan, S Jain, I Mohamed, T Amuthalingam, R Tyler, JC Glasbey, RJW Wilkin, F Pata, M Frasson, R Blanco-Colino, AS Soares, A Bhangu, D Nepogodiev (chair)

Advisory group (conceptualization; methodology; writing – review & editing): EA Griffiths, TD Pinkney, DG Morton

Regional leads (investigation; project administration; writing – review & editing): A Birindelli, NS Blencowe, W Bolton, CJ Bradshaw, CS Chean, S Di Saverio, V Farina, G Harris, JB Haddow, NB Jamieson, S McCain, J Mason, D Milgrom, GR Nana, MS Noor Mohamed, J O'Brien, S Pasquali, JCH Pearce, G Pellino, M Podda, M Rabie, K Sahnan, P Sarmah, A Sgrò, A Simioni, C Skerritt, East Midlands Surgical Academic Network, London Surgical Research Group, Paediatric Surgical Trainee Research Network, West Midlands Research Collaborative

Local leads (investigation; project administration; writing – review & editing): W Moni-Nwinia (Birmingham Children's Hospital); WMA Elzeneini (Bristol Royal Hospital For Children); Z Haveliwala (Cambridge University Hospital); CJ Bradshaw (Chelsea And Westminster; C Skerritt (Evelina Children's Hospital); A Macdonald (Great Ormond Street Hospital); D Fawkner-Corbett (John Radcliffe Hospital); JF Rae (Leicester Royal Infirmary); C Koh (Manchester Children's Hospital); MS Haneef (Queen's Medical Centre, Nottingham); G Winter (Royal Aberdeen Childrens Hospital); C Roy (Royal Hospital For Sick Children Edinburgh); CR Elbourne (Royal London Hospital); CE Jones (Southampton Children's Hospital); C Pardy (St George's Hospital, London); AY Zhang (Letterkenny University Hospital); PM Neary (Mater Misericordiae University Hospital); JA De-Marchi (Our Lady Of Lourdes Hospital); BR O'Connor (Our Lady's Children's Hospital, Dublin); K Wijesundera (Portiuncula Hospital); NM Foley (Sligo University Hospital);ND Kennedy (St Vincent's University Hospital); LA Tiedt (Tallaght Hospital); JC Bolger (University Hospital Limerick); TM Connelly (University Hospital Waterford); OS Ahmed (Wexford General Hospital); S Kabwama (Airedale General Hospital, Keighley); RT Gray (Altnagelvin Area Hospital); IW Kamande (Antrim Area Hospital); S Nazarian (Barnet Hospital); FA Dar (Barnsley District General Hospital); AT Misky (Basildon University Hospital); S Arunachalam (Basingstoke And North Hampshire Hospital); CS Twum-Barima (Bedford Hospital); IM Mohamed (Birmingham Heartlands Hospital); PO Coe (Bradford Royal Infirmary); M Elshaer (Broomfield Hospital); DA Colvin (Causeway Hospital, Coleraine); TEM Morrison (Colchester General Hospital); AJ Thaventhiran (Conquest Hospital, St Leonards On Sea); TM Gilbert (Countess Of Chester Hospital); JM Clements (Craigavon Area Hospital); G Hicks (Croydon University Hospital); S Afshar (Cumberland Infirmary, Carlisle); NC Mckinley (Daisy Hill Hospital, Newry); N Assaf (Darent Valley Hospital, Dartford); T Hanna (Derriford Hospital); EGE Macinnes (Doncaster Royal Infirmary); N Thavanesan (Dorset County Hospital, Dorchester); LK Palani-Velu (Forth Valley Royal Hospital, Labert); LPS Yow (Glan Clwyd Hospital, Rhyl); RW Radwan (Glangwili General Hospital, Carmarthen); M Abdelrahman (Gloucestershire Royal Hospital); KA Lee (Good Hope District General Hospital); Z A Zarka (Great Western Hospital, Swindon); NA Mcdowall (Hairmyres Hospital); CY Tan (Hinchingbrooke Hospital, Huntingdon); ML Venn (Homerton University Hospital); DL Ashmore (Huddersfield Royal Infirmary); A Reddy (James

