Evaluation of clinical examination and preoperative imaging in patients with right iliac fossa pain and a medium or high risk score for appendicitis

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

*Collaborating authors are listed in Appendix 1

Short title: Diagnosis of appendicitis in adults

Category: Original research

Keywords: Appendicitis; appendicectomy; general surgery; emergency surgery; risk prediction models.

Competing interests statement: The authors have declared that no competing interests exist.

Funding: The authors received no specific funding for this work.

Acknowledgements: None

Abstract

Introduction: Clinical risk models can be used to safely stratify UK patients with suspected appendicitis with low scores to ambulatory pathways, preventing admissions and unnecessary surgery. This study aimed to examine the diagnostic pathways in patients with medium or high risk scores.

Methods: This prospective cohort study included patients admitted with suspected appendicitis in the UK. Patients aged ≥16 years were included if they had a medium or high risk score for appendicitis (Appendicitis Inflammatory Response score >2 in men, Adult Appendicitis Score >8 in women). The primary outcome measure was the normal appendicectomy rate (NAR). The AIRS and AAS risk prediction models were validated against pre-determined criteria.

Results: This UK study included 2,231 women and 1,958 men. Overall, 57.7% of patients underwent surgery. The NAR was 18.4% in women aged 16-45 years, 9.5% in men aged 16-45 years, 4.5% in women aged ≥46 years, and 2.1% in men aged ≥46 years. Risk prediction models did not achieve the pre-determined threshold to be used to identify patients with appendicitis. Ultrasound was the most common imaging modality in women aged 16-45 years, whereas CT was most common in the other subgroups. CT was performed in 12.8% women aged 16-45 years, 21.1% in men aged 16-45 years, 69.3% in women aged ≥46 years, and 74.5% in men aged ≥46 years. The overall NAR in patients who had CT imaging alone (3.6%) was lower than in patients who had no imaging (12.4%) or ultrasound imaging alone (19.0%).

Conclusion: Patients with right iliac fossa pain should be risk scored, with low-risk patients triaged to ambulatory management and medium and high-risk patients routinely CT scanned. Normal appendicectomy should become a specific therapeutic option rather than a chance finding.

Introduction

Acute appendicitis remains the most common general surgery emergency in the world and surgery remains the mainstay of treatment[1]. Despite this, investigation and diagnosis remain largely unstandardised, with unclear guidance over when to use imaging and which modalities are best[2]. The UK has continued to struggle with high admission rates for patients with right iliac fossa pain and high normal appendicectomy rates (NAR), underpinned by variable cross-sectional imaging rates[3]. Although some countries have reported high preoperative imaging use, many countries have no national data and are likely to have similar variation in practice to the UK[4-7].

We showed in RIFT-1 that in the UK, patients with right iliac fossa pain who have a low risk score have a very high NAR when operated, and can safely be managed on an outpatient basis[3]. However, there remains an evidence gap around diagnostic pathways for patients with a medium or high risk score. An evidence-based approach to diagnosis would allow for improved shared decision making with patients, discussion around antibiotic-first therapy, and preoperative surgical planning for minimally invasive surgery. It could avoid unnecessary admissions, prevent unnecessary surgery, and ensure that normal appendicectomy is performed as a planned, targeted therapy rather than a result of misdiagnosis.

This study aimed to examine the diagnostic accuracy of clinical risk prediction models and imaging for the diagnosis of appendicitis in patients with RIF pain who had a medium or high risk score.

Methods

The Right Iliac Fossa Pain Treatment Study captured data on patients presenting to hospital with acute right iliac fossa pain across the UK, Italy, Portugal, Republic of Ireland, and Spain. Analyses of this dataset have previously found high levels of data completeness (99.1%) and data accuracy (98.3%)[3].

Study population

UK has a higher NAR than the other countries that participated in RFT, so this analysis was restricted to the UK where there is a specific need to explore strategies to reduce NAR. A previous analysis of RIFT data identified the optimal risk prediction models and associated cut-offs to identify men (Appendicitis Inflammatory Response Score[8], cut-off score ≤2) and women (Adult Appendicitis Score[9], cut-off score ≤8) aged 16-45 years at low risk of appendicitis. This study aimed to focus on diagnostic strategies in medium and high risk patients, so it included men with AIRS score >2 and women with AAS score >8. There are different differential diagnoses for right iliac fossa pain in younger adults (gynaecological pathology in women) versus older adults (malignancy), so analyses were pre-planned to be stratified by age group (≤45 years versus >45 years).

