Evaluation of appendicitis risk prediction models in adults with suspected appendicitis

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

Correspondence to: Mr A. Bhangu, Academic Department of Surgery, University of Birmingham, Heritage Building, Mindelsohn Way, Birmingham B15 2TH, UK (e-mail: a.a.bhangu@bham.ac.uk)

Background: Appendicitis is the most common general surgical emergency worldwide, but its diagnosis remains challenging. The aim of this study was to determine whether existing risk prediction models can reliably identify patients presenting to hospital in the UK with acute right iliac fossa (RIF) pain who are at low risk of appendicitis.

Methods: A systematic search was completed to identify all existing appendicitis risk prediction models. Models were validated using UK data from an international prospective cohort study that captured consecutive patients aged 16–45 years presenting to hospital with acute RIF in March to June 2017. The main outcome was best achievable model specificity (proportion of patients who did not have appendicitis correctly classified as low risk) whilst maintaining a failure rate below 5 per cent (proportion of patients identified as low risk who actually had appendicitis).

Results: Some 5345 patients across 154 UK hospitals were identified, of which two-thirds (3613 of 5345, 67·6 per cent) were women. Women were more than twice as likely to undergo surgery with removal of a histologically normal appendix (272 of 964, 28·2 per cent) than men (120 of 993, 12·1 per cent) (relative risk 2·33, 95 per cent c.i. 1·92 to 2·84; $P < 0·001$). Of 15 validated risk prediction models, the Adult Appendicitis Score performed best (cut-off score 8 or less, specificity 63·1 per cent, failure rate 3·7 per cent). The Appendicitis Inflammatory Response Score performed best for men (cut-off score 2 or less, specificity 24·7 per cent, failure rate 2·4 per cent).

Conclusion: Women in the UK had a disproportionate risk of admission without surgical intervention and had high rates of normal appendicectomy. Risk prediction models to support shared decision-making by identifying adults in the UK at low risk of appendicitis were identified.

Introduction

Acute appendicitis is the most common general surgical emergency worldwide[1], but its diagnosis remains challenging, particularly in young women for whom there is a broader range of differential diagnoses than for men[2]. A key concern is overtreatment, in the form of normal appendicectomy (removal of a histologically normal appendix), which may be associated with increased postoperative complications, duration of hospital stay and healthcare costs compared with diagnostic laparoscopy alone[3−5]. There is little consensus on optimal diagnostic pathways for patients presenting with acute right iliac fossa

(RIF) pain in the UK. Traditionally, concerns over the radiation exposure associated with CT have limited its routine use[6]. Although modern low-dose protocols have reduced radiation exposure whilst maintaining diagnostic performance[7,8], CT rates are lower in the UK than in many other high-income countries, and this is associated with higher normal appendicectomy rates (NARs)[9].

To improve diagnosis of appendicitis, international guidelines[10,11] recommend routine clinical risk scoring. Although the Appendicitis Inflammatory Response (AIRS) and Alvarado scores are recommended most frequently, there is little evidence to support their application, as they

have been inconsistently and poorly validated[12]. None is in routine clinical use[13]. Consequently, although guidelines recommend that patients at low risk of appendicitis should not be admitted routinely to hospital[13], diagnostic uncertainty may lead to patients being admitted for observation, increasing the likelihood of overtreatment and healthcare-related costs. Robustly validated risk prediction models that identify groups at low risk of appendicitis could help to standardize clinical assessment and inform patient and doctor decision-making. In turn, this may reduce hospital admissions, overtreatment and healthcare costs[14].

To influence wide-scale change, a prospective multicentre study including patients presenting with RIF pain was developed to assess whether existing risk prediction models can identify low-risk patients suitable for ambulatory management. This study represents the first planned analysis from the Right Iliac Fossa Treatment (RIFT) Study Group[15]. The aim of this study was to identify the optimal risk prediction model to identify young patients (aged 16–45 years) in the UK at low risk of acute appendicitis. Children and adults aged over 45 years follow distinct management pathways and will be addressed in subsequent preplanned analyses.

Methods

This study was undertaken in two parts. First, a systematic literature search was performed to identify all available risk prediction models for acute appendicitis. This is reported according to the PRISMA guidelines[16]. Second, a multicentre prospective observational cohort study was performed to collect accurate data for validation of these risk prediction models. This is reported according to Standards for Reporting Diagnostic Accuracy (STARD) guidelines[17] for diagnostic accuracy studies.

The RIFT Study[15] captured data in the UK, Italy, Portugal, Republic of Ireland and Spain. Clinical risk score validation was preplanned to be performed in the UK for patients presenting with acute RIF pain. National differences in clinical pathways mean that analyses must be stratified by country. At the time of planning the analysis, the NAR was anticipated to be around 20 per cent in the UK[18], but under 5 per cent in other European countries[14,19]. Given the low baseline NAR, there is little clinical need for risk scoring in European populations, so validation of risk scores in those patients would be unlikely to change clinical practice. Therefore, clinical risk score validation was preplanned to be performed in the UK only, although observed NARs from across all participating countries are presented for context. Future analyses

are planned to explore variation in imaging rates across Europe.

Identifying risk scoring models

Systematic searches of MEDLINE and Web of Science were performed using the search terms ['appendicitis' OR 'appendectomy' OR 'appendicectomy'] AND ['score' OR 'model' OR 'scoring' OR 'nomogram']. These results were supplemented by additional unstructured searches of Google Scholar. Reference lists from relevant articles were hand-searched for eligible studies, as were recent World Society of Emergency Surgery and Royal College of Surgeons (England) Emergency Surgery guidelines[10,13]. No date restrictions were set. The search was last updated on 25 July 2018.

Studies were eligible for inclusion if they described a risk prediction model for the diagnosis of acute appendicitis that used clinical data and/or routine laboratory tests (full blood count, C-reactive protein, liver function tests). As the aim of the study was to identify model(s) that could be used at initial surgical assessment, models were excluded if they relied on radiological investigations or non-routine laboratory tests that are not typically available at the point of first surgical contact. Models that aimed to differentiate simple from complex appendicitis were also excluded. Only English-language articles that provided sufficient information to replicate their clinical risk model algorithm were included. Titles and abstracts were screened, followed by review of full texts of relevant articles. Study selection and data extraction were completed independently by two authors, with any disagreements resolved through discussion with a third author.

Study dissemination

Data collection was conducted according to a prespecified, published protocol[15]. The protocol was disseminated through surgical trainee-led research collaboratives. Any hospital providing acute general surgical services was eligible to participate.

Study approval

As this observational study collected routine, anonymized data with no change to clinical care pathways, lead investigators at participating UK centres registered the study locally as either clinical audit or service evaluation. In the Republic of Ireland, Italy, Portugal and Spain, local lead investigators were responsible for arranging research ethics committee or institutional approval locally, as appropriate.

Patient selection

Eligible patients were identified from participating centres during one of four prespecified 2-week study periods between 13 March 2017 and 18 June 2017. During each interval, all consecutive patients referred by a general practitioner or emergency physician to the on-call surgeon's team with suspected appendicitis or acute RIF pain were identified at the point of admission to the surgical unit. Patients who had previously undergone either a therapeutic appendicectomy or an incidental appendicectomy as part of another procedure were excluded. Pregnant women have significantly different clinical needs to other patients and were therefore excluded. Pregnancy was identified by patients' self-report. The possibility of pregnancy was further excluded by most women undergoing urinalysis. Patients who underwent appendicectomy for whom histological findings were not available were excluded, as it was not possible to determine the underlying diagnosis (appendicitis *versus* normal appendicectomy) for them. Patients who were managed without surgery for acute appendicitis following a positive CT or MRI diagnosis of appendicitis were also excluded, as the diagnosis was not confirmed histologically.

