



# Diagnostic accuracy of a pragmatic, ultrasound-based approach to adult patients with suspected acute appendicitis in the ED

Beat Lehmann , Ursina Koeflerli, Thomas C Sauter , Aristomenis Exadaktylos, Wolf E Hautz 

Handling editor Simon Carley

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/emmermed-2019-208643>).

Department of Emergency Medicine, Inselspital University Hospital Bern, Bern, Switzerland

## Correspondence to

Dr Beat Lehmann, Emergency Medicine, Inselspital University Hospital Bern, Bern 3010, Switzerland; [beat.lehmann@insel.ch](mailto:beat.lehmann@insel.ch)

Received 6 April 2019  
Accepted 22 February 2022  
Published Online First  
17 March 2022

## ABSTRACT

**Background** Systematic imaging reduces the rate of missed appendicitis and negative appendectomies in patients with suspected acute appendicitis (AA). Little is known about the utility of ultrasound as a first diagnostic measure in patients with suspected AA. The aim of this retrospective study is to determine whether ultrasound, performed by emergency physicians or radiologists, can be used as first diagnostic measure in suspected cases to rule out AA and to avoid unnecessary CT.

**Methods** We performed a retrospective analysis at the ED of the University Hospital Bern, Switzerland, from 2012 to 2014. Our standard protocol is that all adult patients suspected of appendicitis receive an ultrasound as their first imaging test, either by an emergency physician or a radiologist. The test characteristics of conclusive and inconclusive ultrasound exams were compared with a pragmatic gold standard.

**Results** The study included 508 patients with suspected AA. 308 patients (60.4%) had a conclusive ultrasound. Among these, sensitivity for appendicitis was 89.6% (95% CI 82.1% to 94.3%), specificity 93.8% (89.1% to 96.6%), the positive predictive value was 87.98 (80.84 to 92.71) and the negative predictive value was 94.65 (91.18 to 96.80). The remaining 200 (39.4%) patients had an inconclusive ultrasound exam. 29% (59/200) of these patients ultimately had appendicitis. Less experienced emergency physician sonographers came to a definitive conclusion in 48.1% (95% CI 36.9% to 59.5%), experienced emergency physician sonographers in 76.0% (68.4% to 82.5%) and radiologists in 52.4% (44.5% to 60.2%).

**Conclusion** A conclusive ultrasound of the appendix performed by either emergency physicians or radiologists is a sensitive and specific exam to diagnose or exclude AA in patients with suspected AA. Because of 6% false negative exams, clinical follow-up is mandatory for patients with negative ultrasound. An inconclusive ultrasound warrants further imaging or a follow-up visit, since 29% of patients with inconclusive ultrasound had an AA.

## INTRODUCTION

Acute abdominal pain is among the most common complaints of patients visiting the ED and represents around 8% of the total number of all ED visits.<sup>1</sup> The most common diagnoses are 'non specific abdominal pain' (44%), followed by acute appendicitis (AA) (16%), although there is substantial variation between different studies.<sup>2</sup>

## Key messages

### What is already known on this subject

⇒ Acute appendicitis is the most frequent specific pathology in patients with an acute abdomen presenting to an ED. In order to minimise the negative appendectomy rate, systematic imaging of the appendix is recommended. Several studies compared sensitivity and specificity of ultrasound exams versus CT, not further specifying conclusive versus inconclusive ultrasound exams. Therefore, little is known about the utility of ultrasound as a first diagnostic measure in patients with suspected acute appendicitis in the ED.

### What this study adds

⇒ In this retrospective, single-centre study of 508 patients who received abdominal ultrasound as first investigation for suspected acute appendicitis, conclusive ultrasound (negative or positive) performed by either emergency physicians or radiologists had a good sensitivity and specificity (90% and 94%, respectively). Of all patients with an inconclusive ultrasound, almost one-third had an acute appendicitis. An inconclusive ultrasound therefore requires further imaging or a clinical follow-up. Patients with negative ultrasounds should have follow-up as well.

### How this study might affect research, practice and policy

⇒ Sonography, along with other imaging modalities, plays an important role in the evaluation of patients with suspected acute appendicitis. The present study shows that conclusive point-of-care ultrasound performed by emergency physicians is sensitive and specific in order to confirm or to exclude acute appendicitis. Further studies are necessary to validate these results.