Data collection

The RIFT Study protocol has been published[10]. All UK hospitals providing acute general surgery were invited register for the study. Participating hospitals collected data on all consecutive patients presenting with suspected appendicitis during one or more of four pre-specified two-week study periods between 13 March 2017 and 18 June 2017. Patients were included if they were referred by either a general practitioner or an emergency physician to the on-call surgical team with either acute right iliac fossa pain or suspected acute appendicitis. Consecutive patients were identified at the point of admission to the surgical unit. Patients who had previously undergone appendicectomy were excluded. Pregnant women were excluded, due to the distinct clinical pathways that these patients follow. As the AIRS and AAS models were used to define the medium- and high-risk patient populations, patients missing variables required to calculate AIRS / AAS were excluded from this analysis. As it was not possible to determine the final diagnosis (appendicitis versus normal appendicectomy) for patients treated for presumed appendicitis for whom histopathology was not available, these patients were excluded from the study. This applied both to patients who underwent appendicectomy for whom histology was not available and to patients treated non-operatively for acute appendicitis.

Teams of up to three investigators based on the acute surgical unit collected data over each two-week period. To ensure bedside accurate contemporaneous data collection, a case report form was designed to be completed at the point of initial surgical assessment. Data collection was supervised by a consultant surgeon at each hospital. Data items required to calculate the AIRS and AAS models were collected, along with data on ultrasound, CT, and MRI imaging results taken from formal radiology reports. In patients who underwent surgery, details of the procedure were recorded along with any subsequent histopathology results. Electronic and paper records were used to follow up patients at 30-days following initial presentation, to capture any surgical procedures performed following readmission.

Clinical outcomes

Patients were classified as having appendicitis if they underwent appendicectomy or right hemicolectomy for presumed acute appendicitis within 30 days of initial presentation, and histopathological examination of the appendix confirmed a diagnosis of acute appendicitis. Appendicitis was subcategorised based on histopathology reports as either simple or complex (gangrenous, perforated) appendicitis.

The numerator for the NAR was the number of patients with normal appendix histology, and the denominator was the total number of all patients who underwent appendicectomy. Patients with appendix pathology other appendicitis (e.g. tumour) were included in the denominator but not the numerator.

Statistics and validation of risk prediction models

Baseline characteristics were described by presenting simple counts and percentages. This study evaluated the performance of AIRS in men and AAS in women for identifying patients at high-risk of appendicitis. The reference standard was histopathological diagnosis of acute appendicitis within 30 days of initial assessment.

The performance of the risk prediction models was evaluated by the positive predictive value (PPV, the proportion of patients stratified to the high-risk group who have appendicitis) and sensitivity (proportion of patients with appendicitis who were stratified to the high-risk group). For each patient sub-group score cut-offs were systematically varied and PPV and sensitivity calculated at each cut-off. Based on a consensus exercise with senior surgeons, it was pre-determined that PPV ≥95% would need to be paired with a sensitivity ≥30% to embed use of risk prediction models to identify patients with appendicitis.

Validation of imaging investigations

The performance of imaging (ultrasound and CT) was assessed by calculation of area under the curve (AUC), sensitivity, specificity, NPV and PPV. In addition, 95% confidence intervals (95% CI) were calculated for all measures of diagnostic performance. Analyses were carried out in Stata (Version 15, Stata Corp., College Station, Texas).

Study approval and reporting

This observational study made no changes to clinical care pathways, with no additional follow-up required. Only anonymised data was collected. Therefore, the study was registered local at each participating hospital as either clinical audit or service evaluation. This study is reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines and the Standards for Reporting Diagnostic Accuracy (STARD) guidelines for diagnostic accuracy studies[11,12].

Results

This study was based on 4,189 UK medium and high risks patients. This included 1,569 women aged 16-45 years and 662 women aged ≥46 years had AAS >8, and 1,417 men aged 16-45 years and 541 men aged ≥46 years had AIRS >2 (Figure 1). Women were less likely to be operated than men at both age 16-45 years (51.9% versus 66.8%, Table 1) and ≥46 years (50.6% versus 59.0% of women).

