Protocol

Optimizing diagnostics to distinguish between complicated and uncomplicated appendicitis; a double-edged sword

Version 1.0, 25-10-2019

PI:

Prof. dr. M.A. Boermeester

Study coordinators:

Jochem Scheijmans & Wouter Bom

Protocol committee:

WJ Bom[1], JCG Scheijmans[1], S Ubels[2], AAW van Geloven[3], SL Gans[1], KMAJ Tytgat[4], CC van Rossem[5], J Stoker[6], WA Bemelman[1], C Rosman[2], MG Dijkgraaf[7], MA Boermeester[1]

1 Department of Surgery, Amsterdam UMC, location Academic Medical Center

2 Department of Surgery, Radboud University, Nijmegen

3 Department of Surgery, Tergooi Hospital, Hilversum

4 Department of Gatroenterology and Hepatology, Amsterdam UMC, location Academic Medical Center

5 Department of Surgery, Maasstad Hospital, Rotterdam

6 Department of Radiology, Amsterdam UMC, location Academic Medical Center

7 Clinical Research Unit, Amsterdam UMC, Location Academic Medical Center

Abstract

Objective

To optimize diagnostics in distinguishing uncomplicated from complicated appendicitis in adult patients.

Background

Current diagnostics have proven to be insufficiently accurate to discriminate between patients with uncomplicated and complicated appendicitis. RCT's studying non-surgical (antibiotics only) versus surgical treatment for uncomplicated appendicitis report that 16.9% of all patients diagnosed as having uncomplicated appendicitis, actually do have complicated appendicitis. Failure to diagnose complicated appendicitis will negatively influence outcome since complicated appendicitis should be treated by surgery first and foremost. In our unpublished preliminary data from a Dutch appendicitis SNAPSHOT database(1), we found that patients with complicated appendicitis benefit from timely surgery within 8 hours, reducing complications from 27.5% to 18.3%, whereas surgery for uncomplicated appendicitis can be delayed - if needed - safely up to 24 hours.

To improve the differentiation between uncomplicated and complicated appendicitis, we combined clinical findings, laboratory test results, and imaging features, and developed and internally validated the Scoring system for Appendicitis Severity (SAS). By using the SAS we were able to retrospectively rule out patients with a complicated appendicitis in 96% (negative predictive value) and detect 95% of patients with a complicated appendicitis (sensitivity). In the present study, we aim to provide external and prospective validation of SAS in a new cohort of patients with clinical suspicion of appendicitis, and optimize the score if needed.

Methods

The SAS will be validated with 708 patients and clinical data will be recorded to analyze whether it can distinguish between complicated or uncomplicated appendicitis. If not adequately sufficient, the SAS will be optimized and sequentially a second external validation will be performed in a new group of 255 patients if needed. Patient preferences for treatment will also be assessed.

Anticipated results

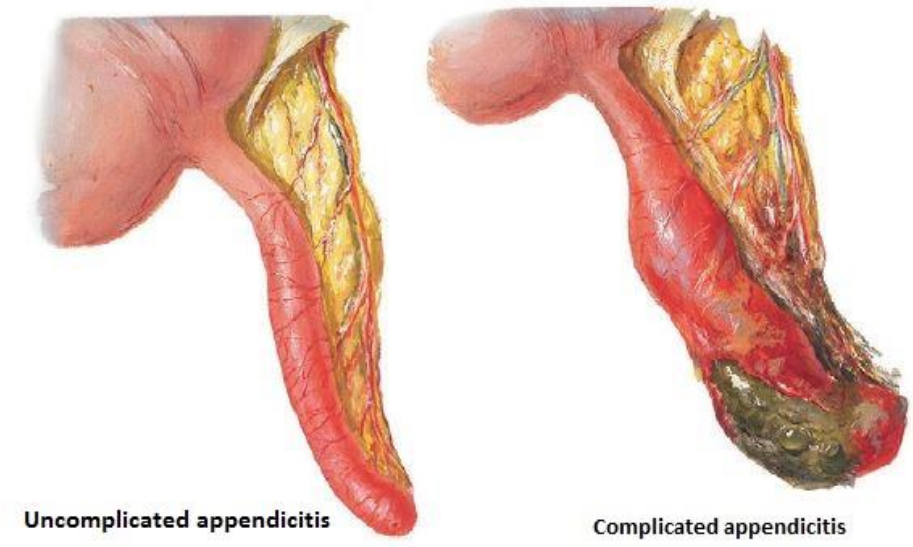
The SAS is able to achieve a NPV of 95% and a sensitivity of 95%, and is thereby sufficient to rule out complicated appendicitis.

The SAS could be helpful in quicker recognition of complicated appendicitis, timely surgery and thereby fewer complications. Moreover, it could facilitate the potential of non-surgical treatment for uncomplicated appendicitis.