Cook University Hospital, Middlesbrough); C Delimpalta (James Paget Hospital, Great Yarmouth); P Sarmah (Kettering General Hospital); SM Shah (King George Hospital, Ilford); C Doyle (King's Mill Hospital, Sutton In Ashfield); V Tudyka (Kingston Hospital); HK Brewer (Lincoln County Hospital); EZ Farrow (Lister Hospital, Stevenage); NS Taylor (Luton & Dunstable Hospital); CJ Smart (Macclesfield District General Hospital); C Halkias (Medway Maritime Hospital, Gillingham); P Vitish-Sharma (Milton Keynes University Hospital); NG Mowbray (Morriston Hospital, Swansea); JB Olivier (Musgrove Park Hospital, Taunton); KJ Lee (New Cross Hospital, Wolverhampton); KD Clement (Ninewells Hospital, Dundee); N Panda (North Manchester General Hospital); MJ Connor (North Middlesex University Hospital); SE Fahmy (Northampton General Hospital, Northampton);SO Adegbola (Northwick Park Hospital, Harrow); EM Vaughan (Perth Royal Infirmary); W Stupalkowska (Peterborough City Hospital); LFC Simmonds (Pilgrim Hospital, Boston); A Malik (Pinderfields Hospital); A Hussein (Princess Alexandra Hospital, Harlow); MJ Karim (Princess Of Wales Hospital, Grimsby); T Singhal (Princess Royal University Hospital, Orpington); R Ormiston (Queen Alexandra Hospital, Portsmouth); VWS Kung (Queen Elizabeth Hospital, Woolwich); MA Rabie (Queen Elizabeth The Queen Mother Hospital); N Lal (Queen's Hospital); A Ang (Rotherham District General Hospital); M Orizu (Royal Alexandra Hospital, Paisley); E Gammeri (Royal Berkshire Hospital, Reading); ECS Clough (Royal Bolton Hospital); CH Choy (Royal Cornwall Hospital, Truro); PJJ Herrod (Royal Derby Hospital); S Lawday (Royal Devon & Exeter Hospital); AJ Hann (Royal Free Hospital); D Robinson (Royal Glamorgan Hospital, Ynysmaerdy); BG Wardle (Royal Gwent Hospital, Newport); D Mcdonnell (Royal Hampshire County Hospital, Winchester); DG Rutherford (Royal Infirmary Of Edinburgh); LM Hickey (Royal Lancaster Infirmary); S Rezvani (Royal Oldham Hospital); CR Bell (Royal Preston Hospital); F Mahmood (Royal Stoke University Hospital); G Donaldson (Royal Victoria Hospital); A Peleki (Russells Hall Hospital); OL Sharp (Salisbury District Hospital); S Jain (Sandwell General Hospital, West Bromwich); S Singh (Scarborough General Hospital); DB Thompson (Scunthorpe General Hospital); O El-Tayar (South West Acute, Enniskillen); MP Murray (St Helier Hospital, Carshalton); RJ Mersh (St Mary's Hospital, London); RC Newton (St Peter's Hospital, Chertsey); W Al-Khyatt (St Richard'S Hospital, Chichester); GF Stephens (St Thomas' Hospital); SH Abbas (Stoke Mandeville Hospital, Aylesbury); MR Iqbal (Tunbridge Wells Hospital, Pembury); CE Brown (Ulster Hospital, Dundonald); S Renshaw (University College Hospital); KA Hureibi (University Hospital Coventry and Warwickshire); UP Pullabatla-Venkata (University Hospital Crosshouse, Kilmarnock); N O Donohoe (University Hospital Lewisham); A Myatt (University Hospital Of North Tees); RJ Egan (University Hospital Of Wales, Cardiff); M Trail (Victoria Hospital, Kirkcaldy); SC Mckay (Walsall Manor Hospital); NFT Engall (Warrington Hospital); E Jerome (Warwick Hospital); DC Townsend (Watford General Hospital); BY Patel (West Middlesex University Hospital); JH Choong (Whipps Cross University Hospital); TM Mohamed (Whittington Hospital); NJ Hudson-Peacock (William Harvey Hospital); R Manson (Wishaw General Hospital); S Bhanderi (Worcestershire Royal Hospital); KRM Hebbar (Worthing Hospital); BS Mothe (Wythenshawe Hospital, Manchester); CR Weegenaar (Yeovil District Hospital); CR Bowman (Ysbyty Wrexham Maelor Hospital).

Supervising consultants (investigation; supervision; writing – review & editing): O Gee (Birmingham Children's Hospital); MS Shalaby (Bristol Royal Hospital For Children); CR Jackson (Cambridge University Hospital); SA Clarke (Chelsea And Westminster); N Bouhadiba (Evelina Children's Hospital); K Cross (Great Ormond Street Hospital); S Wagener (John Radcliffe Hospital); H Dagash (Leicester Royal Infirmary); D Wilkinson (Manchester Children's Hospital); DW Colliver (Queen's Medical Centre, Nottingham); Y Patel (Royal Aberdeen Childrens Hospital); A Harris (Royal Hospital For Sick Children Edinburgh); SJ Cleeve (Royal London Hospital); NJ Hall (Southampton Children's Hospital); D Rex (St George's Hospital, London); MA Aremu (Letterkenny University Hospital); J Mulsow (Mater Misericordiae University Hospital); S El-Masry (Our Lady Of Lourdes Hospital); J Gillick (Our Lady's Children's Hospital, Dublin); J Garvin (Portiuncula Hospital); M Caldwell (Sligo University Hospital); DC Winter (St Vincent's University Hospital); CB Peirce (University Hospital Limerick); F Cooke (University Hospital Waterford); K Mealy (Wexford General