Diagnosis of appendicitis

A diagnosis of acute appendicitis was confirmed if within 30 days of enrolment in the study the patient had excision of the appendix with postoperative histological examination confirming acute appendicitis. Patients who underwent right hemicolectomy for presumed acute appendicitis were pooled with those who had appendicectomy. Based on the original histopathology report, patients were classified as having either simple or complex (gangrenous, perforated) appendicitis[2]. The NAR was calculated as patients with normal appendix histology as a proportion of all patients who had an appendicectomy. Patients with appendix pathology other than appendicitis (such as appendix tumour) were included in the denominator but not the numerator.

Data collection

Data were collected using standardized case report forms (CRFs) by teams of up to three investigators per 2-week period. The CRF was designed to be completed at the patient bedside. A large number of variables have been proposed to predict appendicitis. Therefore, to ensure feasibility of data collection, the CRF was designed to collect the data points required for the four most common adult risk prediction models identified in international guidelines[10].

Patient-level variables collected included age, sex, clinical symptoms and examination findings, urinalysis and blood test results. Data were collected on ultrasound, CT or MRI use, and whether these tests were positive for appendicitis (diagnosis of appendicitis in the formal radiology report). If the patient had surgery, the procedure, operative findings and histopathology results were recorded. Patients were followed during their initial admission, during any subsequent hospital admissions within 30 days of initial presentation, and then at 30 days after initial presentation using a combination of electronic and paper hospital records.

Data integrity

Multiple strategies were used to ensure accurate data collection. A supervising consultant surgeon at each hospital oversaw study conduct and was responsible for overall quality assurance of submitted data. To ensure data completeness, before locking of the online database, local lead investigators were contacted with specific details of missing data. Participating sites voluntarily identified independent data validators who had not been involved in the initial data collection. Data accuracy was determined by review of the following key data fields against the original clinical records for enrolled patients: whether the patient had undergone surgery; whether the patient had been readmitted within 30 days of index admission; and histopathological results, if applicable. The data accuracy rate was defined as the proportion of validated data fields that was recorded correctly. Where incorrect data were identified, validators were asked to amend those data points on the study database.

Validation of risk prediction models and statistics

The development of the statistical analysis plan is summarized in *Appendix S2* (supporting information).

Risk prediction models were validated if patients could be scored with the data points available. As there are distinct differential diagnoses for RIF pain in women, it was preplanned to stratify risk prediction model validation by sex.

For maximum clinical impact, the ideal risk score would classify as many patients as possible as being at low risk of appendicitis (true negatives), but not at the expense of missing significant numbers of patients with appendicitis (false negatives). The clinical performance of each score was therefore evaluated in terms of failure rate and specificity. The failure rate is the false-negative rate: the proportion of patients stratified to the low-risk group who actually have appendicitis (false negatives/(true negatives + false negatives)). Before analysis, a modified Delphi exercise (*Appendix S2*, supporting information) amongst 24

experienced UK general surgeons agreed the maximum acceptable failure rate to be 5 per cent. Specificity is the proportion of patients who do not have appendicitis who were stratified to the low-risk group (true negatives/(true negatives + false positives)).

The main outcome measure for evaluating each risk prediction model was the best achievable specificity whilst maintaining a failure rate of less than 5 per cent. The overall ability of the risk prediction models to discriminate between patients with and without acute appendicitis was determined by calculation of the area under the receiver operating characteristic (ROC) curve (AUC).

Analyses were carried out in Stata® version 15 (Stata-Corp, College Station, Texas, USA).

Handling of missing data

As surgical collaborative cohort studies have been completed with very low rates of missing data, an overall incomplete data rate of under 2·5 per cent was anticipated. Therefore, there was no plan to impute missing data. A complete case analysis (list-wise deletion) was performed, and preplanned sensitivity analyses: missing data points

scored as zero, representing what would happen in normal clinical practice and providing a scenario that would underestimate appendicitis risk; missing data were scored with maximum applicable points.

Results

The systematic search identified 26 risk prediction models (*Fig. 1*). Fifteen of these could be validated with the data collected in the cohort study[20–34] (*Table 1*). These risk prediction models were based on 17 clinical parameters (*Table S1*, supporting information). The other 11 models could not be validated owing to the data set lacking specific variables[35–45] (*Table S2*, supporting information).

Normal appendicectomy rate benchmarking across participating countries

The overall NAR in patients aged 16–45 years was 20·0 per cent (392 of 1957) in the UK and 6.2 per cent (54 of 868) in the other countries. In women aged 16–45 years, the NAR was 28·2 per cent (272 of 964) in the UK, 5·0 per cent (6 of 121) in Italy, 6 per cent (3 of 51) in Portugal, 25 per cent (15
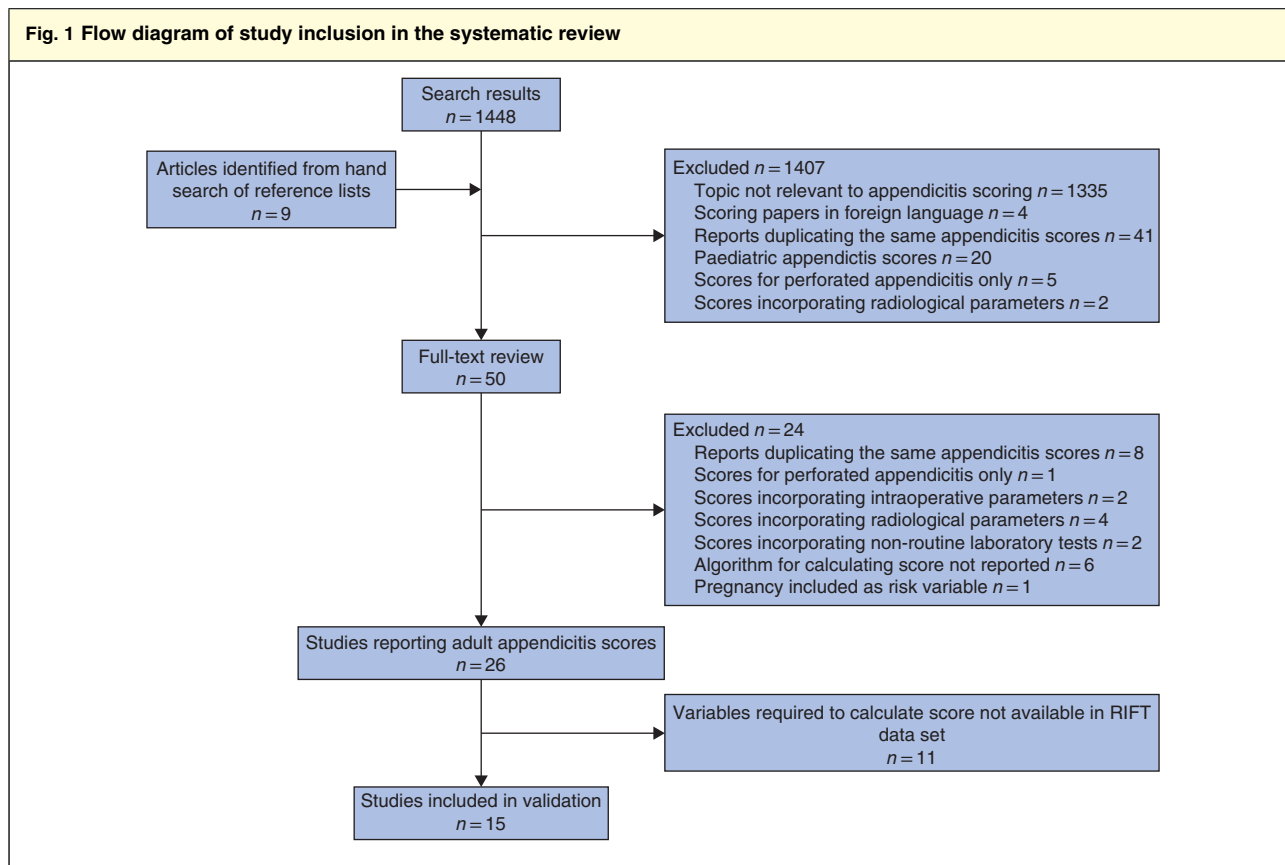