Amongst operated patients, women were more likely than men to undergo procedures other than appendicectomy, most frequently diagnostic laparoscopy (7.0% [80/1149] of all procedures in women of all ages) or gynaecological procedures (3.1% [36/1149]). Amongst patients who underwent appendicectomy, the NAR was higher in women than men at both age 16-45 years (18.4% versus 9.5%, Table 1) and ≥46 years (4.5% versus 2.1%, Table 1).

Women who did not undergo appendicectomy most frequently received a diagnosis of non-specific abdominal pain (40.0% [496/1241]) or gynaecological pathology (25.5% [317/1241], Table S1). Men who did not undergo appendicecomy most frequently received a diagnosis of non-specific abdominal pain (44.0% [334/759]) or non-appendix gastrointestinal pathology (32.9% [250/759]).

Performance of risk prediction models

Risk prediction models did not achieve the pre-determined threshold to be used to identify patients with appendicitis in any of the four patient subgroups (Tables S2-S5). Considerable NARs were recorded even patients with high scores, especially in younger patients (Table 2, Figure S1).

Preoperative imaging

Women had higher rates of preoperative imaging than men at both ages 16-45 years (70.9% versus 35.8%, Table 3) and ≥46 years (87.2% versus 80.6%). Whereas ultrasound was the most common imaging modality in women aged 16-45 years (80.3% [894/1113] of all scans in this group), CT was most common in women aged ≥46 years (86.0% [496/577]), men aged 16-45 years (67.1% [341/508]), and men aged ≥46 years (94.5% [412/436]).

Overall, amongst patients who underwent ultrasound, 30.4% (384/1262) of results were either positive or negative for appendicitis, 69.3% (875/1262) of results were equivocal and 3 results were missing (Table 4). Across all patients, ultrasound scans were more frequently reported to show gynaecological pathology than appendicitis (210 versus 158 patients, Table S6).

The diagnostic performance of ultrasound was poor with PPV of 82.9% and AUC of 0.68 (Table 4). In the subgroup of patients who had an ultrasound scan with findings other than appendicitis, the NPV was 89.1% and AUC was 0.55 (Table S7); findings of non-appendix pathology on ultrasound did not exclude appendicitis.

Overall, amongst patients who underwent a CT scan, 87.9% (1377/1567) of results were either positive or negative for appendicitis, 11.7% (184/1567) of results were equivocal and 6 results were missing (Table 4). When pathology other than appendicitis was identified on CT scans, this was most frequently gastrointestinal pathology in all groups, apart from women aged 16-45 years in whom gynaecological pathology was most common (Table S8). The diagnostic performance of CT was good with PPV of 91.7%, sensitivity of 94.8%, and AUC of 0.94 (Table 4).

The overall NAR in patients who had CT imaging alone (3.6%) was lower than in patients who had no imaging (12.4%) or ultrasound imaging alone (19.0%). This pattern was consistent across all age and sex subgroups (Table 5, Figure 2a), with the lowest NARs being in men aged ≥46 years (1.3%) and women aged ≥46 years (2.6%) who underwent CT imaging alone.

Clinical observation

Across all age and sex subgroups, longer preoperative clinical observation was associated with higher NAR (Figure 2b, Table S9). Overall, NAR was 8.8% in patients observed for less than 24 hours, compared to 13.4% in patients observed for 24-47 hours, and 20.0% in patients observed for ≥48 hours. Patients with a normal appendicectomy showed the highest rates of postoperative readmission (10.2%, 23/226, Table S10).

Intraoperative diagnosis

When surgeons made an intraoperative diagnosis of appendicitis, this was correct on 95.4% (1889/1979, Table S11) of occasions. However, when an intraoperative diagnosis was made of a normal appendix, this was incorrect and the patient had appendicitis on 30.4% of occasions (63/207).

Discussion

This study shows that, in medium and high risk UK patients with right iliac fossa pain, relying on clinical evaluation, risk scoring, and ultrasound results in high NAR in both men and women. The thresholds to introduce risk scoring to identify appendicitis in this population were not met. CT scan use in this real-world population was associated with lower NAR. In older patients, where risks of cancer diagnosis and operative risks are higher, CT scan use was more frequent and NAR was lower.