Hospital); L Whittaker (Airedale General Hospital, Keighley); AL Gidwani (Altnagelvin Area Hospital); CK Byrnes (Antrim Area Hospital); SMF Saunders (Barnet Hospital); MH Shiwani (Barnsley District General Hospital); N Ashraf (Basildon University Hospital); AK Venkatasubramaniam (Basingstoke And North Hampshire Hospital); KE Bevan (Bedford Hospital); D Mcarthur (Birmingham Heartlands Hospital); JP Griffith (Bradford Royal Infirmary);A Charalabopoulos (Broomfield Hospital); W Campbell (Causeway Hospital, Coleraine); D Menzies (Colchester General Hospital); D Jenner (Conquest Hospital, St Leonards On Sea); NJ Eardley (Countess Of Chester Hospital); S Yoong (Craigavon Area Hospital); M Abulafi (Croydon University Hospital); H Avalapati (Cumberland Infirmary, Carlisle); R Thompson (Daisy Hill Hospital, Newry); P Nastro (Darent Valley Hospital, Dartford); R Kochupapy (Derriford Hospital); D Beral (Doncaster Royal Infirmary); BM Stubbs (Dorset County Hospital, Dorchester); J Crozier (Forth Valley Royal Hospital, Labert); V Pento (Glan Clwyd Hospital, Rhyl); WD Beasley (Glangwili General Hospital, Carmarthen); ST Hornby (Gloucestershire Royal Hospital); H Youssef (Good Hope District General Hospital); R Alexander (Great Western Hospital, Swindon); S Denley (Hairmyres Hospital); F Di Franco (Hinchingbrooke Hospital, Huntingdon); A Quddus (Homerton University Hospital); A Saha (Huddersfield Royal Infirmary); A Reddy (James Cook University Hospital, Middlesbrough); VR Velchuru (James Paget Hospital, Great Yarmouth); Y Salama (Kettering General Hospital); A Bhargava (King George Hospital, Ilford);N Watson (King's Mill Hospital, Sutton In Ashfield); V Tudyka (Kingston Hospital); PG Dunning (Lincoln County Hospital); Geraint Lloyd (Lister Hospital, Stevenage); SV Gurjar (Luton & Dunstable Hospital); CJ Smart (Macclesfield District General Hospital); B Andrews (Medway Maritime Hospital, Gillingham); A Singh (Milton Keynes University Hospital); G Shingler (Morriston Hospital, Swansea); D Mahon (Musgrove Park Hospital, Taunton); S Elgaddal (New Cross Hospital, Wolverhampton); CJ Payne (Ninewells Hospital, Dundee); S Dalmia (North Manchester General Hospital); MS Nair (North Middlesex University Hospital); JG Finch (Northampton General Hospital, Northampton); J Watfah (Northwick Park Hospital, Harrow); CA Carden (Perth Royal Infirmary); R Makhija (Peterborough City Hospital); M Rao (Pilgrim Hospital, Boston); J Sarveswaran (Pinderfields Hospital); V Vijay (Princess Alexandra Hospital, Harlow); S Rekhraj (Princess Of Wales Hospital, Grimsby); T Singhal (Princess Royal University Hospital, Orpington); B Knight (Queen Alexandra Hospital, Portsmouth); MN Siddiqui (Queen Elizabeth Hospital, Woolwich); JF Sebastian (Queen Elizabeth The Queen Mother Hospital); S Vakis (Queen's Hospital); R Rajaram (Rotherham District General Hospital); J Gray (Royal Alexandra Hospital, Paisley); D Mcgrath (Royal Berkshire Hospital, Reading); G Faulkner (Royal Bolton Hospital); S Gopalswamy (Royal Cornwall Hospital, Truro); JN Lund (Royal Derby Hospital); NJ Smart, I Daniels (Royal Devon & Exeter Hospital); M Varcada (Royal Free Hospital); A Woodward (Royal Glamorgan Hospital, Ynysmaerdy); GL Williams (Royal Gwent Hospital, Newport); K Szentpali (Royal Hampshire County Hospital, Winchester); R Ravindran (Royal Infirmary Of Edinburgh); C Bronder (Royal Lancaster Infirmary); A Rate (Royal Oldham Hospital); V Daya Shetty (Royal Preston Hospital); VSR Rao (Royal Stoke University Hospital); B Clements (Royal Victoria Hospital); RT Patel (Russells Hall Hospital); G Branagan (Salisbury District Hospital); A Torrance (Sandwell General Hospital, West Bromwich); K Maude (Scarborough General Hospital); G Kaur (Scunthorpe General Hospital); AZ Zadi (St Helier Hospital, Carshalton); J Kinross (St Mary's Hospital, London); S Irukulla (St Peter's Hospital, Chertsey); W Hawkins (St Richard'S Hospital, Chichester); S Wheatstone (St Thomas' Hospital); T Magro (Stoke Mandeville Hospital, Aylesbury); STR Bailey (Tunbridge Wells Hospital, Pembury); G Marshall (Ulster Hospital, Dundonald); J Mccullough (University College Hospital); G Marangoni (University Hospital Coventry and Warwickshire); EL Leung (University Hospital Crosshouse, Kilmarnock); CM Borg (University Hospital Lewisham); S Gopinath (University Hospital Of North Tees); J Torkington (University Hospital Of Wales, Cardiff); S Yalamarthi (Victoria Hospital, Kirkcaldy); S Mirza (Walsall Manor Hospital); M Brett (Warrington Hospital); S Ramcharan (Warwick Hospital); V Pandey (Watford General Hospital); B Thava (West Middlesex University Hospital); SM Andreani (Whipps Cross University Hospital); SJ Sahay (Whittington Hospital); B Aravind (William Harvey Hospital); M Downey (Wishaw General Hospital); D Nicol (Worcestershire Royal Hospital); P Whitehouse (Worthing Hospital); A

Sharma (Wythenshawe Hospital, Manchester); N Francis (Yeovil District Hospital); DJ Stewart (Ysbyty Wrexham Maelor Hospital).