Background

Appendicitis is one of the most common abdominal infectious diseases. It has an incidence of 77-89 per 100,000 cases and life-time prevalence as high as 9%.(2, 3) Since 1886 it is common to perform an appendectomy for appendicitis.(4) It was a longtime belief that an uncomplicated appendicitis could become a complicated (perforated) appendicitis and therefore the appendix should be removed. This theory most likely is not true and complicated and uncomplicated appendicitis may be two different entities (see figure 1).(5, 6) A recent meta-analysis also has shown that delayed surgery does not significantly affect the perforation or complication rate, which is in line with the concept of two different diseases instead of a continuum.(7) To date, an appendectomy is the standard therapy worldwide for acute appendicitis. In The Netherlands each year about 16,000 appendectomies are performed.(8) Although an appendectomy is common practice nowadays, it has a complication rate between 5-28%.(9)

Figure 1, Uncomplicated appendicitis vs Complicated appendicitis
(Source: http://cysurgery.com/wp-content/uploads/2018/02/appendicitis.jpg)



New insights have proven that conservative treatment with antibiotics in patients with truly uncomplicated appendicitis may be just as safe and effective as surgical treatment without the risk of surgical complications, just like management strategies of other colonic diseases such as diverticulitis or colitis. For this, it is mandatory to accurately identify patients with uncomplicated appendicitis. Major complications are seen in 8.4% of patients after appendectomy for uncomplicated appendicitis versus 4.9% after antibiotic treatment without appendectomy, focusing on results from RCTs only.(9) For minor complications, these percentages - in a meta-analysis of RCTs only - are 12.5% versus 2.2% in favor of antibiotic treatment.(9) These differences in complications between appendectomy and

antibiotic treatment were not significant, most likely due to the fact that 1 in 4 to 5 patients (17.9 – 27.3%) were wrongfully diagnosed with uncomplicated instead of complicated appendicitis, and should not have been included initially. (See Figure 2 and 3).

Figure 2, Percentage of complicated appendicitis in systematic review from Varadhan et al (BMJ, 2012)

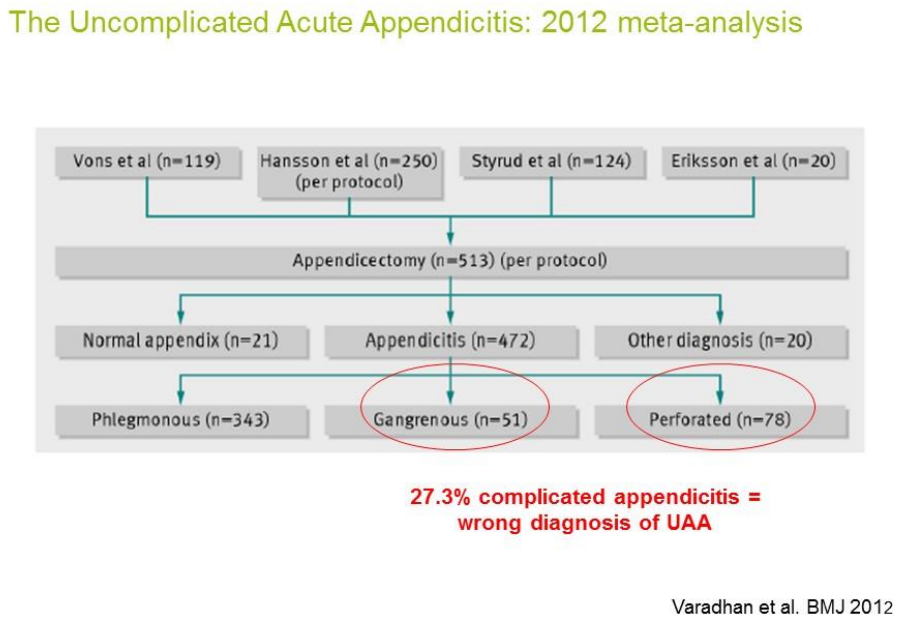
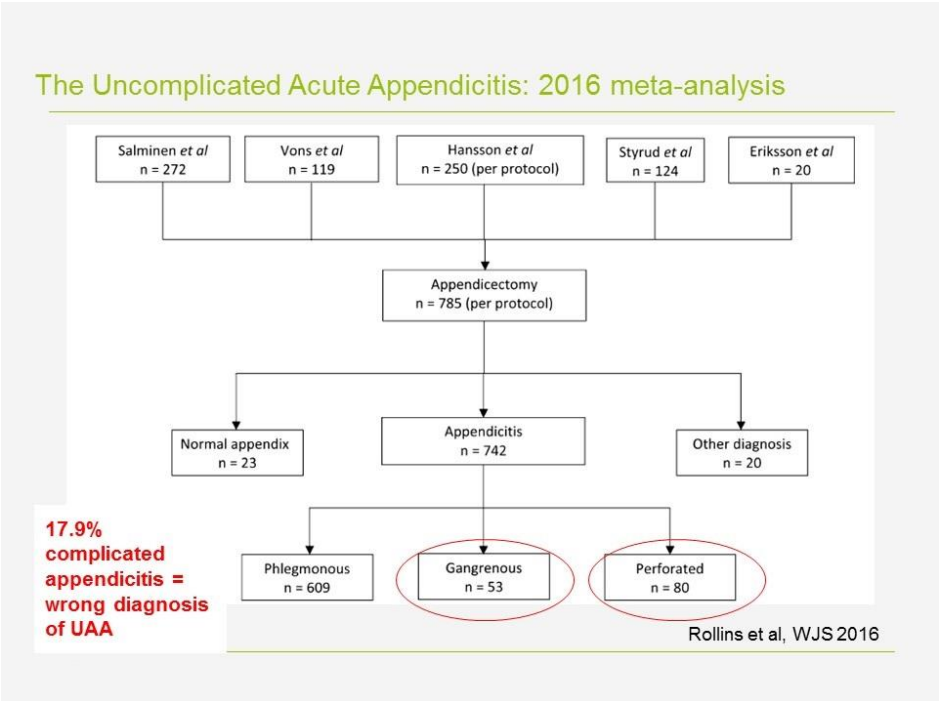
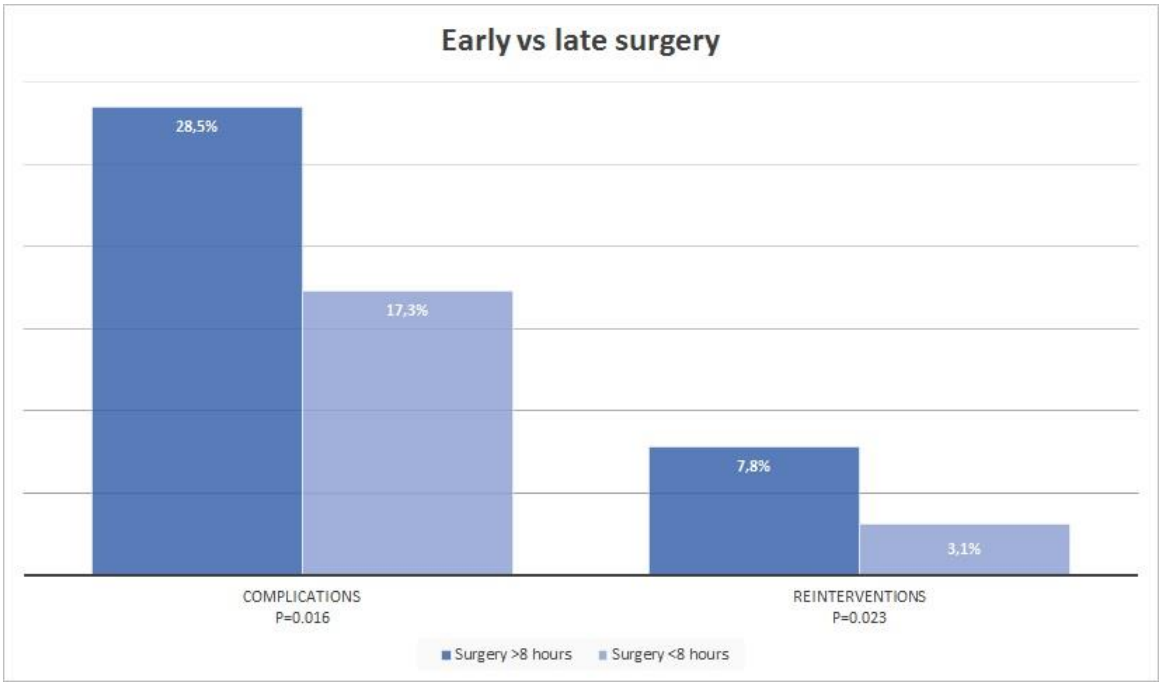