Ultrasound was not associated with low NAR in this UK population. When ultrasound was combined with CT, NAR remained high, suggesting this is a group of patients with more challenging diagnosis. Similarly, longer preoperative observation was associated with higher NAR, also likely to reflect a challenging diagnostic group. Paitents who are admitted for over 24 hours should be prioritised for a CT scan.

Improved diagnosis would allow better shared decision making with patients. For example, it would enable informed discussion about antibiotic therapy in patients with early appendicitis, no faecolith, and the desire to avoid surgery[13]. It would also enable better surgical planning, including early stratification of perioperative antibitoics, and reduced preoperative delays[14], improving surgical bed management, which is critical to elective surgery performance[15].

This study had limitations. In this observational study risk scoring did not directly influence decision making; an interventional study may find risk scoring less effective[16]. We observed clinical outcomes in patients who were selected for imaging by surgeons, so the performance of imaging in these patients may not be generalisable to patients more broadly.

There is evidence from other countries including the United States that CT can reduce NAR[6]. However, most countries have not published robust NAR outcome data and high-quality prospective global studies are needed to establish NAR and imaging use around the world, particularly in lower resource settings.

Primary antibiotic therapy became commonplace for appendicitis during the COVID-19 pandemic[17] but post-pandemic surgery has been re-established as the gold-standard treatment. It is now time to embed evidence-based approaches to diagnosis. Patients with right iliac fossa pain should be risk scored, with low-risk patients triaged to ambulatory management and medium and high-risk patients routinely CT scanned. Patients who have been admitted for 24 hours should be prioritised for CT as further delay is unlikely to be beneficial. Ultrasound should be reserved for investigation of possible gynaecological pathology in women. Normal appendicectomy should become a specific therapeutic option rather than a chance finding, reserved for selected patients with undiagnosed on-going pain; restricting normal appendicectomy in this way would ensure improved informed consent, a more successful surgical encounter from the patient perspective, and a more predictable postoperative course. Implementation studies are now needed to change clinical practice in the NHS; the post-pandemic period is a good opportunity for change whilst surgical systems are being re-developed[18].

Table 1: Operative management				
	Age 16-45 years		Age ≥46 years	
	Men	Women	Men	Women
Operated	66.8% (947/1417)	51.9% (814/1569)	59.0% (319/541)	50.6% (335/662)
Appendicectomy	95.7% (906/947)	86.0% (700/814)	91.9% (293/319)	86.6% (290/335)
Appendicitis	87.6% (794/906)	76.1% (533/700)	91.8% (269/293)	93.1% (270/290)
Simple appendicitis	63.1% (501/794)	66.2% (353/533)	36.8% (99/269)	47.0% (127/270)
Complex appendicitis	36.9% (293/794)	33.8% (180/533)	63.2% (170/269)	53.0% (143/270)
Adenocarcinoma	0% (0/906)	0.1% (1/700)	0.7% (2/293)	0.7% (2/290)
Carcinoid	0.3% (3/906)	1.0% (7/700)	1.0% (3/293)	0.3% (1/290)
Crohn's disease	0% (0/906)	0.1% (1/700)	0% (0/293)	0% (0/290)
Other abnormal histology	2.5% (23/906)	4.1% (29/700)	4.4% (13/293)	1.4% (4/290)
Normal appendix	9.5% (86/906)	18.4% (129/700)	2.1% (6/293)	4.5% (13/290)
Other operations	4.3% (41/947)	14.0% (114/814)	8.1% (26/319)	13.4% (45/335)
Diagnostic laparoscopy	20	71	4	9
Colonic resection	2	3	5	17
Other gastrointestinal procedures	15	4	13	12
Urological procedures	1	1	2	1
Gynaecological procedures	-	32	-	4
Missing	3	3	2	2
Not operated	33.2% (470/1417)	48.1% (755/1569)	41.0% (222/541)	49.4% (327/662)

Table 2: Real-world NAR rates stratified by patient score				
	Men 16-45 years		Men >46 years	
AIRS score	Total pts	NAR	Total pts	NAR
3	75	33.3%	12	8.3%
4	89	14.6%	23	0.0%
5	166	14.5%	49	2.0%
6	181	5.0%	62	3.2%
7	175	4.0%	51	3.9%
8	118	2.5%	45	0.0%
9+	102	4.9%	51	0.0%
	Women 16-45 years		Women >46 years	
AAS score	Total pts	NAR	Total pts	NAR
9	72	27.8%	4	25.0%
10	92	27.2%	20	25.0%
11	95	21.1%	23	8.7%
12	117	20.5%	42	0.0%
13	91	12.1%	42	0.0%
14	76	15.8%	26	7.7%
15	65	12.3%	40	0.0%
16	42	11.9%	31	3.2%
17+	50	8.0%	62	3.2%

NAR: normal appendicectomy rate.