Collaborators (investigation; writing – review & editing): B Pisavadia, Y Abou El Ella, W Moni-Nwinia, A Champsi (Birmingham Children's Hospital); WMA Elzeneini, DL Hughes, EJ Griggs, HL Fuller (Bristol Royal Hospital For Children); Z Haveliwala (Cambridge University Hospital); SJ Mccluney, CJ Bradshaw, JE McEntee (Chelsea And Westminster); S Kumar, ND Kennedy (Evelina Children's Hospital); B Ertansel, A Macdonald (Great Ormond Street Hospital); D Fawkner-Corbett, C Shah, K Ford, EP Redman (John Radcliffe Hospital); LT Henderson, S Patel (Leicester Royal Infirmary); A Pal, C Koh (Manchester Children's Hospital); SG Tiboni, B Johnson, TD Boam, M Haneef, HS Perera, OH Elsharnoby, A La Valle (Queen's Medical Centre, Nottingham); R Smith, G Winter (Royal Aberdeen Childrens Hospital); E Teasdale, L Solinas, A Harris (Royal Hospital For Sick Children Edinburgh); JD Foster, CP Major, A Budzanowski, L Tullie, SL Lee, K Sloan, R King, A Kelay (Southampton Children's Hospital); C Pardy, A Mears, J Galea (St George's Hospital, London); U Kabir, CLL Wong, LS Rahmani, S Tan, S Chng, B Jasinski, SA Cheng, S Mardhiah, K McGlynn (Letterkenny University Hospital);C Hayes, AMW Fox (Mater Misericordiae University Hospital); A Zaborowski, RM Mitru, JA De-Marchi, A Mc-Dermott (Our Lady Of Lourdes Hospital); D Coyle, I Stoica, SV McMahon, BR O'Connor, D Mc Laughlin, M Kannegieser-Bailey, R Murphy, A Muntean (Our Lady's Children's Hospital, Dublin); K Wijesundera, S Shet, L Thomas (Portiuncula Hospital); S Quill, A Aljorfi, NM Foley (Sligo University Hospital);DP Ahern, AE Redmond (St Vincent's University Hospital); A Szucs, SE Edwards, LA Tiedt, P Manoharan, S Brennan (Tallaght Hospital); JC Bolger, AM Abdelgadir, KL Mckevitt, MA Zarog (University Hospital Limerick); G Ahmed, W Bukhari (University Hospital Waterford); A Ahad, OS Ahmed (Wexford General Hospital); N Aruparayil, P Christopoulos, B Chambers (Airedale General Hospital, Keighley); N O'Neill, RH Long, RT Gray, P Mccaughey, MJ Lynch Wong, RS Mccain, HC Lennox-Warburton, C Moore, KM Manektella (Altnagelvin Area Hospital); IW Kamande, S Mcilwaine, S Rupani, DJ Simpson, J Wauchope, M Ng, L Christian, A Crone (Antrim Area Hospital); JB Haddow, R Sacks, S Nazarian, NRA Symons, A Lazzaro (Barnet Hospital); SD Patil, FA Dar, S Roomi (Barnsley District General Hospital); I Silva, JM Hodgson, C Ly, AT Misky, H Froud, H Patel (Basildon University Hospital); S Arunachalam, P Cay, RS Karwal, S Arunachalam, D Danquah-Boateng (Basingstoke And North Hampshire Hospital); BJM Berry, HD Esmail, CS Twum-Barima, H Maripi, D Bilku, MA McKelvie, K Miller, A Maina (Bedford Hospital); R Velho, R Hasan, R Clingan, SMH Jah, K Waite, A Jones (Birmingham Heartlands Hospital); S Buckley-Jones, L Lecky-Thompson; N Saghir, S Mansoor, D Mistry, R Brown, A Wong, S Gurung (Bradford Royal Infirmary);M Elshaer, G Govind, O Niaz (Broomfield Hospital); DA Colvin, ARS Dyal (Causeway Hospital, Coleraine); TEM Morrison, MJ Mitchard, B Ribeiro, H Merai, EA Dean, S Khan (Colchester General Hospital); A Baginski, C Mann, W Foers, L Jones, B Woodward (Conquest Hospital, St Leonards On Sea); DM Mcwhirter, AT Thomas, TG Gilbert, LW Weatherburn, JP Pilkington, FC Cameron, TM Gilbert, CS Chean (Countess Of Chester Hospital); JM Clements, JD Clements, C Mccann, S Davidson, L Hackney, JS Clements (Craigavon Area Hospital); A Martin, G Hicks, DT Du (Croydon University Hospital); Z Shakoor, SK Yen, M Adnan, S Ranathunga, S Sana, YH Tay, MY Chin, M Gillespie (Cumberland Infirmary, Carlisle); AG Brown, U Campbell (Daisy Hill Hospital, Newry); N Assaf, M Chatzikonstantinou (Darent Valley Hospital, Dartford); B Mahendran, T Flack (Derriford Hospital); M Chowdhary, JM Lim, EGE Macinnes, E Whiteman, JA Shepherd, A Pedder (Doncaster Royal Infirmary); N Thavanesan, KL Siggens, CW Lai, V Morrison-Jones, Z Hayat (Dorset County Hospital, Dorchester); A Brown, G Neagle, LK Palani-Velu (Forth Valley Royal Hospital, Labert); LPS Yow, V Pento, RE Adams, PL Davies, SY Wong, R Amarasinghe, LE Tague, E Jones (Glan Clwyd Hospital, Rhyl); RW Radwan, A Martin, E Jones (Glangwili General Hospital, Carmarthen); NS Blencowe, M Abdelrahman, EJ Wilson, OO Awokoya, SN Griffiths, EN Kirkham, AE Cotton, A Adimonye, PA Leighton (Gloucestershire Royal Hospital); KA Lee, A Abdelrahman, H Cartwright, Z Gates, M Miguras (Good Hope District General Hospital); ZA Zarka, K Khan, C Louw, T Grove, T Badenoch, J Mckeon (Great Western Hospital, Swindon); NA Mcdowall, CS Wood, RP Leitch (Hairmyres