A missed diagnosis of AA results in an increased rate of perforation, morbidity and mortality.<sup>3–5</sup> However, it is also important to minimise the negative appendectomy rate (NAR), because negative appendectomies are associated with increased morbidity and mortality.<sup>6</sup>



© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Lehmann B, Koeflerli U, Sauter TC, et al. *Emerg Med J* 2022;**39**:931–936.

Dahlberg and colleagues<sup>7</sup> demonstrated a reduction in the NAR from 10.9% to 1.7% in a Scandinavian hospital between 2007 and 2014. In the same period, preoperative imaging increased from 30% to 93%. These data suggest that the reduction in NAR was due to systematic preoperative imaging.

CT is the gold standard for the diagnosis of AA, with a pooled sensitivity of 94% and pooled specificity of 95% in a recent meta-analysis.<sup>8</sup> However, its irradiation, cost, the limited availability in certain settings and the resulting delay limit its use. Ultrasound has been established as a safe, non-irradiating technology, which is widely used and immediately available. Due to its dependence on examiner skill and a somewhat lower diagnostic accuracy compared with CT, the role of ultrasound in the diagnosis of AA is still unclear.

Studies investigating the sensitivity and specificity of a specific imaging modality for a given disease are of somewhat limited clinical use, because, in clinical practice, patients do not present with AA that requires imaging confirmation, but present with abdominal symptoms that are more or less specific for one or many diseases. Imaging, among other tools, is used to increase or decrease the probability of one of these diseases. Which images are acquired by which type of examiner and on which patients is in reality determined by physician expertise, patient characteristics and contextual factors such as imaging availability and local guidelines. The interplay between these factors results in what is called 'clinical or diagnostic practice'. The evaluation of such practice is potentially more informative for patient care than the post hoc determination of one imaging modality's performance over another for patients where the diagnosis is now established.

In the ED of the University Hospital in Bern, Switzerland, diagnostic practice for suspected AA consists of a clinical exam, standard lab followed by an abdominal ultrasound, performed by either an ED physician or a radiologist, depending on the ED physician's level of training in ultrasound. Our surgeons do not require a CT scan prior to surgery in patients with conclusive ultrasounds and clinical findings, which correlate with the diagnosis (such as pain in typical location and signs of infection). Patients with positive ultrasounds will proceed to surgery; patients with negative ultrasounds attend a follow-up visit at the ED's Fast Track (minor area) within the next 1–2 days. If the initial ultrasound is inconclusive (ICUS), most often because the appendix has not been visualised, either additional imaging (CT and MR) or a follow-up visit are scheduled, based on clinician judgement, depending mainly on the acuity of the patient's presentation and laboratory findings (figure 1).

The aim of this study was to evaluate this strategy, determine the rate of conclusive exams and the sensitivity and specificity of point-of-care ultrasound (POCUS) for AA. We hypothesise

that in case of a conclusive ultrasound (positive or negative), no further imaging is necessary.

## MATERIAL, METHODS AND PATIENTS

### Setting

This is a retrospective analysis performed at the ED of the University Hospital (Inselspital) Bern, Switzerland. We are a level I, university-affiliated trauma centre that sees around 48 000 adult patients annually, mostly from urban Bern, which has around 135 000 inhabitants and some from the surrounding catchment area of around 1 million people.

### Participants

We retrieved all data of adult patients who presented with suspected AA between May 2012 and July 2014 from our electronic patient documentation system (E.care, Turnough, Belgium). Specifically, we searched the documentation full text for the terms 'appendix' and 'appendicitis' (their German equivalent, actually) and spelling variations thereof.

As noted earlier, our standard protocol is that patients receive an ultrasound as their first imaging test in cases of suspected appendicitis. Ultrasounds are performed either by emergency medicine physicians or by radiologists at the discretion of the treating physician. Typically, an examiner would diagnose AA on ultrasound based on the consideration of the following signs: enlarged appendix (>6 mm diameter), non-compressibility, pain on compression, increased periappendicular echogenicity and fluid collection.<sup>9</sup> Where the appendix was not visualised, the examiner would consider it an ICUS.