Figure 2, Percentage of complicated appendicitis in systematic review from Rollins et al (WJS, 2016)



In patients with *complicated appendicitis*, e.g. perforated or gangrenous appendicitis, appendectomy remains the golden standard and needs to be performed within 8 hours after diagnosis. A correct diagnosis of complicated appendicitis is important to prevent delayed surgery, which is associated with more postoperative complications. In the preparation for this present study proposal we performed a secondary data analysis using the Dutch SNAPSHOT appendicitis database; a previously published Dutch prospective cohort study comprising 1975 patients with the suspicion of appendicitis for which an appendectomy was planned.(1) We found an increase in complication rate if surgery was delayed for more than 8 hours after diagnosis in patients with complicated appendicitis. In these patients, 60 out of 218 (27.5%) had a postoperative complication or adverse event when surgery was delayed for more than 8 hours versus 47 out of 257 patients (18.3%) when surgery was performed within 8 hours (p=0.016). Delay by more than 8 hours was also associated with more surgical and/or radiological interventions (7.8% vs 3.1%, p=0.023) and longer hospital stay (5 vs 4 days, p=0.003). (See Figure 4). For *uncomplicated appendicitis* no significant increase in postoperative complications was seen in patients with delay of surgery beyond 8 hours after diagnosis (7.4 vs. 9.3xxx, p=0.31, surgery within and after 8 hours respectively).

Figure 3, Percentage of complications and reinterventions in surgery within or after 8 hours (preliminary data from the Dutch SNAPSHOT appendicitis study)



Thus, appropriate identification of complicated appendicitis is essential, but current diagnostic strategies have insufficient discriminatory value in distinguishing between patients with uncomplicated and complicated appendicitis.(10, 11) Several studies have analyzed the value of imaging modalities such as CT, ultrasound (US) or MRI for this discrimination. These studies all

demonstrate that the accuracy of current imaging modalities is good to excellent for the diagnosis of appendicitis in general, but *insufficient for the discrimination between complicated and uncomplicated appendicitis*. (11-17) Leeuwenburgh et al (11) have demonstrated in a prospective multicenter diagnostic accuracy study that both MRI and conditional CT (CT only after negative or inconclusive US) are unable to accurately discriminate between complicated and uncomplicated appendicitis. Both strategies have a low positive predictive value (68% for US with conditional CT and 57% for MRI) and a low negative predictive value (84% for US with conditional CT and 86% for MRI) for complicated appendicitis. Sensitivity for complicated appendicitis was 48% for US with conditional CT and 57% for MRI. In other words, 52% of all patients with complicated appendicitis were missed by US with conditional CT and 43% for MRI. (See figure 5)

Figure 4, accuracy of MRI and ultrasonography with conditional CT for perforated appendicitis by Leeuwenburgh et al (BJS, 2014)

	No. of patients	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Ultrasound imaging with conditional CT*	125	48 (32, 65) 15 of 31	93 (85, 96) 87 of 94	68 (47, 84) 15 of 22	84 (76, 90) 87 of 103
MRI	120	57 (39, 73) 17 of 30	86 (77, 91) 77 of 90	57 (39, 73) 17 of 30	86 (77, 91) 77 of 90
P†		0·517	0·127	0·399	0·833

Values in parentheses are 95 per cent confidence intervals; proportions used to calculate diagnostic indices are also shown. Only imaging results positive for perforated or simple appendicitis were used for these calculations. *All patients underwent initial ultrasonography, then computed tomography (CT) if the ultrasound results were negative or inconclusive. PPV, positive predictive value; NPV, negative predictive value; MRI, magnetic resonance imaging. †χ^2 test.