Fig. 1 Flow diagram of study inclusion in the systematic review

Table 1 Characteristics of risk prediction models identified in systematic review				
Model	Year	Country	Derivation cohort	Patients included in derivation cohort
Van Way et al.[34]	1982	USA	Retrospective single-centre	476 patients who underwent appendicectomy; included children and adults, and both sexes*
Alvarado[20]	1986	USA	Retrospective single-centre	305 patients presenting with abdominal pain; age range 4–80 years, 42% of patients were women
Izbicki et al.[28]	1992	Germany	Retrospective single-centre	536 patients who underwent appendicectomy; included both sexes*
Eskelinen et al.[25]	1992	Finland	Prospective multicentre	1333 patients presenting with abdominal pain; mean age 38 years, 52 per cent of patients were women
Christian and Christian[24]	1992	India	Prospective single-centre	58 patients presenting with abdominal pain; age range 10–56 years, 22% of patients were women
Modified Alvarado[29]	1994	UK	n.a.†	n.a.†
Eskelinen et al.[26]	1994	Finland	Prospective multicentre	636 patients presenting with abdominal pain; included men only*
van der Broek et al.[33]	2002	Holland	Prospective single-centre	577 patients presenting with abdominal pain; all patients aged above 10 years, 59% of patients were women
Birkhahn et al.[22]	2006	USA	Prospective single-centre	439 patients presenting with abdominal pain; age range 3–93 years, 65% of patients were women
AIRS[21]	2008	Sweden	Prospective multicentre	316 patients presenting with abdominal pain; mean age 26 years, 54% of patients were women
RIPASA score[23]	2010	Brunei	Retrospective single-centre	312 patients who underwent appendicectomy; mean age 26 years, 42% of patients were women
Ting et al.[32]	2010	Taiwan	Retrospective single-centre	532 patients who underwent appendicectomy; 39% of patients were women*
AAS[31]	2014	Finland	Prospective single-centre	725 patients presenting with abdominal pain; age range 16–97 years, 58% of patients were women
Goh[27]	2017	China	n.a.†	n.a.†
Mikaere et al.[30]	2018	New Zealand	Retrospective single-centre	885 patients presenting with abdominal pain; all patients aged over 15 years, 59% of patients were women

*Overall age and/or sex information not reported. †Not applicable (n.a.) as modification of a previously published score (no derivation cohort). AIRS, Appendicitis Inflammatory Response Score; RIPASA, Raja Isteri Pengiran Anak Saleha Appendicitis; AAS, Adult Appendicitis Score.

of 59) in the Republic of Ireland and 10·1 per cent (18 of 179) in Spain. In men in the same age group, the NAR was 12·1 per cent (120 of 993) in the UK, 0·6 per cent (1 of 157) in Italy, 6 per cent (2 of 32) in Portugal, 3·4 per cent (7 of 208) in the Republic of Ireland and 3·3 per cent (2 of 61) in Spain. As anticipated, the NAR was higher in the UK than in other countries, so risk score validation proceeded in UK patients. All of the following data are for UK patients only.
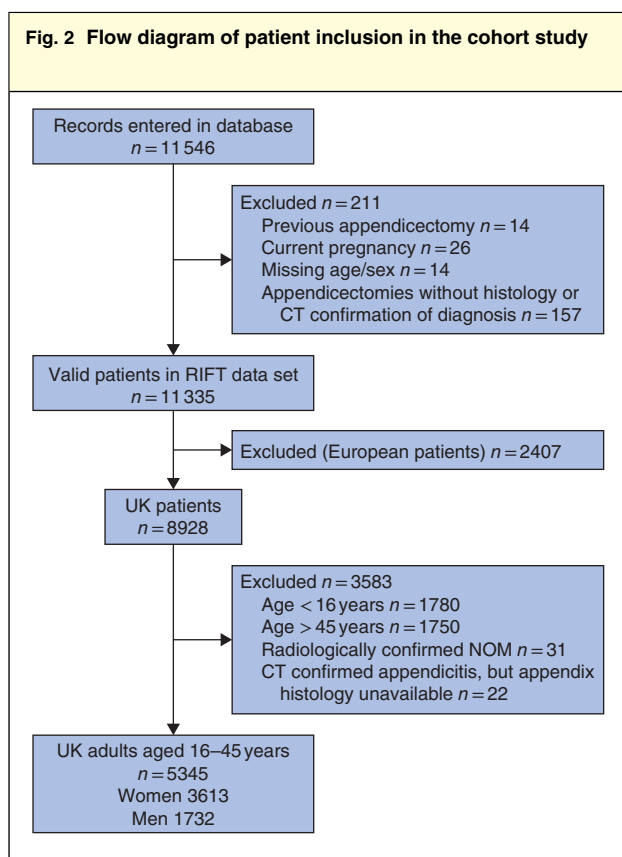
Data integrity

A total of 5345 patients were included in the study (*Fig. 2*), from across 154 UK hospitals. Of the 17 clinical parameters required to validate the risk prediction models, 0·9 per cent (794 of 90 865) were missing. Some 35·3 per cent (4461 of 12 647) of eligible data fields were assessed by independent data validators, finding overall data accuracy to be 98·3 per cent (4384 of 4461).

Patient characteristics and outcomes

Two-thirds (3613 of 5345, 67·6 per cent) of patients were women (*Table 2*). Women were less likely to undergo any surgery than men (32·0 *versus* 59·8 per cent respectively; relative risk 0·53, 95 per cent c.i. 0·50 to 0·57, *P* < 0·001). Amongst women undergoing appendicectomy, 97·6 per

Fig. 2 Flow diagram of patient inclusion in the cohort study

Records entered in database
$n = 11\,546$

Excluded $n = 211$
Previous appendicectomy $n = 14$
Current pregnancy $n = 26$
Missing age/sex $n = 14$
Appendicectomies without histology or
CT confirmation of diagnosis $n = 157$

Valid patients in RIFT data set
$n = 11\,335$

Excluded (European patients) $n = 2407$

UK patients
$n = 8928$

Excluded $n = 3583$
Age < 16 years $n = 1780$
Age > 45 years $n = 1750$
Radiologically confirmed NOM $n = 31$
CT confirmed appendicitis, but appendix
histology unavailable $n = 22$

UK adults aged 16–45 years
$n = 5345$
Women 3613
Men 1732

NOM, non-operative management of appendicitis.