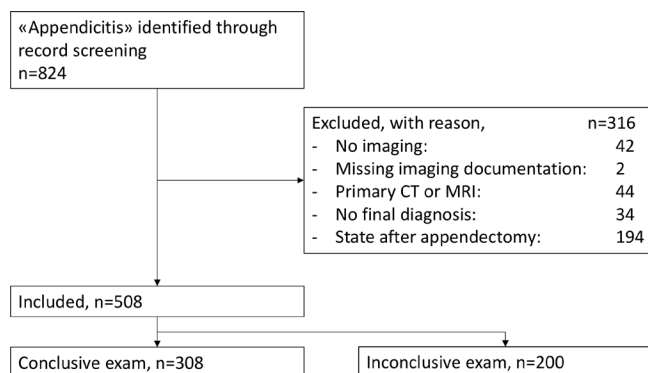
In order to compare accuracy of different groups of sonographers, we classified sonographers as emergency physicians or radiologists. Emergency physicians were further divided into experienced versus inexperienced examiners on the basis of the number of examinations conducted and progress in their training in sonography. Those licenced to perform unsupervised exams in Switzerland require 200 supervised abdominal exams and successful completion of a 3-day course). These individuals were classified as experienced, and other emergency physicians were classified as inexperienced.

### Data collection

From all patient records, we extracted demographic information, patient history, clinical characteristics, imaging studies and reported results, and the patient's course during and after the ED visit. We calculated the Alvarado Score based on the available information from the patient charts, although not all elements were routinely documented.<sup>10</sup>

Sonographic examinations were classified in three groups based on the sonographer's report: AA, normal appendix or inconclusive exam (the latter including an appendix that was not visualised). The sonography result was compared with the gold standard depending on clinical course. When intraoperative or histopathological findings were available, these determined the gold standard. If CT was performed after sonography but appendectomy was not performed, the CT was the reference standard. For patients with a sonographically normal appendix and no further imaging or surgery, the patient's medical record was examined for planned and unplanned follow-up visits and their results.

Sonography results were characterised as true positive and true negative where the ultrasonography report matched with the results of the gold standard; false positive and false negative cases where the diagnosis of the ultrasonography differed from



**Figure 1** Patient enrolment flow chart.

the gold standard, and ICUS. Test characteristics (sensitivity, specificity, negative and positive predictive values and likelihood ratios) were calculated for sonograms, and individually for those performed by emergency physicians (EP) and radiologists. The proportion of ICUS for radiologists, experienced EPs and inexperienced EPs was also determined.

To calculate the required sample size, we assumed an estimated sensitivity of the ultrasound of 85%, an acceptable width of the 95% CI of 10% and a prevalence of disease of 15% in the examined population. We further assumed that two-thirds of all patients examined with ultrasound would have a conclusive exam, resulting in a total required sample size of 523 patients according to the formula by Buderer.<sup>11</sup> Given that we examine around 250 patients with suspected appendicitis per year, we determined a 25-month duration for record evaluation.

Descriptive statistics were used when applicable. For group comparisons, Pearson's  $\chi^2$  test was used for nominal variables and analysis of variance test for metric variables. Sensitivity was calculated as true positive / (true positive + false negative) and specificity as true negative / (true negative + false positive). CIs were determined by means of bootstrapping (with 1000 repetitions). The SPSS Statistics program, version 25 (IBM Corp) was used for all statistical calculations, and a two tailed p value of <0.05 was considered statistically significant. Online

supplemental table 1 lists the specific statistical tests and their findings.

### Patient and public involvement

No patient involved.

### RESULTS

A total of 824 patients were identified through screening of medical records. Three hundred and sixteen patients were excluded (figure 1). Thus, the data of 508 patients (62%) were available for analysis. Table 1 summarises the demographic characteristics and clinical findings of the sample.

The clinical presentation of AA varied considerably and a significant number of patients did not show typical findings of AA (eg, RLQ pain, RLQ pain on palpation, rebound tenderness, etc). The most common final diagnosis among patients with abdominal pain presenting in our ED were non-specific abdominal pain (44%), followed by AA (34%), enteritis (12%), UTI (2%) and others (8%). Prevalence of AA among all patients included in this study was 34% (95% CI 29.6% to 37.9%).