Three recent meta-analyses of studies comparing outcome of appendectomy versus antibiotics only in patients diagnosed with uncomplicated appendicitis have demonstrated that up to 27% of patients included in these studies had complicated instead of uncomplicated appendicitis at surgery and were therefore wrongfully included.(9, 18, 19) The trial contamination of inclusion of patients with initial complicated appendicitis likely has skewed trial results in a negative direction and may have underestimated the potential benefit of antibiotics only treatment. Although initial failures of antibiotic treatment (within 1 month) were below 10%, within 1 year about 22% of patients still end with an appendectomy. Better selection of patients with truly uncomplicated appendicitis is needed to see the merits of antibiotic only treatment as a replacement of appendectomy in simple appendicitis.

From a prospective database a scoring system (Scorings system of Appendicitis Severity, SAS) has been developed, based on clinical features combined with imaging (US or CT) features.(10) Two scoring systems were developed: SAS-US and SAS-CT. SAS has been internally validated and is able to rule out patients with a complicated appendicitis in 96.3% (NPV) and detect 94.6% of patients with a complicated appendicitis (sensitivity).(10)

SAS may act as a double-edged sword. By better triage for rapid surgery, SAS can potentially decrease the number of complications in patients diagnosed with a complicated appendicitis. Secondly, by improving accuracy for the diagnosis *uncomplicated appendicitis* SAS can aid in increasing the role for antibiotics-only treatment in patients with true uncomplicated appendicitis versus a traditional appendectomy.

Appendectomy is common practice for a frequently occurring disease in The Netherlands. It is important to acknowledge patient preferences for eventual concerns of non-operative and operative treatment. Three recently published studies reported a wide range in proportions of patients preferring non-operative treatment; 9.4%-57%. (20-22). In order to facilitate shared decision making for uncomplicated appendicitis in the Dutch population, we will also assess the Dutch patient preference in this study.

Scoring system of Appendicitis Severity (SAS)

In 2015 we developed the SAS to distinguish between patients with complicated and uncomplicated appendicitis based on clinical parameters combined with imaging features.(10) Patients suspected of appendicitis were retrospectively selected from two previously published prospective databases. (11, 16) A total of 395 patients with suspected appendicitis, based on clinical parameters and imaging, were included. Of these patients 27.8% had complicated appendicitis, 60.5% uncomplicated appendicitis and 11.6% had an alternative disease. Based on these results two appendicitis scoring systems were made; one combining clinical and ultrasonography features (appendicitis US scoring system, SAS-US) and one combining clinical and CT features (appendicitis CT scoring system, SAS-CT) (See Figure 6).

Figure 5, The Scoring system of Appendicitis Severity (SAS) by Atema et al (BJS, 2015)

	Points scored	
	Clinical and CT features	Clinical and ultrasound features
Age ≥ 45 years	2	2
Body temperature		
≤ 37·0	0	0
37·1–37·9	2	2
≥ 38·0	4	4
Duration of symptoms ≥ 48 h	2	2
WBC count > 13×10^9/l	2	2
C-reactive protein (mg/l)		
≤ 50	0	0
51–100	2	4
> 100	3	5
Extraluminal free air on imaging	5	–
Periappendiceal fluid on imaging	2	2
Appendicolith on imaging	2	2
Maximum score	22	19

WBC, white blood cell.

The SAS-US has a maximum possible score of 19 points and consists of seven items; age, body temperature, duration of symptoms, WBC count, CRP level, and periappendiceal fluid and presence of an appendicolith on ultrasound examination. (Figure 6)

The SAS-CT has a maximum possible score of 22 points and consists of eight items; age, body temperature, duration of symptoms, WBC count, CRP level, and extraluminal free air, periappendiceal fluid and presence of an appendicolith on CT. (Figure 6)

The SAS-US reaches a sensitivity of 96.6%, a specificity of 45.7%, a positive predictive value of 41.6%. With a *cut-off score of 5 or less*, only 2.9% had complicated appendicitis, resulting in a negative predictive value (NPV) of 97.1% for complicated appendicitis. (Figure 7) (10)

The SAS-CT reaches a sensitivity of 90.2%, specificity of 70.3%, positive predictive value of 55.2%. With a *cut-off score of 6 points or less*, only 5.3% of patients had complicated appendicitis, resulting in a NPV of 94.7% for complicated appendicitis. (Figure 7) (10)

SAS has a high accuracy for the discrimination between complicated and uncomplicated appendicitis. (See Figure 7)

Figure 6, Diagnostic accuracy parameters of SAS for US en CT by Atema et al (BJS, 2015)

	US-SAS	*CT-SAS*
Sensitivity	96.6%	90.2%
Specificity	45.7%	70.3%
Positive predictive value	41.6%	55.2%
Negative predictive value	97.1%	94.7%

Based on the ratio US and CT from the Dutch appendicitis SNAPSHOT study we calculated the target in this present study, resulting in a target sensitivity of 95% and a NPV of 95%.