Table 3: Preoperative imaging rates

Preoperative imaging	Age 16-45 years		Age ≥46 years	
	Men	Women	Men	Women
Imaging performed	35.8% (508/1417)	70.9% (1113/1569)	80.6% (436/541)	87.2% (577/662)
Ultrasound only	11.1% (158/1417)	50.0% (784/1569)	4.2% (23/541)	11.5% (76/662)
CT only	21.1% (299/1417)	12.8% (201/1569)	74.5% (403/541)	69.3% (459/662)
Ultrasound and CT	3.0% (42/1417)	7.0% (110/1569)	1.7% (9/541)	5.6% (37/662)
MRI*	0.6% (9/1417)	1.2% (18/1569)	0.2% (1/541)	0.8% (5/662)
No imaging performed	64.2% (909/1417)	29.1% (456/1569)	19.4% (105/541)	12.8% (85/662)

CT: computed tomography; MRI: Magnetic resonance imaging

*Includes patients who had MRI ± ultrasound ± CT

Table 4: Overall diagnostic performance of ultrasound and CT imaging for diagnosis of appendicitis		
	Ultrasound	Computed tomography
Proportion of scanned patients with final diagnosis of appendicitis	26.4% (333/1262)	45.2% (708/1567)
Scan findings for appendicitis		
Positive	12.6% (158/1262)	46.7% (732/1567)
Equivocal	69.3% (875/1262)	11.7% (184/1567)
Negative	17.9% (226/1262)	41.2% (645/1567)
Missing	0.2% (3/1262)	0.4% (6/1567)
Performance for appendicitis		
AUC	0.68 (0.66-0.7)	0.94 (0.93-0.95)
Sensitivity	39.3% (34.1%-44.8%)	94.8% (92.9%-96.3%)
Specificity	97.1% (95.8%-98.1%)	92.8% (90.9%-94.5%)
PPV	82.9% (76.1%-88.4%)	91.7% (89.4%-93.6%)
NPV*	81.7% (79.2%-83.9%)	95.5% (93.9%-96.8%)

AUC: Area under curve; NPV: Negative predictive value; PPV: Positive predictive value.
95% confidence intervals given in parentheses.

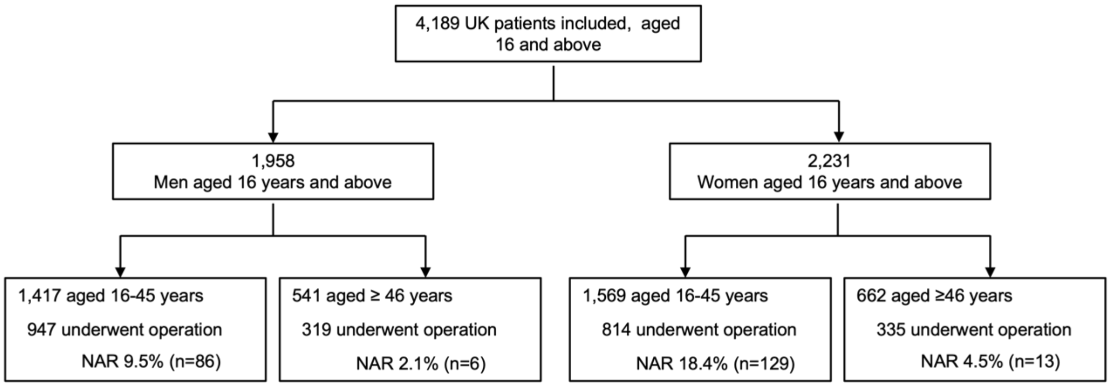
*Main analysis based on scans reported by as either ‘negative’ (i.e. excluding appendicitis) or ‘equivocal’ (i.e. unable to either exclude or confirm appendicitis) both being classified as negative scans; yielding NPV 81.7% for ultrasound, and 95.5% for CT. If only scans reported as ‘negative’ are classified as negative, then the NPV for ultrasound was 96.9% (95% confidence interval 93.7%-98.7%), and the NPV for CT was 99.2% (98.2%-99.7%).