### Test characteristics for ultrasound

Of the 508 analysed patients, 308 (60.6%) had a conclusive ultrasound during their first visit. Sonography was read as

**Table 1** Demographic and clinical characteristics of 508 patients with suspected appendicitis

	True positive (n=103)	True negative (n=181)	False positive (n=12)	False negative (n=12)	Inconclusive (n=200)	All (n=508)
Age (mean (SD))	31.8 (12.04)	30.5 (13.4)	28.6 (9.76)	44.9 (24.18)	31.1 (12.64)	31.3 (13.24)
Sex (M/F)	58 (56%)/45 (44%)	55 (30%)/126 (70%)	3 (25%)/9 (75%)	6 (50%)/6 (50%)	77 (39%)/123 (62%)	199 (39%)/309 (61%)
History, n (%)						
Abdominal pain						
Right lower quadrant (RLQ)	85 (83)	112 (62)	11 (92)	9 (75)	155 (78)	372 (73)
Left lower quadrant (LLQ)	7 (7)	37 (20)	2 (17)	5 (42)	27 (14)	78 (15)
Epigastric pain	3 (3)	22 (12)	1 (8)	1 (8)	16 (8)	42 (8)
Diffuse abdominal pain	14 (14)	30 (17)	1 (8)	2 (17)	28 (14)	75 (15)
Flank pain	3 (3)	13 (7)	–	1 (8)	14 (7)	31 (6)
Start as epigastric pain	44 (43)	21 (12)	7 (58)	3 (25)	45 (23)	120 (24)
Nausea	35 (34)	65 (36)	7 (58)	4 (33)	62 (31)	173 (34)
Vomiting	42 (41)	47 (26)	3 (25)	6 (50)	57 (29)	155 (31)
Diarrhoea	17 (17)	46 (25)	1 (8)	1 (8)	38 (19)	103 (20)
Dysuria	4 (4)	17 (9)	3 (25)	1 (8)	14 (7)	39 (8)
Medical findings						
BMI (mean (SD))	23.5 (2.98); n=80	22.9 (3.74); n=119	23.1 (4.21); n=10	23.3 (3.93); n=8	23.7 (4.39); n=145	23.3 (3.88); n=362
Temperature (mean (SD))	37.1 (0.71); n=103	36.9 (0.74); n=176	37.2 (0.81); n=12	37.6 (0.96); n=12	37.0 (0.71); n=196	37.0 (0.73); n=499
Pain on palpation, n (%)						
RLQ	80 (78)	99 (55)	10 (83)	8 (67)	138 (69)	335 (66)
LLQ	9 (9)	26 (14)	2 (17)	1 (8)	27 (14)	65 (13)
Left upper quadrant/Right upper quadrant	4 (4)	23 (13)	3 (25)	3 (25)	20 (10)	53 (10)
Painful flank percussion	9 (9)	25 (14)	2 (17)	2 (17)	29 (15)	67 (13)
Rebound tenderness	63 (61)	52 (29)	8 (67)	6 (50)	66 (33)	195 (38)
Psoas sign	15 (15)	15 (8)	3 (25)	2 (17)	27 (14)	62 (12)
Alvarado Score	5.6 (1.5)	3.6 (1.69)	5.1 (1.93)	5.1 (2.46)	4.3 (1.84)	4.4 (1.89)
Laboratory findings						
CRP in mg/dL (median, IQR)	14 (51.5)	3 (16)	13 (65)	51 (180)	5 (37.3)	6 (33)
Leucocyte count in G/l (mean, (SD))	13.55 (3.95)	9.54 (4.09)	10.63 (3.52)	13.77 (6.47)	11.02 (3.98)	11.08 (4.32)
Final diagnosis acute appendicitis after all imaging/follow-up, n (%)	103	–	–	12	59 (29)	174 (34)

BMI, body mass index; CRP, C reactive protein.