Hospital); P Sgardelis, MI Perera, D Nagarajan, Y Malam (Hinchingbrooke Hospital, Huntingdon); K Theodoropoulou, ML Venn, S Rajagopal, S Kaptanis, D Popova, O Olagbaiye, S Tayeh, S Rigby (Homerton University Hospital); DL Ashmore, MP Harris (Huddersfield Royal Infirmary); V Kanakala, A Hollingsworth, KR Paramasevon, J Milward, S Ahmed (James Cook University Hospital, Middlesbrough); C Delimpalta, BF Fanibi, N Ferguson (James Paget Hospital, Great Yarmouth); P Sharma, N Islam, V Tabain, R Keegan, P Sarmah, J Ahel, T Alhammali (Kettering General Hospital); J Graveston, EJ Balai, J Graveston, SM Shah, K Rothnie (King George Hospital, Ilford); C Doyle, S Dharamavaram, M Seth, F Chowdhury (King's Mill Hospital, Sutton In Ashfield); M Rezacova, N Seneviratne, EJ Turner (Kingston Hospital); GR Nana, V Rodger, N Ting, HK Brewer, D Photiou (Lincoln County Hospital); EZ Farrow, D Taze, S Lodhia, L Earnshaw, K Kumar, A Neale, J Bastianpillai, M Cipparrone, A Barrie (Lister Hospital, Stevenage); NS Taylor, Z Nash, L Anandan, K Tailor, Z Vinnicombe, S Krivan, R Kuo, A Giorga, H Habib (Luton & Dunstable Hospital); K Malik, M Bogdan (Macclesfield District General Hospital); O Ashour, E Iosif, M Fadel, K Gopalakrishnan, N Orme, S Williams, M Rashid (Medway Maritime Hospital, Gillingham); A Sultana, N Patel (Milton Keynes University Hospital); T Yasin, V Bevan, NG Mowbray, B Al-Sarireh (Morriston Hospital, Swansea); JB Olivier, M Brown, N Mohd, EE Howie, M Poudevigne, C Paget (Musgrove Park Hospital, Taunton); H Rallage, K Chui, F Fawzi, S Layman, E Okorocha, GM Jama (New Cross Hospital, Wolverhampton); P Orawiec, O Kouli, A Hassane, KD Clement, J Kilkenny, AA Devine (Ninewells Hospital, Dundee); N Panda, S Hajibandeh, J Shah, MM Mansour, SN Malik, S Davis, F Trew, SK Bandyopadhyay, K Dart (North Manchester General Hospital); MJ Connor, S Guru-Naidu, R Callan, MK Nair (North Middlesex University Hospital); M Alani, E Sezen, S Salim, SE Fahmy (Northampton General Hospital, Northampton); A Acharya, A Jain, SO Adegbola, K Sahnan, PJ Tozer, J Warusavitarne (Northwick Park Hospital, Harrow); KM Emslie, O Collier-Wakefield, P Sivaloganathan, C Dobson, M Elseedawy, EM Vaughan (Perth Royal Infirmary); W Stupalkowska, L McNally, M Williams, FH Motiwala, S Choi, A Ahmad Asmadi, D Burnside, A Everden, S Suriyakumar,L Sandu (Peterborough City Hospital); DA Kent, J Bowen, LFC Simmonds, P Long, A Khair (Pilgrim Hospital, Boston); A Malik, K Shah, L Phelan (Pinderfields Hospital); R Pierre, AA Dhari, A Hussein, M Hoff (Princess Alexandra Hospital, Harlow); MJ Karim, S Nickson, T Setshwaelo, A Chalk, MJ Parkola, A Harlinska (Princess Of Wales Hospital, Grimsby); TCW Chan, JG Dudek, R Rolph (Princess Royal University Hospital, Orpington); M Allen, H Pollard, R Ormiston, R Gormely, H Finlayson, G Ljungqvist (Queen Alexandra Hospital, Portsmouth); VWS Kung, EL Springate, W Barnieh, AS Patel, ZA Siddiqui, IA Chishti, J Ayube-Brown (Queen Elizabeth Hospital, Woolwich); MR Rabie, L Blake, E Yardimci, S Nagendram (Queen Elizabeth The Queen Mother Hospital); N Lal, J Evans, N Hussain (Queen's Hospital); A Ang (Rotherham District General Hospital); M Orizu, R Stevenson, A Nair, C Murphy, O Mcgucken (Royal Alexandra Hospital, Paisley); E Gammeri, R Pandya, H Bowerman, G Lafaurie, GI Van Boxel, K Shanmugarajah, D Maragouthakis, Z Hanif (Royal Berkshire Hospital, Reading); JD Evans, S Yoganathan, TD Richardson (Royal Bolton Hospital); CH Choy, V Cook, GL Clark, B Rigney, E O'Neill, J Guliani (Royal Cornwall Hospital, Truro); PJJ Herrod, D Chan, F Harper, T Sian, C Boereboom, JEM Blackwell, E Hardy, H Boyd-Carson, DG Couch,CA Barter,SN Thoukididou,JR Hatt(Royal Derby Hospital); CS Jones, S Dean, N Rajaretnam, S Lawday, M Masood, N Thakral, D Griffith, C Doherty, A Longshaw (Royal Devon & Exeter Hospital); D Peprah, G Mathew, AJ Hann (Royal Free Hospital); A Hook, D Robinson, J Vance-Daniel (Royal Glamorgan Hospital, Ynysmaerdy); YMS Ibrahim, BG Wardle, KJ Walters, HE Whewell, MA Sherif, M Mckenna, JCH Pearce (Royal Gwent Hospital, Newport); D O'Sullivan, C Woodrow (Royal Hampshire County Hospital, Winchester); S Gill, A Johnstone, R Gentry, R Irwin, A Forgie, S Welsh (Royal Infirmary Of Edinburgh); LM Hickey, P Ivey, JK Bullivant (Royal Lancaster Infirmary);N Sihra, D Shreshta (Royal Oldham Hospital); B Chong, A Nadeem, JA Fasuyi, MM Patel (Royal Preston Hospital); R Daureeawoo, B Okekunle, M Cheruvu, E Mazumdar, F Mahmood, A Hussain, C Patel, M Mcquaid, A Banks,A Robinson (Royal Stoke University Hospital); G Donaldson, I Ghaffari, J Moradzadeh (Royal Victoria Hospital); A Peleki, M Kamal, M Gulamhussein (Russells Hall Hospital); L Pearce, E Gaines, S Ghatorae, OL Sharp, A Savill, B Hutchinson, J Chapman, F Wu, W Creasy, M Raymond, S