As described above, a secondary analysis of the data from the Dutch SNAPSHOT appendicitis database showed an increase in complication rate if surgery was delayed for more than 8 hours in patients with complicated appendicitis (27.5% versus 18.3%, p=0.016). Immediate recognition of a complicated versus an uncomplicated appendicitis is not only needed for correct and prompt treatment, but may also facilitate antibiotic instead of surgical treatment in patients with uncomplicated appendicitis. The Scoring system of Appendicitis Severity (SAS) offers that opportunity, but is in need of prospective external validation in a newly recruited cohort of patients diagnosed with appendicitis.

Methods

Design

A multicenter prospective cohort study of patients with suspicion of acute appendicitis after diagnostic imaging will be conducted in hospitals within our consortium of Dutch hospitals for studies on acute abdominal pain.

Study population

For the validation and optimization (if needed) of the SAS, at first 708 patients with an imaging-confirmed diagnosis of acute appendicitis will be included (scenario A). If the validation results do not reach the target 95% sensitivity and 95% NPV, optimization of SAS will be performed. Next, if optimization is needed (scenario B), the optimized SAS will be externally validated. For this second validation another 255 patients are needed (see sample size calculations). Therefore, a minimum of 708 patients (scenario A: no optimization of SAS needed) and a maximum of 963 patients (scenario B: optimization of SAS needed) will be included.

Inclusion criteria:

Patients aged ≥18 years with a US and/or CT imaging-confirmed or -suspected diagnosis of acute appendicitis and who undergo surgery with the intention to perform an appendectomy will be included.

Exclusion criteria:

- The few patients that may be diagnosed only by MRI or ultrasound with conditional MRI will not be included, as SAS is not developed for MRI.
- When the ultrasonography was inconclusive or negative, SAS-US cannot be calculated, and diagnostics need to progress to CT. In some cases the decision is made to follow-up disease course instead of having an immediate CT. Present study does not dictate type of imaging, but follows clinical practice (see 'Intervention').
- Patients who receive antibiotics without surgery because of an appendicular infiltrate will not be included, as for the present study a surgical specimen is needed as a reference standard.
- Patients undergoing surgery 48 hours after diagnosis (last performed imaging) will not be included, since we consider that after 48 hours the pre-operative diagnostic results and thereby the associated SAS are not representative for the intraoperative diagnosis.

Intervention

Diagnostic work-up will be performed according to the current standards (Dutch guidelines). Additionally, the treating physician at the emergency department (ED) will be asked to complete the data for the appendicitis scoring system SAS, without actually being aware of the scoring system itself. This will be executed by filling in a standard report in the patient file specified to this patient.

As the first step of the standard report, the physician will be asked whether to classify the patient's clinical presentation as 'complicated' or 'uncomplicated' appendicitis, based on clinical, biochemical and imaging results. Next, they will enter the individual scoring parameters. We do not hand a scoring card to the treating physicians, nor will the final score be shown in the acute setting. The patient will be asked to participate and to fill in the informed consent. After permission is given, the patient data will be collected from the patient file into the data collection program CASTOR. When importing in CASTOR, the data is pseudoanonymised. Thereby, the data in CASTOR is not reducible to the patients.

Both ultrasonography (US) as well as computed tomography (CT) are widely accepted in the diagnostic work-up of patients with acute appendicitis. A positive US diagnosis of appendicitis is as reliable as a positive CT diagnosis(17); that is why two variants of the appendicitis scoring system (using either US or CT) have been developed (see figure 6). Points are scored for age ≥45 years, body temperature (37.1 ^{o}C -37.9^{o}C and ≥38^{o}C), duration of symptoms ≥48 hours, White Blood Cell count >13x109/L, CRP level (>50mg/L and >100mg/L), periappendiceal fluid on imaging and appendicolith on imaging and for CT also extraluminal air. Standardized method of reporting will be used by the radiologists. With a cut-off score of 5 points for SAS-US and 6 points for SAS-CT, patients will be classified as *complicated appendicitis* (SAS-US > 5 points or SAS-CT > 6 points) or *uncomplicated (simple) appendicitis* (SAS-US ≤ 5 points or SAS-CT ≤ 6 points).

As a negative US has a high false-negative rate, those patients undergo a (conditional) CT, according to the Dutch Appendicitis Guideline and Dutch Guideline for Diagnostics of Acute Abdominal Pain. (8, 23) Patients who undergo both US and CT will be scored using SAS-CT.

The treating physician will be blinded for points given for every entered single parameter and the total SAS score. The outcome of SAS will not be used for treatment choices, as this is a validation study.

Reference standard

For both US and CT, the following parameters will be determined and reported in a standardized way: visualisation of the entire appendix, appendiceal diameter, presence of periappendiceal fat infiltration, periappendiceal fluid, appendicolith, destruction of appendiceal wall (perforation), abscesses. For CT, presence of extraluminal air will be scored as well. The radiologist will also make a judgment whether the appendicitis is complicated or not. The standardized radiology reports also register the level of certainty of the final severity diagnosis 'uncomplicated appendicitis' or 'complicated appendicitis'.

Pathology reports and operative notes also will be standardized (standardized checklist; CRF), and surgeon or pathologist will be asked to indicate whether they classify the appendicitis to be complicated or uncomplicated.

In this validation study, all patients diagnosed with appendicitis, whether uncomplicated or complicated, will undergo an appendectomy according to present guidelines.

To assess the accuracy of SAS in discriminating between uncomplicated and complicated appendicitis this diagnosis will be compared to a **reference standard**. This reference standard is the final overall diagnosis including 3 months of follow-up: either confirmation of the diagnosis 'acute appendicitis', or establishment of an alternative diagnosis; for those with the diagnosis of acute appendicitis a final diagnosis of *'uncomplicated appendicitis'* or *'complicated appendicitis'* is assigned.