Table 5: Normal appendicectomy rates by preoperative imaging modality					
Preoperative imaging	Age 16-45 years		Age ≥46 years		Overall
	Men	Women	Men	Women	
No imaging	9.8% (64/656)	19.5% (57/292)	5.4% (3/56)	14.3% (4/28)	12.4% (128/1032)
Ultrasound only	16.4% (11/67)	20.9% (58/277)	0% (0/9)	9.5% (2/21)	19.0% (71/374)
CT only	5.5% (9/164)	8.6% (8/93)	1.3% (3/224)	2.6% (6/234)	3.6% (26/715)
Ultrasound and CT	7.1% (1/14)	15.2% (5/33)	0% (0/4)	14.3% (1/7)	12.1% (7/58)
MRI*	20.0% (1/5)	20.0% (1/5)	n/a	n/a	20.0% (2/10)

CT: computed tomography; MRI: Magnetic resonance imaging; n/a: not applicable as no scans performed

*Includes patients who had MRI ± ultrasound ± CT

Figure 1: Flowchart of included patients



NAR: normal appendicectomy rate.

Figure 2a: Normal appendicectomy rates by pre-operative imaging modality

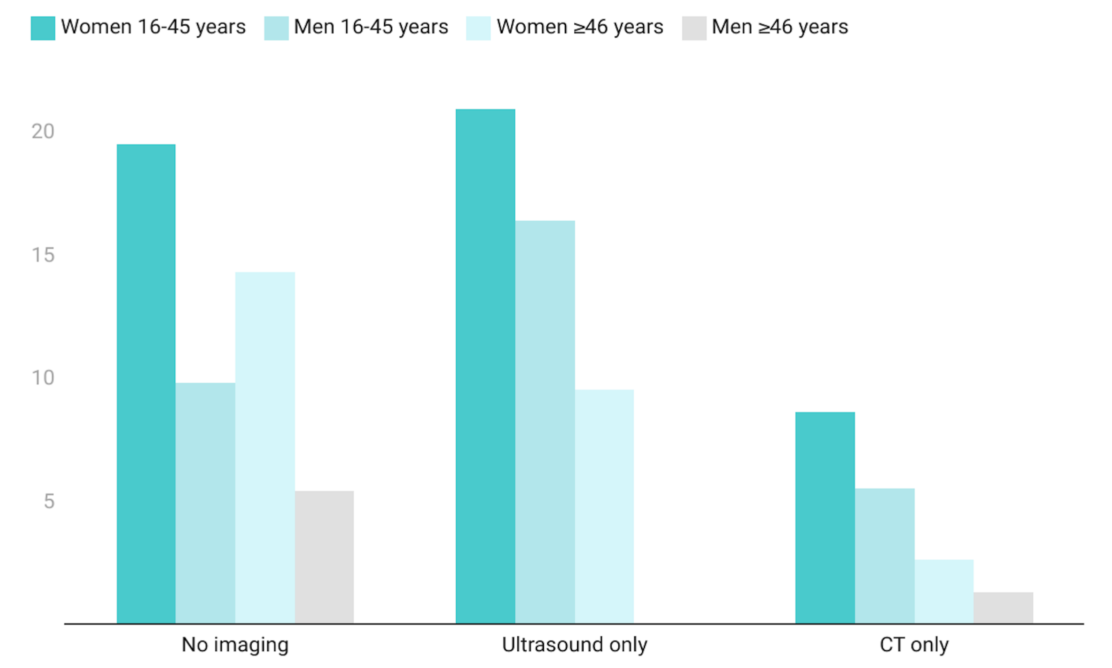
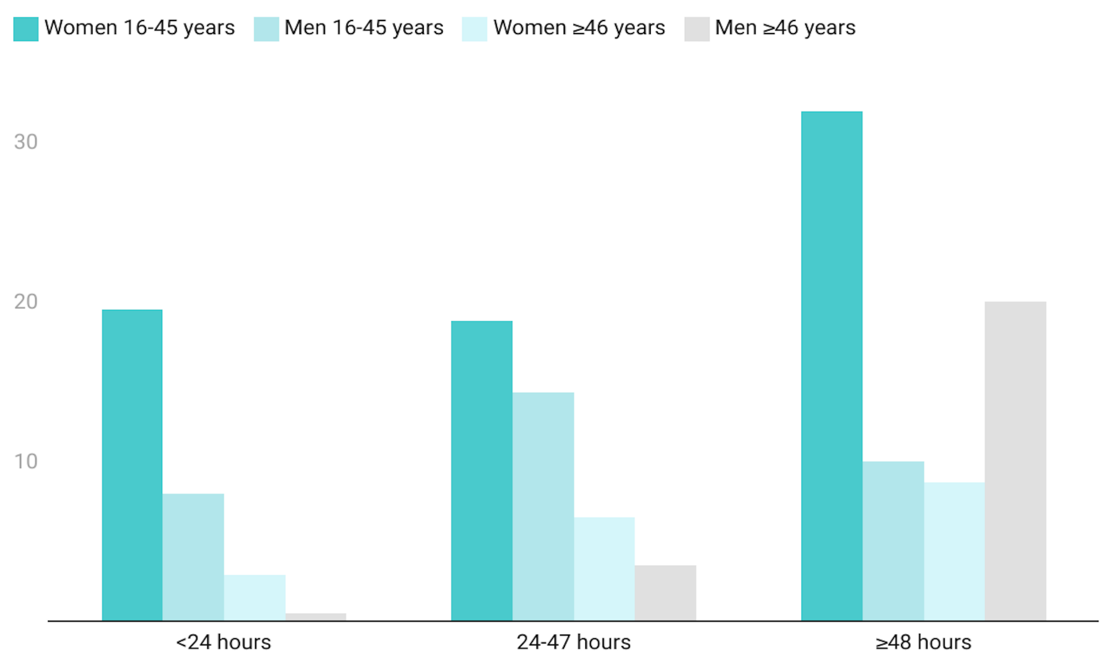