Grosvenor (Salisbury District Hospital); A Odeh, Y Malik, H Bansal (Sandwell General Hospital, West Bromwich); C Grant, A Raofi, B Ahmed, D Mai (Scarborough General Hospital); DB Thompson, JEH Souter, RN Hamelmann, S Ikram (Scunthorpe General Hospital); O El-Tayar, A De-Manzoni-Garberini (South West Acute, Enniskillen); M Oakey, H Smoker (St Helier Hospital, Carshalton); PR Boshier, M Fehervari, S Easdon, RJ Mersh, T Ilozue, ME Adam (St Mary's Hospital, London); RC Newton, S Jokhan, A Foster, K Nambiar, P Bohra, S Janardanan, V Shanmuganathan, F Maqboul (St Peter's Hospital, Chertsey); C Ettles, SD Wardle, E Martinou (St Richard'S Hospital, Chichester); GF Stephens, A Khasria, R Bagga, D Motter, N Mundkur, Y Pan (St Thomas' Hospital); SH Abbas, K Akbari, SM Farrell, A Rahim, A Gummaraju (Stoke Mandeville Hospital, Aylesbury); MR Iqbal, A Mahmoud, O Akinsola, N Smallcombe, M Tarazi, C Hanley (Tunbridge Wells Hospital, Pembury); CE Brown, UM Campbell, NC Mckinley (Ulster Hospital, Dundonald); S Renshaw, D Franklin, JR Davidson (University College Hospital); KA Hureibi, SS Raza, A Krishnamoorthy, Y Rajjoub, M Ali, TC Seddon, RE Payne, A Das (University Hospital Coventry and Warwickshire); LM Martin, KN Naismith, UP Pullabatla Venkata, VM Manda (University Hospital Crosshouse, Kilmarnock); N O Donohoe, KM Burns, L Huang (University Hospital Lewisham); M Samuel, JA Docherty (University Hospital Of North Tees); WL Cheah, R Ooi, M Nyeko-Lacek, L Marsh, A Prideaux, CH Li, A Poacher, M Lee, M Muzaffar,RJ Egan (University Hospital Of Wales, Cardiff); H Sunter, NPG Roth, C Roy, D Mcmorran, A Turnbull, M Trail (Victoria Hospital, Kirkcaldy); GR Layton, JE Archer, P Yang (Walsall Manor Hospital); NFT Engall (Warrington Hospital); L Henderson, E Douka, V Amin, K Borghol, P Marriott (Warwick Hospital); DC Townsend, OD Blackford, S Bond, B Baker, WA Mohamed (Watford General Hospital); R Williams, J Garnham, BY Patel, HD Robb, J Allington, L Cloney (West Middlesex University Hospital);JH Choong, SM Chok (Whipps Cross University Hospital); GB Peiris, TM Mohamed, S Gunning, E Nimako, D Pandya (Whittington Hospital); P Hever, A Amayo, C Bull, NJ Hudson-Peacock, C Clements, M Al-Sheikh (William Harvey Hospital); F Savioli, M Long, E Horsfield, C Robertson, S Ogboru, AC Mcilwrath (Wishaw General Hospital); S Bhanderi, S Bailey, J Bell (Worcestershire Royal Hospital); KRM Hebbar, C Limb, N Obeid, JE Rich, A Balasubramaniam (Worthing Hospital); BS Mothe, R Mashar (Wythenshawe Hospital, Manchester); M Taylor, JR Schuster Bruce, CR Weegenaar, G Dennison, NJ Curtis (Yeovil District Hospital);S Ng, CFW Carder, A Abdelwahed, CR Bowman, SB Chandler, BJ Tinsley (Ysbyty Wrexham Maelor Hospital).