Phlegmonous inflammation of the appendix is diagnosed as '*uncomplicated appendicitis*'. This only applies when no signs of gangrene or necrosis are seen microscopically and macroscopically. Transmural inflammation or ulceration can occur in this diagnosis. (5)

The *'complicated* appendicitis' contains severe inflammation of the appendix with signs of gangrene or if a perforation is seen. (5) Microperforations described in histopathologic report are only diagnosed as *'complicated appendicitis'* if the surgeon treats the patient with antibiotics due to intraoperative findings.

This reference standard will be established by the consensus of an expert panel, consisting of 2 surgeons, 2 radiologists, 1 pathologist and 1 ED physician / surgical resident who will review a structured summary of clinical information during admission, operative notes, pathology report, imaging findings and CRFs from surgeon and pathologist.

A picture from the appendix intraoperatively will be stored in the medical record and eventually saved in CASTOR. One picture will be made before removing the appendix, thereafter a picture of the

specimen on a white background will be made by the clinical assistant or the surgeon after removal of the appendix. These pictures can be used to come to a consensus in cases of doubt of the diagnosis.

Patients will have follow-up for 3 months. Follow-up data will be extracted from the medical records. In case of disagreement among expert panel members a final diagnosis will be assigned during a consensus meeting of the expert panel concerning the disagreement cases.

Primary outcome

The *sensitivity* and *negative predictive value* of the Scoring system of Appendicitis Severity (SAS) for excluding complicated appendicitis.

Secondary outcomes

- Specificity and positive predictive value of SAS for excluding complicated appendicitis.
- Sensitivity, specificity, NPV, PPV for excluding complicated appendicitis for SAS-US and SAS-CT separately.
- The discriminatory capacity of SAS, SAS-US and SAS-CT by calculating the area under the curve, indicating how well the model distinguishes patients with complicated appendicitis from uncomplicated appendicitis.
- Optimization of the model and scoring system. The relative weight of each variable will be analyzed with a multivariable regression analysis. All predictive values used in the model will be expressed in odds ratio's. For details see Data-analysis section.
- Patient reported preferred treatment (see PROM paragraph below).
- Number of complications for both complicated and uncomplicated appendicitis, and will be stratified per time period 'surgery ≤ 8 after ED admission' and 'surgery > 8 hours after ED admission'.
- Multivariable regression analyses will be performed for possible other risk factors for complicated appendicitis. This will include additional clinical features and radiological features.
- Diagnostic accuracy and discriminatory capacity of the physician at ED in distinguishing a complicated from an uncomplicated appendicitis compared to SAS.

A standardized patient-reported outcome measure (PROM) questionnaire will be disseminated at the end of 3-months follow-up to explore the patients' experience with and perception of the treatment they underwent and their expectations if their treatment had been non-operative. Based on different structured scenarios displaying various failure rates of conservative treatment and various complication rates of appendectomy, we will ask them to give their preference for non-surgical or

surgical treatment in uncomplicated disease and to substantiate their choice by checking a list of prespecified arguments.

All complications for three months of follow-up will be scored using the Clavien-Dindo classification. Patient files will be hand searched. In addition, patients will be asked through the end-of-follow-up questionnaires whether they experienced complications or not and/or consulted a physician because of suspicion of complications.

Hypothesis

It is hypothesized that the Scoring system of Appendicitis Severity (SAS) reaches a negative predictive value of at least 95% and a sensitivity of 95% for excluding complicated appendicitis.

Sample size calculation

Based on the data from the appendicitis SNAPSHOT study we know that for patients who underwent CT or US 68.2% of patients will undergo US only, 7.9% CT primarily and 24.0% will undergo US followed by CT. The SAS previously reached a sensitivity of 96.6 for US and 90.2% for CT. (10) An average sensitivity of 95% is targeted. For the negative predictive value, the SAS-US reaches 97.1% and SAS-CT 94.7%. Therefore, a NPV of at least 95% is targeted.

Validation period

We target a sensitivity of 95% and a NPV of 95%, and consider a lower limit of 3 percent as the only limit of the corresponding one-sided 97.5% confidence interval (CI) as the bare minimum. When the number of complicated appendicitis equals 203, the one-sided 97.5% CI will extend 3% from the observed percentage for an expected percentage of at least 95% for sensitivity. Given a prevalence of 28.7% complicated cases in the target population, about 708 patients need to be included initially to reach a minimum of 203 patients with complicated appendicitis among all included appendicitis. With the same extend of 3% from the observed percentage for an expected percentage of at least 95% for NPV, a total of 511 patients with appendicitis is needed.

To report both reliable sensitivity and reliable NPV, we need to include the highest number of those two. After including at least 203 patients with complicated appendicitis to achieve the target sensitivity and at least 203 patients with a SAS score predicting an uncomplicated appendicitis to achieve the target NPV, we will test our hypothesis and validate the SAS. The expected required number of patients is 708 for this validation cohort.

Optimization period

If the validation does not reach the targeted sensitivity or NPV of the point estimate of at least 92%, optimization of the SAS will be executed. Optimization will be performed by using the data of the 708 patients. To external validate the optimized SAS, we need a new cohort of patients. We will include these patients after including the primary 708 patients. We again intend to achieve a sensitivity of 95%, now with a lower limit of 5% as the only limit of the corresponding one-sided 97.5% CI as the bare minimum. Because of the large cohort in which the SAS has been validated, we consider that the lower limit of 5% instead of 3% will suffice. We calculated that 255 patients are needed for this second validation cohort, in case of need for optimization of SAS.