Table 2 Patient management stratified by sex

	Women (*n* = 3613)	Men (*n* = 1732)
No. who had surgery	1156 (32·0)	1036 (59·8)
No. of appendicectomies performed	964 of 1156 (83·4)	993 of 1036 (95·8)
Confirmed appendicitis	625 (64·8)	841 (84·7)
Other appendix pathology	67 (7·0)	32 (3·2)
Histologically normal appendix	272 (28·2)	120 (12·1)
No. not operated on	2457 (68·0)	696 (40·2)

Values in parentheses are percentages.

cent (941 of 964) had a laparoscopic procedure, of which 2·2 per cent (21 of 941) were converted to an open operation. Amongst men undergoing appendicectomy, 93·9 per cent (932 of 993) had a laparoscopic procedure, of which 6·0 per cent (56 of 932) were converted to an open procedure.

The NAR was higher in women than in men (28·2 *versus* 12·1 per cent respectively; relative risk 2·33, 95 per cent c.i.

1·92 to 2·84, $P < 0.001$). Within the 30-day follow-up, 27·4 per cent (1466 of 5345) of all patients presenting with acute RIF pain had undergone appendicectomy and been confirmed to have appendicitis on histological examination. Of all patients presenting with RIF pain, women were less likely to have a confirmed diagnosis of appendicitis than men (17·3 *versus* 48·6 per cent respectively; relative risk 0·36, 0·33 to 0·39, $P < 0.001$).

Clinical use of risk prediction models

Only 0·6 per cent of patients (32 of 5345) were recorded as having been formally risk-scored on admission by their clinical team. When performed, the Alvarado score was used most frequently (29 of 32, 91 per cent).
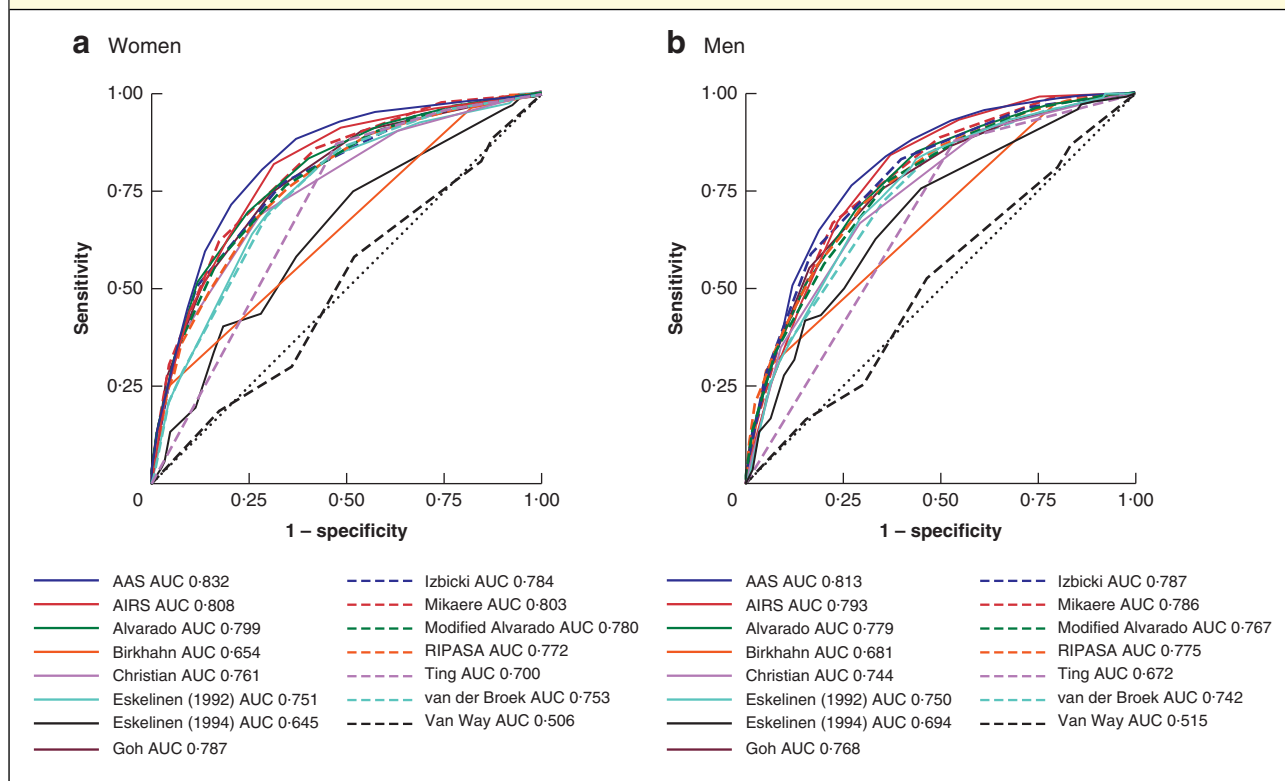
Validation of risk prediction models

Of the 15 risk prediction models, 11 showed consistently good discrimination for identifying appendicitis (AUC above 0·7) across both women and men (*Fig. 3* and *Table 3*). In women, the Adult Appendicitis Score[31] (AAS) achieved the highest specificity whilst maintaining a failure rate of less than 5 per cent in low-risk patients (*Table 3*). At a cut-off score of 8 or less, the AAS triaged 63·1 per cent of women who did not have appendicitis to the low-risk group (specificity), and amongst all women in the low-risk group 3·7 per cent in fact had appendicitis (failure rate). In men, the optimal model was the Appendicitis Inflammatory Response Score[21] (AIRS), with a cut-off score of 2 or less associated with a specificity of 24·7 per cent and a failure rate of 2·4 per cent.

Sensitivity analyses

Overall, 0·9 per cent (294 of 32 517) of the variables required to calculate AAS were missing in women. The main complete-case analysis was based on the 95·0 per cent (3433 of 3613) of women for whom all variables were available. In sensitivity analysis, the failure rate was found to range from 3·6 to 3·8 per cent, if missing variables were scored with either the maximum or minimum possible point value respectively. Specificity was found to range from 61·3 to 64·2 per cent respectively.

In total, 1·0 per cent (156 of 15 588) of the variables required to calculate AIRS were missing in men. The complete-case analysis was based on the 93·9 per cent (1627 of 1732) of men for whom all variables were available. In sensitivity analysis, the failure rate ranged between 2·4 and 2·9 per cent, and specificity from 23·3 to 26·4 per cent.

Fig. 3 Receiver operating characteristic (ROC) curves for 15 appendicitis risk prediction models in women and men

a Women; **b** men. These receiver operating characteristic (ROC) curves are based on a complete-case analysis for all 15 risk prediction models. As a result they may be based on fewer patients than when each model was validated individually (*Table 3*). This accounts for any minor discrepancies between the data presented here and those in *Table 3*. AAS, Adult Appendicitis Score; AUC, area under the ROC curve; AIRS, Appendicitis Inflammatory Response Score; RIPASA, Raja Isteri Pengiran Anak Saleha Appendicitis.

Patients stratified to low-risk groups

Of the 1856 women identified as low risk by the AAS in the complete-case analysis, 1560 (84·1 per cent) did not undergo surgery (*Table 4*). Of women who either did not undergo surgery or had a procedure other than appendicectomy, at 30 days the final diagnoses were non-specific abdominal pain (851 of 1627, 52·3 per cent), benign gynaecological pathology (430 of 1627, 26·4 per cent), urinary tract infection (99 of 1627, 6·1 per cent) and miscellaneous other (247 of 1627, 15·2 per cent). A full breakdown of diagnoses is provided in *Table S3* (supporting information).