Data validators (validation; writing – review & editing): RL Cresner (Chelsea And Westminster); A Suliman (Evelina Children's Hospital); F Cooper (Royal Aberdeen Childrens Hospital); A Alshafei (Our Lady's Children's Hospital, Dublin); N Varma (Addenbrooke's Hospital); L Cousins (Altnagelvin Area Hospital); A Busuttil (Basildon University Hospital); C Baird (Birmingham Heartlands Hospital); N Drye (Bradford Royal Infirmary);S Mansour (Broomfield Hospital); O Anderson (Colchester General Hospital); B Woodward (Conquest Hospital, St Leonards On Sea); R Mahapatra (Countess Of Chester Hospital); JA Clements (Craigavon Area Hospital); N D'Souza (Croydon University Hospital); R Thompson (Daisy Hill Hospital, Newry);AM Tang (Glangwili General Hospital, Carmarthen); BE Byrne (Gloucestershire Royal Hospital); CY Tan (Hinchingbrooke Hospital, Huntingdon); P Cunha (Homerton University Hospital); C Ogbuokiri (Huddersfield Royal Infirmary); T Alhammali (Kettering General Hospital); F Chowdhury (King's Mill Hospital, Sutton In Ashfield); H Kho (Lincoln County Hospital); S Lodhia (Lister Hospital, Stevenage); T Doulias (Medway Maritime Hospital, Gillingham); MSJ Wilson (Ninewells Hospital, Dundee); N Panda (North Manchester General Hospital); J Wolff (Northampton General Hospital, Northampton); K Akhtar (Peterborough City Hospital); JM Tognarelli (Queen Elizabeth Hospital, Woolwich); SA Hussaini (Rotherham District General Hospital); E Wright (Royal Alexandra Hospital, Paisley); E Gammeri (Royal Berkshire Hospital, Reading); H Brewer (Royal Derby Hospital); S Rinkoff (Royal Free Hospital); RL Harries (Royal Gwent Hospital, Newport); CJ Fairfield (Royal Infirmary Of Edinburgh); LM Hickey (Royal Lancaster Infirmary); A Jackson (Scunthorpe General Hospital); HL Wright (St Thomas' Hospital); U Walters (University College Hospital); K Carney (University Hospital Of North Tees); PC Logan (Walsall Manor Hospital); Z Mughal

(University Hospital Coventry and Warwickshire); E Strachan (West Middlesex University Hospital); SM Andreani (Whipps Cross University Hospital); B Chasty (Whittington Hospital); A Sharma (Wythenshawe Hospital, Manchester); JKY Ma (Ysbyty Wrexham Maelor Hospital).

Rapid review: D Nepogodiev (search, resolution of disagreements during study selection), P Chauhan (study screening, data extraction), D Naumann (study screening), A Ball (data extraction).

S2 Appendix: Data quality and sensitivity analyses for missing data

Data accuracy

Independent data validators who had not been involved in the initial data collection determined data accuracy rates by review of the following key data fields against the original clinical records for enrolled patients: (1) whether the patient had undergone surgery; (2) whether the patient had been readmitted within 30 days of index admission; (3) histopathological results, if applicable. The data accuracy rate was defined as a proportion of validated data fields that was correctly recorded. Any incorrect data identified were amended on the database by the validators.

Independent validators examined 29.9% (1281/4284) of eligible data fields, determining data accuracy to be 99.0% (1268/1281).

Data completeness

Data completeness was calculated as the proportion of data fields required for calculation of the validated appendicitis risk prediction scores that were missing in the overall dataset.

Data completeness for the 13 variables required to validate the risk prediction models was 97.8% (23223/23751).

Overall, 211 patients did not have data available for at least one variable required to calculate one or more of the risk prediction models. The number of patients missing each variable is reported in Supplementary Table 7. The data fields most commonly not available were neutrophil count (WCC, n=173), c-reactive protein (CRP, n=161), and white cell count (n=153). In total 195 patients had one or more blood test results not available. There were very low rates of missing data for the non-laboratory variables. Just 16 patients were excluded from risk prediction model validation due to missing clinical variables (i.e. patients with all laboratory test data available, but one or more missing clinical variables).

Missing data for variables required to calculate risk scores

Laboratory data (CRP, WCC, neutrophils) can be retrieved relatively easily from electronic records. So rather than investigators having ‘missed’ this data during the data collection process, it is likely that these blood tests were not actually performed as part of routine clinical care [this observational study did not require patients to undergo additional tests] and so it was not possible to collect this data at all.

One possible explanation for blood tests not being performed is that they may be less likely to be performed in younger children, in whom there is a preference to avoid invasive tests. This is reflected in more children aged 5-10 years than children aged 11-15 years missing one or

more laboratory tests (13.5% [112/829] versus 8.3% [83/998], respectively, p<0.001). Although this imbalance could theoretically have influenced the overall validation result, this has been accounted for by stratification of Shera score performance by age-sex sub-group (Table 3).