Table 2 Overview of diagnostic characteristics of all conclusive sonographies

	N total	N conclusive (% of total)	Sensitivity (95% CI)	Specificity (95% CI)	PPV* (95% CI)	NPV* (95% CI)	LR+ (95% CI)	LR- (95% CI)	N inconclusive (% of total)
<b>Overall</b>	508	308	89.6 (82.1 to 94.3)	93.8 (89.1 to 96.6)	87.98 (80.84 to 92.71)	94.65 (91.18 to 96.80)	14.4 (8.3 to 25.01)	0.11 (0.06 to 0.19)	200
Radiologists	166	87(52.4)	90.32 (74.25 to 97.96)	91.07 (80.38 to 97.04)	83.72 (68.85 to 92.29)	94.88 (86.29 to 98.20)	10.12 (4.35 to 25.54)	0.11 (0.04 to 0.31)	79 (47.6)
ED physicians overall	342	221(64.6)	89.29 (80.63 to 94.98)	94.89 (89.76 to 97.92)	89.88 (81.13 to 94.83)	94.57 (90.37 to 97.0)	17.47 (8.46 to 36.1)	0.11 (0.06 to 0.21)	121 (35.4)
Advanced ED sonographers	154†	117(75.97)	90.91 (78.33 to 97.47)	94.52 (86.56 to 98.49)	89.4 (76.4 to 95.65)	95.34 (88.92 to 98.12)	16.59 (6.37 to 43.22)	0.1 (0.04 to 0.25)	37 (24.03)
Unexperienced ED sonographers	81†	39(48.15)	87.50 (61.65 to 98.45)	91.3 (71.96 to 98.93)	83.65 (57.32 to 95.12)	93.49 (79.62 to 98.14)	10.06 (2.64 to 38.32)	0.14 (0.04 to 0.5)	42 (51.85)

\*Based on an overall prevalence of appendicitis in this study of 33.7%.

†Examiner experience missing for 107 exams.

LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

positive in 115 patients, and 103 (89.6%) had appendicitis based on the reference standard. The sonogram was read as negative in 193 patients, and 12 (6.2%) of these had appendicitis by the reference standard.

The overall sensitivity of POCUS for AA of conclusive examinations was 89.6% (95% CI 82.1% to 94.3%), specificity was 93.8% (89.1%–96.6%), respectively (table 2). The positive likelihood ratio (LR+) for a conclusive exam (ie, the improvement in post-test probability for AA resulting from a positive conclusive exam) thus is  $LR+ = 14.4$  (95% CI 8.3 to 25.01), while LR- (ie, the decrease in post-test probability for AA resulting from a negative conclusive exam; a large decrease) was 0.11 (0.06 to 0.19). Both LRs are considered to have a large impact on the pretest probability.

Two hundred (39.4%) patients had inconclusive primary ultrasound exams, and among these, 59 (29%) were ultimately diagnosed with AA. The details regarding the diagnosis and follow-up of all patients are presented in figure 2.

### Association of ultrasound findings and examiner

The ultrasound exam was performed in 259 cases (51%) by emergency physicians and in 249 cases (49%) by radiologists. Among the emergency physicians, 65% were classified as experienced and 35% as unexperienced examiners, and examiner experience was missing for 107 exams.

Radiologists had a slightly higher sensitivity (90% vs 89%) and slightly lower specificity (91% vs 95%) than emergency physicians. The small number of conclusive exams per group when split into examiner groups means these numbers should be interpreted with caution (see table 2).

The probability of an ICUS depended significantly on the examiner. Two hundred and twenty-one (64.6%) of the studies performed by an emergency physician were conclusive.

Among the emergency physician sonographers, unexperienced examiners reached a definitive conclusion in 48.1% (n=39; 95% CI 36.9% to 59.5%) of exams, while advanced or certified sonographers reported conclusive exams in 117 studies (76.0%; 95% CI 68.4% to 82.5%) (table 2).

Among the variables body mass index (BMI), gender, age, temperature, CRP, leucocyte count and Alvarado score, there was no statistical difference between conclusive and inconclusive ultrasound.