Amongst the 12·3 per cent (229 of 1856) of women who underwent appendicectomy, 57·6 per cent (132 of 229) had a normal appendicectomy, 26·2 per cent (60 of 229) had simple appendicitis, 3·9 per cent (9 of 229) had complex appendicitis and 12·2 per cent (28 of 229) had other abnormal appendix pathology. Amongst the 28 women with other pathology, four were found to have carcinoid and one to have Crohn's disease. The readmission

rate for ongoing RIF pain amongst low-risk women who were not operated on in their index admission was 8·2 per cent (130 of 1586). Of the women who had undergone appendicectomy on index admission, 6·8 per cent (14 of 207) were readmitted with postoperative complications.

In the complete-case analysis, AIRS identified 209 men as being at low risk (*Table 4*). Of these, only 34 (16·3 per cent) had an appendicectomy and one man (0·5 per cent) underwent non-appendix surgery. At 30 days, the final diagnosis for most men who did not undergo appendicectomy was non-specific pain (110 of 175, 62·9 per cent) (*Table S4*, supporting information). Of men who had an appendicectomy, five were found to have appendicitis on histological examination, with no complex appendicitis identified. The NAR in low-risk men was 74 per cent (25 of 34). The readmission rate for ongoing RIF pain in low-risk men who were not operated on in their index admission was 7·3 per cent (13 of 178), and the readmission rate for postoperative complications among those who had an appendicectomy was 23 per cent (7 of 31).

Table 3 Validation and identification of optimal thresholds for risk prediction models, stratified by sex				
Model	AUC	Optimal threshold	Failure rate (%)	Specificity (%)
AAS[31]				
Women	0·83 (0·82, 0·85)	≤ 8	3·7	63·1
Men	0·81 (0·79, 0·83)	≤ 6	5·0	20·4
AIRS[21]				
Women	0·81 (0·79, 0·83)	≤ 3	3·5	51·6
Men	0·79 (0·77, 0·82)	≤ 2	2·4	24·7
Alvarado[20]				
Women	0·80 (0·78, 0·82)	≤ 3	3·7	40·8
Men	0·78 (0·76, 0·80)	≤ 1	0	6·2
Birkhahn et al.[22]				
Women	0·66 (0·64, 0·67)	1	3·8	18·1
Men	0·68 (0·66, 0·70)	1	10·1	21·6
Christian and Christian[24]				
Women	0·76 (0·74, 0·78)	0	0·9	3·9
Men	0·75 (0·72, 0·77)	0	2·1	5·5
Eskelinen et al.[25]				
Women	0·76 (0·74, 0·78)	≤ 45·3	4·4	9·6
Men	0·75 (0·73, 0·78)	≤ 39·4	1·4	8·4
Eskelinen et al.[26]				
Women	0·65 (0·62, 0·67)	≤ − 5·8	3·6	6·4
Men	0·70 (0·67, 0·72)	≤ − 7·7	10·5	2·0
Goh[27]				
Women	0·79 (0·77, 0·81)	≤ 2	4·1	46·3
Men	0·77 (0·75, 0·79)	0	5·3	8·4
Izbicki et al.[28]				
Women	0·79 (0·77, 0·80)	≤ 1	3·3	21·2
Men	0·79 (0·77, 0·81)	≤ 1	6·1	3·6
Mikaere et al.[30]				
Women	0·80 (0·78, 0·82)	≤ 1	4·9	57·9
Men	0·79 (0·76, 0·81)	≤ 1	4·3	21·7
Modified Alvarado[29]				
Women	0·78 (0·76, 0·80)	≤ 3	4·5	43·6
Men	0·77 (0·75, 0·79)	≤ 2	11·8	25·4
RIPASA score[23]				
Women	0·77 (0·75, 0·79)	≤ 5·5	4·7	44·2
Men	0·78 (0·76, 0·80)	≤ 4	0	3·8
Ting et al.[32]				
Women	0·70 (0·68, 0·72)	0	19·8	46·9
Men	0·67 (0·65, 0·69)	0	4·8	52·5
van der Broek et al.[33]				
Women	0·76 (0·74, 0·77)	0	3·3	24·6
Men	0·75 (0·72, 0·77)	≤ 2	14·3	23·1
Van Way et al.[34]				
Women	0·51 (0·48, 0·53)	32	15·5	12·9
Men	0·52 (0·49, 0·54)	32	42·9	16·4

Values in parentheses are 95 per cent confidence intervals. AUC, area under the curve; AAS, Adult Appendicitis Score; AIRS, Appendicitis Inflammatory Response Score; RIPASA, Raja Isteri Pengiran Anak Saleha Appendicitis.

Table 4 Management and readmissions in patients scored as low risk, stratified by sex

	Women with AAS ≤ 8 (*n* = 1856)	Men with AIRS ≤ 2 (*n* = 209)
Patient management		
No. who had surgery	296 (15·9)	35 (16·7)
No. of appendicectomies performed	229 of 296 (77·4)	34 of 35 (97)
Simple appendicitis	60 (26·2)	5 (15)
Complex appendicitis	9 (3·9)	0 (0)
Other appendix pathology	28 (12·2)	4 (12)
Histologically normal appendix	132 (57·6)	25 (74)
No. not operated on	1569 (84·1)	174 (83·3)
Readmission		
Ongoing RIF pain in patients not operated on in index admission	130 of 1586 (8·2)	13 of 178 (7·3)
Postoperative complications following appendicectomy on index admission	14 of 207 (6·8)	7 of 31 (23)

Values in parentheses are percentages. AAS, Adult Appendicitis Score; AIRS, Appendicitis Inflammatory Response Score; RIF, right iliac fossa.

Table 5 Use of imaging and performance, stratified by sex and risk category

	Women		Men	
	Low risk (AAS ≤ 8) (*n* = 1856)	High risk (AAS > 8) (*n* = 1577)	Low risk (AIRS ≤ 2) (*n* = 209)	High risk (AIRS > 2) (*n* = 1418)
Ultrasound imaging	1291 (69·6)	916 (58·1)	59 (28·2)	205 (14·5)
AUC*	0·63 (0·57, 0·70)	0·68 (0·65, 0·71)	1·00 (n.a.)	0·66 (0·60, 0·72)
Sensitivity	0·28	0·38	1·00	0·37
Specificity	0·99	0·98	1·00	0·95
NPV	0·97	0·82	1·00	0·74
PPV	0·54	0·84	1·00	0·79
CT	208 (11·2)	316 (20·0)	41 (19·6)	341 (24·0)
AUC*	0·99 (0·98, 1·00)	0·93 (0·90, 0·96)	1·00 (n.a.)	0·92 (0·90, 0·95)
Sensitivity	1·00	0·92	1·00	0·94
Specificity	0·97	0·95	1·00	0·91
NPV	1·00	0·96	1·00	0·94
PPV	0·62	0·91	1·00	0·90
No imaging	434 (23·4)	456 (28·9)	111 (53·1)	910 (64·2)