A maximum of 963 patients will be included. This total consists of the primary cohort of the validation/optimization cohort of 708 patients and, if needed, a second external validation cohort of 255 patients.

Data collection

Data from the following resources will be collected.

CRF's

Scores from the SAS will be collected in CASTOR, just as the CRF's for surgeon and pathologist. As described previously, the individual parameters of SAS will be entered in the online CRF by the physician at the ED preoperatively. Imaging results will be descripted according to a standardized radiology report. These reports will be interpreted by the physician at the ED. CRF's will be send to the surgeon and pathologist shortly after the operation.

Medical records

Perioperative data will be extracted by automatic extraction from the Electronic Patient Files. A SQL code will be constructed in a collaboration with a Business Intelligence Specialist. For variables which are not able to extract automatically, medical records will be hand searched. Data will be pseudoanonymised and saved in CASTOR.

Questionnaires

Questionnaires will be distributed by e-mail via CASTOR or, if the patients prefer, paper versions by post. Answers will be digitalized and stored within CASTOR.

Data analysis

Primary outcomes and validation

Scores will be calculated as described above. With a cut-off score of 5 points for SAS-US and 6 points for SAS-CT, patients will be classified as complicated (SAS-US > 5 points or SAS-CT > 6 points) or uncomplicated (SAS-US ≤ 5 points or SAS-CT ≤ 6 points). Contingency tables will be constructed for SAS-US and SAS-CT. Sensitivity, specificity, PPV and NPV will be calculated. The area under the curve of the SAS will be plotted and calculated in a receiver operating characteristic (ROC) curve. A similar analysis will be performed for SAS-US and SAS-CT separately.

Secondary outcomes

Questionnaires will be analyzed and the number and percentage of patients choosing for antibiotic treatment, surgery or patients without a preference will be displayed. The most influencing arguments for this choice will be presented.

The number of patients with a complication will be calculated and will be stratified per time period (8 hours within ED admission and 8 hours after ED admission).

Possible other risk factors for complicated appendicitis than already used in the SAS will be searched through univariable regression analyses.

The diagnostic accuracy and discriminatory capacity of the physician at ED in distinguishing a complicated from an uncomplicated appendicitis will be compared to SAS by calculating sensitivity and specificity for both 'tests'. The significance of the differences will be calculated by the chi-square test.

Optimization

In case the sensitivity or negative predictive values are below 92%, we will continue with an optimization analysis. We will perform the optimization of SAS using the data of the first cohort of 708 patients. For external validation of this cohort, another 255 patients will be included (see Sample Size Calculation). Data from these patients will be collected in the same manner as in the validation period. For the optimization, possible variables collected from the CRF's and medical reports will be used. The variables which are included in the SAS will be reevaluated or rescaled. Additionally, literature will be searched for other predictors of complicated appendicitis. If available and applicable, those variables will be added into the optimized scoring model.

Continuous variables will be categorized. An optimal cut-off score will be chosen with the use of restricted cubic spline functions. A multivariable logistic binary regression model with the before mentioned predictors, including the parameters used in the SAS, will be constructed and reduced

with backwards selection, excluding parameters with a p value <0.15. The model will be transformed into a clinical applicable scoring system, multiplying the adjusted coefficient of each parameters and rounding it to the nearest integer. Total scores for every patient will be calculated and a cut-off analysis will be performed to select patients with predicted complicated appendicitis, not exceeding 5% of false negatives.

The optimized model will be performed for both US and CT. Diagnostic accuracy measures will be calculated and the score will be externally validated on our second cohort of 255 patients.

Patient recruitment

Patients will be recruited 24 hours a day. Informed consent will be obtained at the ED or at the ward, both pre- and postoperatively. Information about the study will be given and questions will be answered by a medical doctor before signing the informed consent. If a patient leaves the hospital before informed consent was obtained, informed consent will be obtained by letter or e-mail. In consultation with our juridical department this is in line with the design of this study, due to the fact that this study does not subject to the Medical Research Involving Human Subjects Act (WMO).

Intervention and risk

The Scoring system of Appendicitis Severity (SAS) can be applied without adding diagnostics other than standard diagnostic work-up protocols. This is a purely diagnostic study without direct management consequences for the included patients. Participants will receive diagnostics and treatment according to current standards and there is no additional burden except for a single time point patient-reported outcome and preferences questionnaire. Participation will not result in any risks for the patients.

Compensation

No financial compensation will be provided. There is no indication for travel allowance.

Patient privacy

From the electronic patient file only relevant data will be collected, such as patient characteristics and primary and secondary outcomes specified as above. These data will be encrypted. The encrypted data will be stored in a private storage, only available for involved researchers. The encryption code will be secured by a password and is accessible for the (local)head researcher only. The patient questionnaires will be anonymous and will only be marked by the study number. If available, pictures of the appendices intraoperatively will be collected too. The filing of these pictures

with a unique study number will not be traceable to the patient. Data will be collected in multiple centers and will be shared after encryption via data collection program Castor, after pseudoanonymisation. The data will be stored for 15 years. After this period the data will be destroyed. When patients give their permission, the data could be used in other subject related studies for a longer period. The collection of patient data will be reported to the local privacy officer. Data are open for re-use for research in the topic of appendicitis.

Publication and implementation

The results will be published in an international peer reviewed journal. They will also be disseminated through international conferences, (inter)national guidelines, and will be the foundation for further research and a change in practice. Data will be open for re-use after publication of our results.

After completion of the study the national guideline can be adjusted according to the findings of this study. If SAS shows to be accurate enough to differentiate uncomplicated from complicated appendicitis, non-surgical treatment is likely more effective than published results have shown to date. A new RCT comparing appendectomy with antibiotic treatment using this more accurate way to select uncomplicated appendicitis may be needed to see the actual potential of non-surgical treatment of true uncomplicated appendicitis.