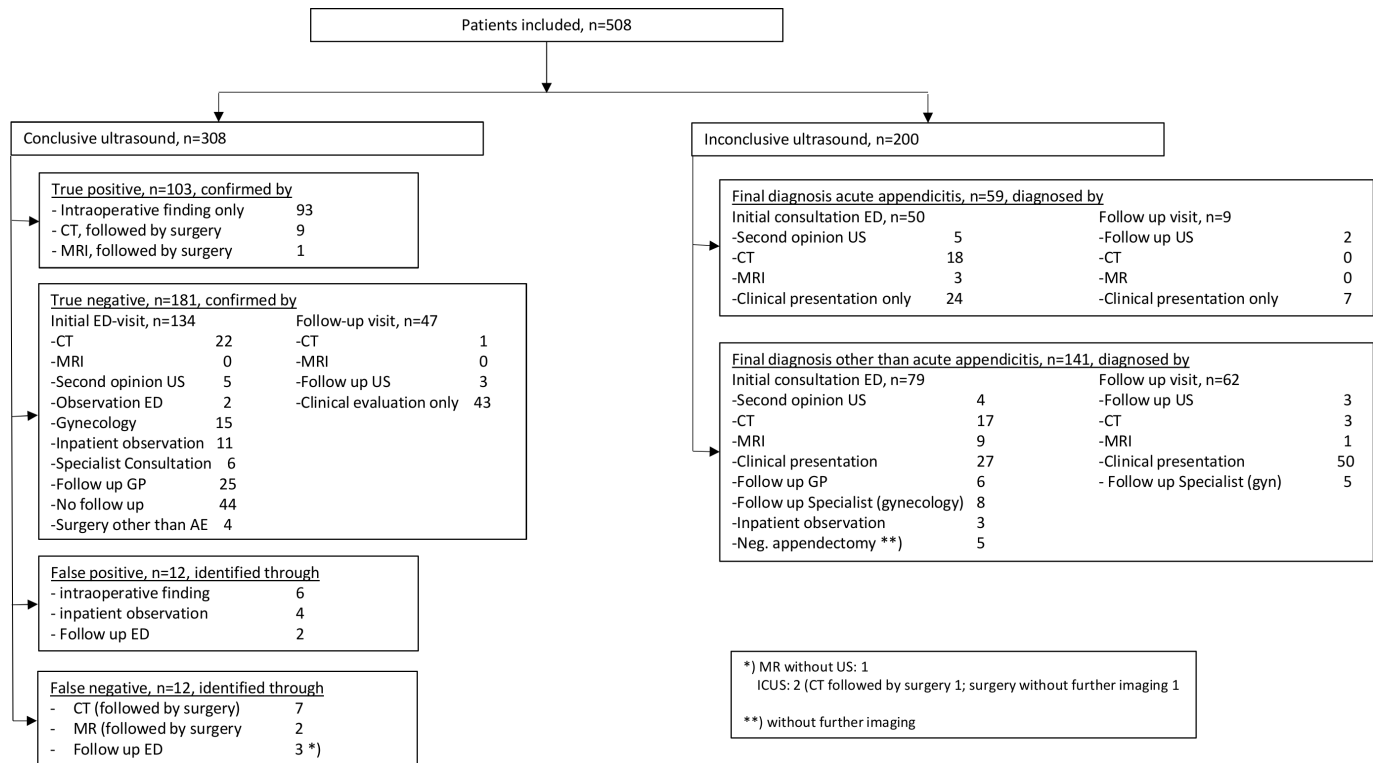
Patients examined with ultrasound by a radiologist were no different from patients examined by emergency physicians in age or BMI (all  $p > 0.05$ , see online supplemental file 1).

The NAR was 5.9% overall. Among the 11 patients with negative appendectomy, six patients had a false positive ultrasound and five had a non-conclusive ultrasound. Among the latter group, a diagnostic laparoscopy was performed without further preoperative imaging. NAR drops to 3.5% when only patients with a conclusive ultrasound exam are considered.

### DISCUSSION

In this retrospective analysis, we investigated the role of ultrasound in the evaluation of suspected AA in the ED. A conclusive ultrasound (either positive or negative) for patients in our study resulted in clinically meaningful likelihood ratios that substantially affect the post-test probability. However, where ultrasound was inconclusive, 29% of patients eventually had a confirmed diagnosis of AA. An ICUS therefore does not allow exclusion of AA and further imaging or a close follow-up-visit is mandatory.

These findings also suggest it is crucial that the sonographer determines one's own confidence in the exam, that is, whether



**Figure 2** Results and diagnostic steps for conclusive and Inconclusive ultrasound.

the exam is conclusive or not, as only conclusive examinations allowed for the confirmation or exclusion of AA as indicated by LR+ of 14.4 and LR- of 0.11. Interestingly, radiologists showed a higher number of ICUS than emergency physicians, although the patients seen by radiologists were of comparable ultrasonographic difficulty as indicated by their age and BMI. One possible reason for this finding may be that emergency physicians performing bedside ultrasound have access to clinical and laboratory data that may strengthen their confidence if concordant with their sonographic findings, thus making them more likely to rate an exam as conclusive. Experienced emergency physician examiners showed a significantly better sensitivity and specificity compared with less experienced providers, which is well known from the literature.<sup>12</sup>