Values in parentheses are percentages unless indicate otherwise; *values in parentheses are 95 per cent confidence intervals. AAS, Adult Appendicitis Score; AIRS, Appendicitis Inflammatory Response Score; AUC, area under the receiver operating characteristic (ROC) curve; n.a., not applicable; NPV, negative predictive value; PPV, positive predictive value.
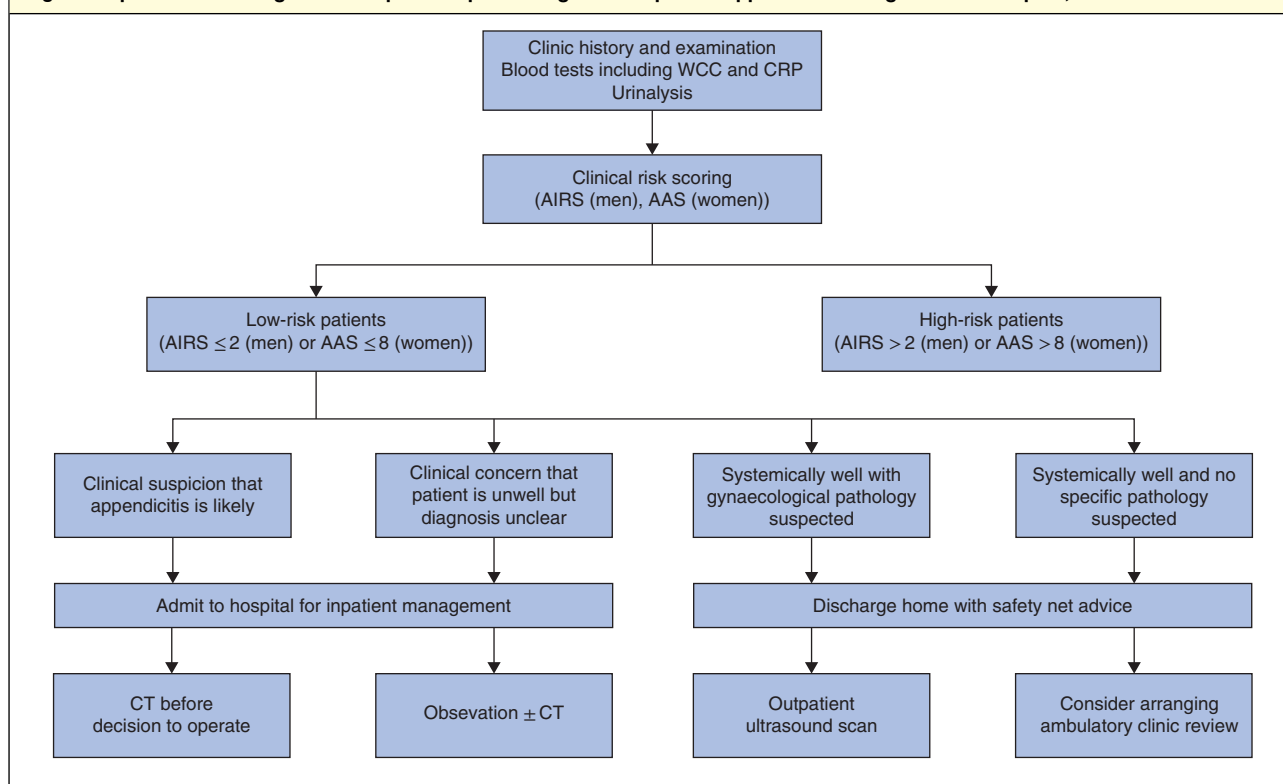
Imaging

Most women (2638 of 3613, 73·0 per cent) had either ultrasound or CT preoperative imaging. The imaging modality used most frequently was ultrasonography (2289 of 3613, 63·4 per cent). Low overall sensitivity (0·36) indicated that triage based solely on ultrasound imaging would misclassify most patients with appendicitis as being low risk. The failure rate for ultrasonography was 8·4 per cent. Stratifying women by AAS score, the AUC for ultrasound imaging was modest in both the low-risk (AUC 0·63) and high-risk (AUC 0·68) groups (*Table 5*). Although CT was performed in only 15·1 per cent (547 of 3613) of women overall, this

imaging modality was sensitive (0·92), with a low failure rate (2·1 per cent). Stratifying by AAS score, the AUC for CT was excellent in both the low-risk (AUC 0·99) and high-risk (AUC 0·93) groups.

Overall, only 36·2 per cent (627 of 1732) of men underwent preoperative imaging. Ultrasound imaging (276 of 1732, 15·9 per cent) was performed less frequently than CT (398 of 1732, 23·0 per cent). The overall sensitivity of ultrasonography (0·38) was low, with a high failure rate (18·8 per cent). In men stratified as high risk by AIRS, ultrasound imaging had a poor AUC (0·66). In contrast, the overall sensitivity of CT (0·94) was high, with a failure rate of 4·5 per cent. In high-risk men, CT had an excellent

Fig. 4 **Proposed clinical algorithm for patients presenting with suspected appendicitis or right iliac fossa pain, stratified as low risk**

WCC, white cell count; CRP, C-reactive protein; AIRS, Appendicitis Inflammatory Response Score; AAS, Adult Appendicitis Score.

AUC (0·92). Analysis of the performance of ultrasound and CT imaging in low-risk men was limited by low numbers (*Table 5*).

AAS and AIRS performance across participating countries

Overall, 17·3 per cent (625 of 3613) of women and 48·6 per cent (841 of 1732) of men in the UK presenting with appendicitis had a histological diagnosis of appendicitis, compared with 42·5 per cent (361 of 849) of women and 68·5 per cent (442 of 645) of men in Ireland, Italy, Portugal and Spain.

To test the potential for international application of risk-scoring, performance of the AAS and AIRS was tested in the Irish, Italian, Portuguese and Spanish data. In women, the AAS (cut-off score 8 or below) had a specificity of 57·5 per cent (226 of 393), with a failure rate of 17·5 per cent (48 of 274). In men, the AIRS (cut-off score 2 or less) had a specificity of 15·6 per cent (25 of 160), with a failure rate of 32 per cent (12 of 37).

Discussion

This study found that women in the UK who presented with acute RIF pain had a disproportionately high rate of admission without surgical intervention. Women who did undergo surgery had high rates of normal appendicectomy. Using the AAS[31] (cut-off score 8 or less) it was possible to stratify almost two-thirds of UK women aged 16–45 years who presented with RIF pain into a low-risk group. This low-risk group had an overall one in 27 risk of appendicitis and one in 200 risk of complex appendicitis. Similarly, the AIRS[21] (cut-off score 2 or less) identified a smaller group of UK men who were at low risk of appendicitis. The performance of ultrasound imaging for diagnosis of appendicitis was poor in both men and women, whereas CT was both sensitive and specific across all subgroups.

The overall NAR in UK adults aged 16–45 years was 20·0 per cent, significantly higher than the rate recorded in other countries that participated in the RIFT Study. This represents one of the world's highest NARs[9,18,19]. Although simple and complex appendicitis may represent distinct pathologies[46], some surgeons believe that

delaying surgery may increase the risk of appendiceal perforation. This leads to some surgeons having a low threshold for surgery, preferring for patients with equivocal presentations to undergo early appendicectomy rather than a period of clinical observation. This may result in potentially unnecessary operations (removal of histologically normal appendices) with associated postoperative morbidity[3−5]. However, leaving a macroscopically normal looking appendix *in situ* may risk missing microscopic inflammation[47] and is associated with an increased readmission rate[48]. These conflicting considerations have resulted in variations in practice, with some surgeons routinely leaving a macroscopically normal appendix *in situ*[49]. Improved preoperative diagnosis could potentially reduce both overtreatment and heterogeneity in practice.