If the SAS is implemented in the guidelines, it will be easier to stimulate its use. Moreover, a web-based application or app could simply aid any doctor involved in diagnosis and treatment of patients with acute abdominal pain and the suspicion of acute appendicitis. Pocket maps can be produced to disseminate the use of SAS.

Funding

The *'Maag Lever Darm-stichting'* provided funding for this study by means of the *'Slimme en snelle diagnostiek'* grant (no. D18-05).

References

1. van Rossem CC, Bolmers MD, Schreinemacher MH, van Geloven AA, Bemelman WA. Prospective nationwide outcome audit of surgery for suspected acute appendicitis. The British journal of surgery. 2016;103(1):144-51.
2. Anderson JE, Bickler SW, Chang DC, Talamini MA. Examining a common disease with unknown etiology: trends in epidemiology and surgical management of appendicitis in California, 1995-2009. World journal of surgery. 2012;36(12):2787-94.
3. Ceresoli M, Zucchi A, Allievi N, Harbi A, Pisano M, Montori G, et al. Acute appendicitis: Epidemiology, treatment and outcomes- analysis of 16544 consecutive cases. World journal of gastrointestinal surgery. 2016;8(10):693-9.
4. Fitz RH. Perforating inflammation of the vermiform appendix: with special reference to its early diagnosis and treatment. Am J Med Sci. 1886;92:321-46.
5. Bhangu A. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. Lancet (London, England). 2015;386.
6. Andersson R, Hugander A, Thulin A, Nystrom PO, Olaison G. Indications for operation in suspected appendicitis and incidence of perforation. BMJ (Clinical research ed). 1994;308(6921):107-10.
7. van Dijk ST, van Dijk AH, Dijkgraaf MG, Boermeester MA. Meta-analysis of in-hospital delay before surgery as a risk factor for complications in patients with acute appendicitis. The British journal of surgery. 2018;105(8):933-45.
8. NVvH. Richtlijn Acute appendicitis [Guideline]. 2019 [Available from: https://richtlijnendatabase.nl/richtlijn/acute_appendicitis/startpagina_-_acute_appendicitis.html.
9. Sallinen V, Akl EA, You JJ, Agarwal A, Shoucair S, Vandvik PO, et al. Meta-analysis of antibiotics versus appendicectomy for non-perforated acute appendicitis. The British journal of surgery. 2016;103(6):656-67.
10. Atema JJ, van Rossem CC, Leeuwenburgh MM, Stoker J, Boermeester MA. Scoring system to distinguish uncomplicated from complicated acute appendicitis. The British journal of surgery. 2015;102(8):979-90.
11. Leeuwenburgh MM, Wiezer MJ, Wiarda BM, Bouma WH, Phoa SS, Stockmann HB, et al. Accuracy of MRI compared with ultrasound imaging and selective use of CT to discriminate simple from perforated appendicitis. The British journal of surgery. 2014;101(1):e147-55.
12. Bixby SD, Lucey BC, Soto JA, Theysohn JM, Ozonoff A, Varghese JC. Perforated versus nonperforated acute appendicitis: accuracy of multidetector CT detection. Radiology. 2006;241(3):780-6.
13. Doria AS, Moineddin R, Kellenberger CJ, Epelman M, Beyene J, Schuh S, et al. US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. Radiology. 2006;241(1):83-94.
14. Foley TA, Earnest Ft, Nathan MA, Hough DM, Schiller HJ, Hoskin TL. Differentiation of nonperforated from perforated appendicitis: accuracy of CT diagnosis and relationship of CT findings to length of hospital stay. Radiology. 2005;235(1):89-96.
15. Fraser JD, Aguayo P, Sharp SW, Snyder CL, Rivard DC, Cully BE, et al. Accuracy of computed tomography in predicting appendiceal perforation. Journal of pediatric surgery. 2010;45(1):231-4; discussion 4-4.

16. Lameris W, van Randen A, van Es HW, van Heesewijk JP, van Ramshorst B, Bouma WH, et al. Imaging strategies for detection of urgent conditions in patients with acute abdominal pain: diagnostic accuracy study. BMJ (Clinical research ed). 2009;338:b2431.

17. van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester MA. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. Radiology. 2008;249(1):97-106.

18. Rollins KE, Varadhan KK, Neal KR, Lobo DN. Antibiotics Versus Appendicectomy for the Treatment of Uncomplicated Acute Appendicitis: An Updated Meta-Analysis of Randomised Controlled Trials. World journal of surgery. 2016;40(10):2305-18.

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

20. Althans AR, Tamer P, Brady JT, Steinhagen E, Ho VP. Surgery versus Antibiotics for Uncomplicated Appendicitis: Which Would a Medical Student Want? Surgical infections. 2017;18(8):868-73.

21. Hanson AL, Crosby RD, Basson MD. Patient Preferences for Surgery or Antibiotics for the Treatment of Acute Appendicitis. JAMA Surg. 2018.

22. Kadera SP, Mower WR, Krishnadasan A, Talan DA. Patient perspectives on antibiotics for appendicitis at one hospital. The Journal of surgical research. 2016;201(2):253-7.

23. NVvH. Richtlijn Diagnostiek acute buikpijn bij volwassenen [Guideline]. 2013 [Available from: https://heelkunde.nl/sites/heelkunde.nl/files/richtlijnen-definitief/Diagnostiek-bij-acute-buikpijn---definitieve-versie.pdf.