In our population, the clinical presentation of AA varied considerably and only few patients had the typical history and/or clinical findings. The wide spectrum of clinical presentations might explain the relatively high NARs that have been reported historically, since in order to minimise the number of missed appendicitis, appendectomy has been performed liberally.<sup>6</sup> This underlines the importance of a systematic imaging protocol in patients with suspected AA. The NAR among our patients was 5.9%, which is comparable with other publications (Güller 2.8%, Schok 12%, Sammakorpi 8.7%).<sup>13-15</sup>

The performance of ultrasound for AA in our patient cohort is somewhat better than seen in many studies.<sup>7 9 16 17</sup> In a meta-analysis by Lee and Yun, the pooled sensitivity was 0.84 and specificity was 0.91.<sup>18</sup> A possible explanation for a more favourable performance in our group is that patients with ICUS have been considered as a separate group.

In 60.6% of all cases, a conclusive diagnosis was made with ultrasound. The rate of ICUS in our population was lower with 39% than in two recent studies that showed an ICUS rate of 49% and 74%.<sup>19 20</sup> We could not find an association between body

mass index and the frequency of ICUS found in some previous studies.<sup>19 20</sup>

In our population, the rate of positive CT scans is relatively high with 42%. This high rate of positive exams is likely due to the fact that CT was limited to only patients with ICUS and corresponds well to a study by Atema *et al*<sup>21</sup> that showed a reduction of CT scans up to 50%, if CT scans are only performed if a ICUS is present.

### Limitations

As a retrospective analysis, there is no guarantee of the completeness and correctness of the recorded patient data. Some patients may have been missed due to our search strategy. We have not been able to track all the patients after their initial ED visit. It might therefore be possible that patients were treated in another hospital, even if we should normally be informed about this. For this reason, we estimate the number of patients being treated in other institutions as being very small and therefore non-relevant. Decisions about imaging modalities and treatment were made by the treating physicians and were not standardised. Since the ultrasound exams in our study were performed by a physician in a real-life ED setting, interobserver and intraobserver variability cannot be excluded, though all physicians performing the ultrasound were specifically trained in abdominal ultrasound. One further limitation is the fact that different gold standards have been used (CT, appendectomy and follow-up visit).

Since the examiners in our study are very well trained, many of them having more than 200 supervised exams, our data cannot be transferred to a population of physicians with a lower training standard.

Lastly, our study lacks external validation or internal cross-validation, a shortcoming that could be addressed in future, preferably multicentric studies.

## CONCLUSION

In our ED, conclusive ultrasound of the appendix is a sensitive (89.6%) and specific (93.8%) exam in order to diagnose or exclude AA. However, a 6% false-negative rate suggests clinical follow-up the next day is mandatory for patients with negative ultrasound exam.

ICUS cannot be interpreted as negative as there is a high rate of appendicitis in this group. The more experienced the examiner, the lower is the rate of ICUS and the better are sensitivity and specificity.

**Contributors** All authors significantly contributed to the final manuscript in its present form, including conception and design of the study, drafting of the manuscript and finally approval of the manuscript. BL is the guarantor for the paper.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and was approved by Kantonale Ethikkommission Murtenstrasse 31 Hörsaaltrakt Pathologie, Eingang 43A, Büro H3723010 Bern Reference number:KEK-BE No 155-15. Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