A large number of risk prediction models for acute appendicitis have been published. Few have been validated robustly, with most validation studies relying on small single-centre retrospective data sets[12]. In the present study of 5345 patients across 154 UK hospitals, most models were unable safely to identify significant numbers of patients at low risk of appendicitis. Given the clinical importance of identifying low-risk patients, this study was preplanned to focus on validating models' prediction of patients who do not have appendicitis (true negatives) rather than prediction of acute appendicitis (true positives), whereas most previous studies have prioritized identification of high-risk patients. In addition to differences in baseline case mix, this explains why the optimal cut-off scores identified in this study differ from those proposed in the original AAS (original study proposed cut-off score of 11 or less *versus* 8 or below identified in the present study) and AIRS (original study proposed cut-off score of 5 or less *versus* 2 or less identified in this study) studies[21,31]. Identification of the optimal cut-off scores for use in the UK population will increase the likelihood of risk prediction models being disseminated widely and implemented safely in the UK National Health Service. Routine risk scoring has been found in prospective studies to be associated with reduced need for imaging and hospital admission, and to reduce the NAR[14,50].

Ultrasound imaging is used frequently in women as it allows effective visualization of gynaecological organs. However, in this national UK cohort it was found to perform poorly in the identification of appendicitis in both women and men. Although CT was performed highly selectively, consistent with previous studies[7,8] it demonstrated excellent discrimination for appendicitis. Routine CT may decrease NARs, but exposes patients to radiation[51]. In the past it has been estimated that there may be one excess cancer for every 12 normal appendicectomies

avoided by routine CT[6], but these concerns are less prominent in the era of low-dose CT protocols. Nonetheless, low-risk patients, particularly women of childbearing age, may choose to avoid ionizing radiation, if there is a low index of suspicion of appendicitis. This study excluded older, postmenopausal women, who are more likely to benefit from routine CT to exclude colonic pathology such as malignancy and diverticulitis[10].

This is the largest prospective multicentre cohort study worldwide of RIF pain in the era of laparoscopic surgery. A total of 154 hospitals contributed data, representing around two-thirds of UK hospitals that provide general surgery[52]. The study's findings are therefore broadly generalizable across the UK. The CRF was designed to be completed at the patient's bedside during their initial assessment. This minimized measurement and recall bias, leading to high data completeness (99·1 per cent) and accuracy (98·3 per cent) rates, ensuring high internal validity.

Previous trainee-led prospective cohort studies achieved high levels of case ascertainment[53], but it is possible that a small number of eligible patients were missed during the study inclusion windows. Follow-up was limited to the index hospital where patients initially presented, so some patients discharged without having undergone appendicectomy may have been readmitted and operated on at another hospital, although this is likely to be infrequent. There is weak epidemiological evidence indicating that there may be seasonal variation in the incidence of appendicitis[54]. Year-long data collection would maximize the study's generalizability, but high-quality data collection within a multicentre collaborative study would not be sustainable for protracted periods.

An additional 15 data items (*Table S2*, supporting information) were required to validate all existing risk prediction models, but collecting these items for each patient would have placed an impractical burden on participating centres. As a consequence, 11 existing models could not be validated. None of these is in clinical use, but there is a hypothetical possibility that one or more may have outperformed AAS and AIRS if tested.

Risk stratification can be performed by the first clinician in contact with the patient, who has blood test results available. However, as this study captured patients at the point of assessment by surgical teams, its findings do not directly support the implementation of risk scoring by general practitioners and emergency physicians. As the incidence of appendicitis in patients presenting to general practice or the emergency department with abdominal pain is lower than that in patients reviewed by surgical teams, it is likely that risk prediction models would perform better in these settings, but further evaluation is required.

The UK has one of the world's highest NARs[14,18,19], so the predefined aim for this study was to evaluate the potential for routine risk scoring to identify low-risk UK patients who are unlikely to have appendicitis. It was predicted that the NAR would be lower in the other participating countries and there would be no need for appendicitis risk scoring. As anticipated, the overall NAR in Italy, Portugal, Republic of Ireland and Spain was lower than that in the UK (10·2 *versus* 28·2 per cent respectively in women, and 2·6 *versus* 12·1 per cent in men). An unexpected finding, however, was that in those countries a greater proportion of all patients who were admitted with RIF pain had a final diagnosis of appendicitis than in the UK (42·5 *versus* 17·3 per cent respectively in women, and 68·5 *versus* 48·6 per cent in men). The differences in the prevalence of appendicitis between the UK and other settings may explain why the failure rates (reciprocal of negative predictive value) were unacceptably high in the Irish, Italian, Portuguese and Spanish patients. Therefore, this study's results should be extrapolated cautiously to settings outside the UK. It is possible that in other countries with high baseline NAR, such as Australia[55], and lower appendicitis prevalence amongst patients admitted with RIF pain there may be a role for clinical risk scoring, but local validation studies are needed.

Risk prediction models, stratified by sex, may act as adjuncts to high-quality serial clinical assessment of patients, rationalizing exposure to ionizing radiation to those patients most likely to benefit from CT. AAS and AIRS can be implemented easily, as they require only simple clinical information and routine blood tests, which were already performed for most patients in the present observational cohort.

The authors propose that all adults presenting with acute RIF pain or suspected appendicitis should be scored routinely using the appropriate risk prediction model. To support calculation and application of appropriate cut-off scores at the patient's bedside, a mobile-, tablet- and desktop-compatible web application has been developed (http://appy-risk.org).

To mitigate against the high risk of normal appendicectomy in low-risk patients, if a patient is stratified as low risk and suspicion of appendicitis remains, low radiation dose CT should be undertaken to confirm the diagnosis before a decision to operate. Ultrasound imaging is preferable in women if the principal differential diagnosis is gynaecological pathology. A very small proportion of low-risk women and men (*Tables S3* and *S4*, supporting information) presenting with RIF pain have serious conditions such as colitis. Therefore, patients with unclear diagnosis or markers of significant illness (such as fever or significantly raised

C-reactive protein level) should be admitted for observation and potentially for inpatient radiological investigation. Patients who are clinically well with a low index of suspicion of pathology requiring inpatient treatment may choose to be managed at home with the safety net of prompt ambulatory reassessment. Ambulatory management reduces rates of inpatient admission and CT[56], and may in turn reduce unnecessary surgery and NARs. A clinical algorithm proposed by the authors is shown in *Fig. 4*.

When diagnostic failure does occur in patients stratified to low-risk groups, the risk of complex appendicitis is very low. Previous studies have suggested that a short delay to appendicectomy does not increase the risk of perforation[57]. It is likely that many patients in these studies were administered antibiotics while they awaited surgery, and it is not known whether patients with appendicitis who initially receive a period of ambulatory management are at increased risk of perforation, so this should be investigated in future studies.

Collaborators

Writing Group: D. Nepogodiev, J. H. Matthews, G. L. Morley, D. N. Naumann, A. Ball, P. Chauhan, S. Bhanderi, I. Mohamed, J. C. Glasbey, R. J. W. Wilkin, T. M. Drake, J. Clements, N. S. Blencowe, P. J. J. Herrod, F. Pata, M. Frasson, R. Blanco-Colino, A. S. Soares, A. Bhangu.