Another possibility is that blood tests are less likely to be performed in children for whom there is a lesser index of suspicion of a serious pathology. This is supported by children for whom one or more laboratory tests were missing having a lower incidence of appendicitis than those children for whom all three tests were available in the dataset (12.3% [24/195] versus 28.9% [472/1632], respectively, p<0.001). Furthermore, more than three quarters of patients missing laboratory test data had three or fewer clinical signs/symptoms (Supplementary Table 8). Conversely, the odds of a patients having all three laboratory tests available significantly increased the more clinical signs and symptoms the patient had (Supplementary Table 8). Therefore, it is possible that a large proportion of the patients excluded from the risk prediction model validations are likely to have been 'true negatives' (low-scoring patients [few clinical signs/symptoms] who did not have appendicitis). The effect of including these patients in the risk prediction model validations would have been to decrease the failure rate, as the proportion of true negative versus false negatives would increase. Therefore, exclusion of these patients may have underestimated performance of the models, with the failure rates for the Shera score at the selected cut-offs remaining <5%.

Head-to-head comparison of risk prediction models using a constant sample of patients

In order to maximise the sample size and precision of our estimates, and to minimise possible residual bias, in the main analysis each risk prediction model was validated based on all patients who had complete data for the variables required to calculate that risk score. Consequently, the sample sizes for validation of the different risk scores varied. In order to enable a head-to-head comparison of the risk prediction models based on a constant sample of participants, we have performed a sensitivity analysis that only includes children who had data available to calculate all the risk prediction models. Of the 1,827 total children, it was possible to calculate all risk prediction models for 1,616 children.

The results of this sensitivity analysis are reported in Supplementary Table 9. The same optimum threshold was identified in the sensitivity analysis for each of the 15 models, as had been identified in the main analysis. For 13 risk prediction models the area under the curves (AUC) in the main and sensitivity analyses were identical; for the Alvarado score the AUC slightly increased from 0.84 in the main analysis to 0.85 in the sensitivity analysis; and for the 1994 Eskelinen score the AUC slightly decreased from 0.79 in the main analysis to 0.77 in the sensitivity analysis. In the sensitivity analysis, the Shera score achieved the highest specificity

(49.2%) whilst maintaining a failure rate under 5%. Therefore, the finding regarding the best performing risk prediction model from the main analysis was robust in the head-to-head comparison of models.

Sensitivity analyses for missing data required to calculate the Shera score

Since surgical trainee collaborative cohort studies have been completed with very low rates of missing data, an incomplete data rate of under 2.5% was anticipated, so imputation of missing data was not planned. We performed a complete case analysis (listwise deletion) and pre-planned sensitivity analyses: (1) missing data points scored as zero, representing what would happen in normal clinical practice and providing a scenario that would underestimate individual patient acute appendicitis risk; (2) missing data scored with maximum applicable points.

Overall, applying the Shera score with cut-off of ≤3 for children aged 5-10 years and females aged 11-15, and a cut-off of ≤2 in males aged 11-15, the failure rate was to 3.0% in both sensitivity analyses, whilst the specificity ranged 43.7-48.9%.

In the sensitivity analyses, the failure rate was found to range 2.3-2.6% in children aged 5-10, 3.4-3.7% in females aged 11-15 years, and 3.4-3.8% in males aged 11-15 years. Specificity was found to range 47.3-53.7% in children aged 5-10, 48.5-53.4% in females aged 11-15 years, and 27.9-31.2% in males aged 11-15 years.

Sensitivity analyses for missing data for the operation field

In this observational study all patients meeting eligibility criteria were included at the point of presenting to the surgical team. Since patient consent was not required in this study, there was no non-participation due to patients withholding consent.

The follow-up pathway was designed to be minimise the burden on investigators and maximise data returns. Patients were followed-up at 30-days by review of in-hospital electronic records, with no requirement for additional patient contact.

For 14 patients data was missing regarding whether they had undergone surgery within 30 days of index admission. Branching logic in the REDCap database in which data was collected required completion of the operation field to access any of the operation specific fields (what operation was completed, preoperative delay, intraoperative findings, histology), therefore it was unlikely that collaborators would miss completion of the operation field by accident for patients who had undergone surgery, as they would be alerted to the omission by the inability to access any of the operation-specific data fields. It was assumed therefore that patients who had missing data for the operation field did not undergo surgery.

Of the 14 patients missing data for the operation field, 5 were also missing data for white cell count and so would have been excluded from the key analyses validating risk prediction model

performance due to inability to calculate the risk scores. Of the remaining 9 patients, the records were locked and confirmed for 3 patients, indicating that the data input had been reviewed and confirmed (at many sites through the data validation process this was completed by a different investigator to the person who initially entered the data). This suggests that it is unlikely that these patients underwent surgery as the lack of operation-specific data input would have been flagged at the data locking stage. For the remaining six patients there is less certainty around whether the original assumption is correct that they did not undergo surgery.

Since to meet the criteria for appendicitis in this study patients had to be recorded as having undergone surgery with subsequent histopathological examination of their appendix and confirmation of the diagnosis, these six patients were coded as not having had appendicitis. In order to determine whether the coding of these six patients made a difference to the results of the Shera score validation a sensitivity analysis was conducted. In this analysis the six patients were excluded and the failure rate for the Shera score was recalculated for each age-sex group and overall. Of the six patients with a missing operation field, four patients had low Shera scores and two patients had medium/high Shera scores. The results of the sensitivity analyses are reported in Supplementary Table 10, demonstrating that exclusion of the six patients made no discernible difference to the results.