## ORCID iDs

Beat Lehmann <http://orcid.org/0000-0003-3218-2470>

Thomas C Sauter <http://orcid.org/0000-0002-6646-5789>

Wolf E Hautz <http://orcid.org/0000-0002-2445-984X>

## REFERENCES

- Hastings RS, Powers RD. Abdominal pain in the ED: a 35 year retrospective. *Am J Emerg Med* 2011;29:711–6.
- Strömberg C, Johansson G, Adolfsson A. Acute abdominal pain: diagnostic impact of immediate CT scanning. *World J Surg* 2007;31:2347–54.
- Öztürk A, Korkmaz M, Atalay T, et al. The role of doctors and patients in appendicitis perforation. *Am Surg* 2017;83:390–3.
- Wen SW, Naylor CD. Diagnostic accuracy and short-term surgical outcomes in cases of suspected acute appendicitis. *CMAJ* 1995;152:1617–26.
- Al-Omran M, Mamdani M, McLeod RS. Epidemiologic features of acute appendicitis in Ontario, Canada. *Can J Surg* 2003;46:263–8.
- Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. *Arch Surg* 2002;137:799–804.
- Dahlberg MJA, Pieniowski EHA, Boström Lennart Å S. Trends in the management of acute appendicitis in a single-center quality register cohort of 5,614 patients. *Dig Surg* 2018;35:144–54.
- Doria AS, Moineddin R, Kellenberger CJ, et al. Us or CT for diagnosis of appendicitis in children and adults? A meta-analysis. *Radiology* 2006;241:83–94.
- Reddan T, Corness J, Harden F, et al. Improving the value of ultrasound in children with suspected appendicitis: a prospective study integrating secondary sonographic signs. *Ultrasonography* 2019;38:67–75.
- Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557–64.
- Buderer NM. Methodology: S I. Statistical methodology: I. incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;3:895–900.
- Kim J, Kim K, Kim J, et al. The learning curve in diagnosing acute appendicitis with emergency sonography among novice emergency medicine residents. *J Clin Ultrasound* 2018;46:305–10.
- Güller U, Rosella L, McCall J, et al. Negative appendectomy and perforation rates in patients undergoing laparoscopic surgery for suspected appendicitis. *Br J Surg* 2011;98:589–95.
- Schok T, Simons PCG, Janssen-Heijnen MLG, et al. Prospective evaluation of the added value of imaging within the Dutch National Diagnostic Appendicitis Guideline--do we forget our clinical eye? *Dig Surg* 2014;31:436–43.
- Sammalkorpi HE, Mentula P, Savolainen H, et al. The introduction of adult appendicitis score reduced negative appendectomy rate. *Scand J Surg* 2017;106:196–201.
- Corson-Knowles D, Russell FM. Clinical ultrasound is safe and highly specific for acute appendicitis in moderate to high pre-test probability patients. *West J Emerg Med* 2018;19:460–4.
- Matthew Fields J, Davis J, Alsup C, et al. Accuracy of point-of-care ultrasonography for diagnosing acute appendicitis: a systematic review and meta-analysis. *Acad Emerg Med* 2017;24:1124–36.
- Lee SH, Yun SJ. Diagnostic performance of emergency physician-performed point-of-care ultrasonography for acute appendicitis: a meta-analysis. *Am J Emerg Med* 2019;37:696–705.
- Keller C, Wang NE, Imler DL, et al. Predictors of nondiagnostic ultrasound for appendicitis. *J Emerg Med* 2017;52:318–23.
- Pelin M, Paquette B, Revel L, et al. Acute appendicitis: factors associated with inconclusive ultrasound study and the need for additional computed tomography. *Diagn Interv Imaging* 2018;99:809–14.
- Atema JJ, Gans SL, van Randen A. Ct strategy for patients with suspected acute appendicitis: comparison of conditional and immediate CT. *Eur Radiol* 2015;25:2445–52.

Groups compared	Compared with regard to	P value	Statistical test
Patients examined by ED physician vs. those examined by radiologist	Proportion of conclusive examinations	0.008	$\chi^2$ (Fishers exact test)
Patients examined by ED physician vs. those examined by radiologist	Age	0.689	Mann-Whitney-U-Test
Patients examined by ED physician vs. those examined by radiologist	BMI	0.162	Mann-Whitney-U-Test
Patients examined by experienced ED sonographers vs. those examined by unexperienced ED sonographers	Proportion of conclusive examinations	<0.001	$\chi^2$ (Fishers exact test)
Conclusive versus inconclusive exams	BMI	0.654	Mann-Whitney-U-Test
Conclusive versus inconclusive exams	gender		
Conclusive versus inconclusive exams	age	0.953	Mann-Whitney-U-Test
Conclusive versus inconclusive exams	temperature	0.662	Mann-Whitney-U-Test
Conclusive versus inconclusive exams	CRP	0.758	Mann-Whitney-U-Test
Conclusive versus inconclusive exams	Leucocyte count	0.926	Mann-Whitney-U-Test
Conclusive versus inconclusive exams	Alvarado score	0.616	Mann-Whitney-U-Test

**Supplemental table 1: statistical tests used and their findings.**