Acknowledgements

References

1 GlobalSurg Collaborative. Mortality of emergency abdominal surgery in high-, middle- and low-income countries. *Br J Surg* 2016; **103**: 971–988.
2 Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet* 2015; **386**: 1278–1287.
3 Bijnen CL, Van Den Broek WT, Bijnen AB, De Ruiter P, Gouma DJ. Implications of removing a normal appendix. *Dig Surg* 2003; **20**: 115–121.
4 Lee M, Paavana T, Mazari F, Wilson TR. The morbidity of negative appendicectomy. *Ann R Coll Surg Engl* 2014; **96**: 517–520.
5 Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and

socioeconomic implications of negative appendectomy. *Am J Surg* 2016; **212**: 1076–1082.

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

7 Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP *et al.* Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med* 2012; **366**: 1596–1605.

8 Yoon HM, Suh CH, Cho YA, Kim JR, Lee JS, Jung AY *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**: 2537–2548.

9 van Rossem CC, Bolmers MD, Schreinemacher MH, Bemelman WA, van Geloven AA, Pinkney TD *et al.* Diagnosing acute appendicitis: surgery or imaging? *Colorectal Dis* 2016; **18**: 1129–1132.

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

11 Gorter RR, Eker HH, Gorter-Stam MA, Abis GS, Acharya A, Ankersmit M *et al.* Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. *Surg Endosc* 2016; **30**: 4668–4690.

12 Kularatna M, Lauti M, Haran C, MacFater W, Sheikh L, Huang Y *et al.* Clinical prediction rules for appendicitis in adults: which is best? *World J Surg* 2017; **41**: 1769–1781.

13 Royal College of Surgeons of England. *Commissioning Guide: Emergency General Surgery (Acute Abdominal Pain)*; 2014. https://www.rcseng.ac.uk/−/media/files/rcs/standards-and-research/commissioning/commissioning-guide--egs-published-v3.pdf [accessed 5 February 2018].

14 Andersson M, Kolodziej B, Andersson RE; STRAPPSCORE Study Group. Randomized clinical trial of appendicitis inflammatory response score-based management of patients with suspected appendicitis. *Br J Surg* 2017; **104**: 1451–1461.

15 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**: e017574.

16 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

17 Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016; **6**: e012799.

18 National Surgical Research Collaborative. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. *Br J Surg* 2013; **100**: 1240–1252.

19 Sartelli M, Baiocchi GL, Di Saverio S, Ferrara F, Labricciosa FM, Ansaloni L *et al.* Prospective Observational Study on acute Appendicitis Worldwide (POSAW). *World J Emerg Surg* 2018; **13**: 19.

20 Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; **15**: 557–564.

21 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**: 1843–1849.

22 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**: 497–502.

23 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 Med J* 2010; **51**: 220–225.

24 Christian F, Christian GP. A simple scoring system to reduce the negative appendicectomy rate. *Ann R Coll Surg Engl* 1992; **74**: 281–285.

25 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. *Theor Surg* 1992; **7**: 86–90.

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

27 Goh PL. A simplified appendicitis score in the diagnosis of acute appendicitis. *Hong Kong J Emerg Med* 2017; **17**: 230–235.

28 Izbicki JR, Knoefel WT, Wilker DK, Mandelkow HK, Müller K, Siebeck M *et al.* Accurate diagnosis of acute appendicitis: a retrospective and prospective analysis of 686 patients. *Eur J Surg* 1992; **158**: 227–231.

29 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**: 418–419.

30 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**: E303–E307.

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

32 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**: 401–406.

33 van den Broek WT, Bijnen BB, Rijbroek B, Gouma DJ. Scoring and diagnostic laparoscopy for suspected appendicitis. *Eur J Surg* 2002; **168**: 349–354.

34 Van Way CW III, Murphy JR, Dunn EL, Elerding SC. A feasibility study of computer aided diagnosis in appendicitis. *Surg Gynecol Obstet* 1982; **155**: 685–688.

35 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**: e0164574.

36 Fenyö G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. *Acta Chir Scand* 1987; **153**: 545–551.

37 Jahn H, Mathiesen FK, Neckelmann K, Hovendal CP, Bellstrøm T, Gottrup F. Comparison of clinical judgment and diagnostic ultrasonography in the diagnosis of acute appendicitis: experience with a score-aided diagnosis. *Eur J Surg* 1997; **163**: 433–443.

38 Jawaid A, Asad A, Motiei A, Munir A, Bhutto E, Choudry H *et al.* Clinical scoring system: a valuable tool for decision making in cases of acute appendicitis. *J Pak Med Assoc* 1999; **49**: 254–259.

39 Lindberg G, Fenyo G. Algorithmic diagnosis of appendicitis using Bayes' theorem and logistic regression. *Bayesian Stat* 1988; **3**: 665–668.

40 Ramirez JM, Deus J. Practical score to aid decision making in doubtful cases of appendicitis. *Br J Surg* 1994; **81**: 680–683.

41 Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. *Ann Surg* 1983; **198**: 753–759.

42 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.

43 Arnbjörnsson E. Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective *versus* retrospective studies. *Ann Chir Gynaecol* 1985; **74**: 159–166.

44 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. *Niger J Med* 2009; **18**: 398–401.

45 Ohmann C, Franke C, Yang Q. Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. German Study Group of Acute Abdominal Pain. *Arch Surg* 1999; **134**: 993–996.

46 Livingston EH, Woodward WA, Sarosi GA, Haley RW. Disconnect between incidence of nonperforated and perforated appendicitis: implications for pathophysiology and management. *Ann Surg* 2007; **245**: 886–892.

47 Strong S, Blencowe N, Bhangu A; National Surgical Research Collaborative. How good are surgeons at identifying appendicitis? Results from a multi-centre cohort study. *Int J Surg* 2015; **15**: 107–112.

48 Bhangu A, Begaj I, Ray D. Population level analysis of diagnostic laparoscopy *versus* normal appendicectomy for acute lower abdominal pain. *Int J Surg* 2014; **12**: 1374–1379.

49 Van Vlodrop V, Pattyn P, Ceelen W. Management of the 'normal' appendix during surgery for right fossa syndrome: proposed clinical algorithm. *Acta Chir Belg* 2007; **107**: 8–11.

50 Sammalkorpi HE, Mentula P, Savolainen H, Leppäniemi A. The introduction of adult appendicitis score reduced negative appendectomy rate. *Scand J Surg* 2017; **106**: 196–201.

51 Lahaye MJ, Lambregts DM, Mutsaers E, Essers BA, Breukink S, Cappendijk VC *et al.* Mandatory imaging cuts costs and reduces the rate of unnecessary surgeries in the diagnostic work-up of patients suspected of having appendicitis. *Eur Radiol* 2015; **25**: 1464–1470.

52 Nepogodiev D, Chapman SJ, Kolias AG, Fitzgerald JE, Lee M, Blencowe NS; National Surgical Research Collaborative. The effect of trainee research collaboratives in the UK. *Lancet Gastroenterol Hepatol* 2017; **2**: 247–248.

53 CholeS Study Group, West Midlands Research Collaborative. Population-based cohort study of outcomes following cholecystectomy for benign gallbladder diseases. *Br J Surg* 2016; **103**: 1704–1715.

54 Fares A. Summer appendicitis. *Ann Med Health Sci Res* 2014; **4**: 18–21.

55 Arthur T, Gartrell R, Manoharan B, Parker D; QUEST Collaboration. Emergency appendicectomy in Australia: findings from a multicentre, prospective study. *ANZ J Surg* 2017; **87**: 656–660.

56 Pidgeon TE, Shariff U, Devine F, Menon V. A report on an acute, in-hours, outpatient review clinic with ultrasonography facilities for the early evaluation of general surgical patients. *Ann R Coll Surg Engl* 2016; **98**: 468–474.

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

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.