Figure 2b: Normal appendicectomy rate by duration of preoperative clinical observation



References

1. GlobalSurg Collaborative. Mortality of emergency abdominal surgery in high-, middle- and low-income countries. *Br J Surg*. Jul 2016;103(8):971-988.
2. Di Saverio S, Podda M, De Simone B, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg*. Apr 15 2020;15(1):27.
3. 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*. Jan 2020;107(1):73-86.
4. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med*. Apr 26 2012;366(17):1596-605.
5. van Rossem CC, Bolmers MD, Schreinemacher MH, et al. Diagnosing acute appendicitis: surgery or imaging? *Colorectal Dis*. Dec 2016;18(12):1129-1132.
6. Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and socioeconomic implications of negative appendectomy. *Am J Surg*. Dec 2016;212(6):1076-1082.
7. Bhangu A, Soreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet*. Sep 26 2015;386(10000):1278-1287.
8. 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*. Aug 2008;32(8):1843-9.
9. Sammalkorpi HE, Mentula P, Leppaniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis--a prospective study. *BMC Gastroenterol*. Jun 26 2014;14:114
10. 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*. Jan 13 2018;8(1):e017574.
11. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. Nov 14 2016;6(11):e012799. doi:10.1136/bmjopen-2016-012799
12. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. Apr 2019;13(Suppl 1):S31-S34.
13. 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*. May 2014;259(5):894-903.
14. de Wijkerslooth EML, Boerma EG, van Rossem CC, et al. 2 days versus 5 days of postoperative antibiotics for complex appendicitis: a pragmatic, open-label, multicentre, non-inferiority randomised trial. *Lancet*. Feb 4 2023;401(10374):366-376.
15. NIHR Global Health Unit on Global Surgery, COVIDSurg Collaborative. Elective surgery system strengthening: development, measurement, and validation of the surgical preparedness index across 1632 hospitals in 119 countries. *Lancet*. Nov 5 2022;400(10363):1607-1617.
16. 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*. Sep 1999;134(9):993-6.
17. COVIDSurg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study. *Lancet*. Jul 4 2020;396(10243):27-38.
18. NIHR Global Health Research Unit on Global Surgery. Strategies to strengthen elective surgery systems during the SARS-CoV-2 pandemic: systematic review and framework development. *Br J Surg*. Jan 31 2024;111(2).