

Wytze Laméris
Adrienne van Randen
Shandra Bipat
Patrick M. M. Bossuyt
Marja A. Boermeester
Jaap Stoker

Graded compression ultrasonography and computed tomography in acute colonic diverticulitis: Meta-analysis of test accuracy

Received: 30 November 2007
Accepted: 22 March 2008
Published online: 4 June 2008
© The Author(s) 2008

W. Laméris (✉)
Department of Surgery
(suite G4–130), Academic
Medical Center,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands
e-mail: W.Lameris@amc.uva.nl
Tel.: +31-205-663405
Fax: +31-205-69243

The Dutch Organisation for Health Research and Development, Health Care Efficiency Research programme, funded the study (ZonMw, grant number 945-04-308).

W. Laméris · M. A. Boermeester
Department of Surgery,
Academic Medical Center,
University of Amsterdam,
Amsterdam, The Netherlands

A. van Randen · S. Bipat · J. Stoker
Department of Radiology,
Academic Medical Center,
University of Amsterdam,
Amsterdam, The Netherlands

P. M. M. Bossuyt
Department of Clinical Epidemiology,
Biostatistics and Bioinformatics,
Academic Medical Center,
University of Amsterdam,
Amsterdam, The Netherlands

Abstract The purpose was to investigate the diagnostic accuracy of graded compression ultrasonography (US) and computed tomography (CT) in diagnosing acute colonic diverticulitis (ACD) in suspected patients. We performed a systematic review and meta-analysis of the accuracy of CT and US in diagnosing ACD. Study quality was assessed with the QUADAS tool. Summary estimates of sensitivity and specificity were calculated using a bivariate random effects model. Six US studies evaluated 630

patients, and eight CT studies evaluated 684 patients. Overall, their quality was moderate. We did not identify meaningful sources of heterogeneity in the study results. Summary sensitivity estimates were 92% (95% CI: 80%–97%) for US versus 94% (95% CI: 87%–97%) for CT ($p=0.65$). Summary specificity estimates were 90% (95% CI: 82%–95%) for US versus 99% (95% CI: 90%–100%) for CT ($p=0.07$). For the identification of alternative diseases sensitivity ranged between 33% and 78% for US and between 50% and 100% for CT. The currently best available evidence shows no statistically significant difference in accuracy of US and CT in diagnosing ACD. Therefore, both US and CT can be used as initial diagnostic tool until new evidence is brought forward. However, CT is more likely to identify alternative diseases.

Keywords Diverticulitis · Ultrasonography · Tomography · X-ray computed · Sensitivity and specificity · Meta-analysis

Introduction

In the Western elderly population acute colonic diverticulitis (ACD) is a common disease of the gastrointestinal tract. The prevalence of diverticulosis, the underlying pathological condition, ranges from 10% in people under 40 years to as high as 60% in people older

than 80 years. Approximately 10% to 20% of affected people will develop one or more episodes of ACD [1, 2].

A widely shared view nowadays is that imaging is mandatory in the initial assessment of patients suspected of ACD [3–5] to cope with clinical misdiagnosis rates, the discrepancy between clinical presentation and the extent of

ACD, and the possibility that other diseases mimicking ACD are missed.

Ultrasonography (US) and computed tomography (CT) are used in daily practice to complement clinical assessment and physical examination in diagnosing ACD. Those in favour of US stress its lower costs, wider availability, and the lack of radiation exposure and use of contrast material. CT imaging is embraced by others because they claim CT is less operator dependent than US in obtaining high diagnostic accuracy, generates fewer inconclusive results, and is able to assist in surgical planning when intervention is needed [2, 5, 6].

Reported sensitivities and specificities vary, both for US and CT [5, 7]. A systematic review of prospective studies may be able to summarise the diagnostic accuracy of both investigations, providing estimates with greater precision. Ideally, this analysis would merely include studies investigating the diagnostic accuracy of US and CT in the same population (head-to-head comparison). Since such comparative studies are scarce, we performed a systematic review and meta-analysis of prospective comparative studies, as well as prospective studies investigating US or CT separately.

Methods

Search strategy and study eligibility

We performed a literature search to identify studies investigating the diagnostic accuracy of US and CT in human subjects suspected of ACD. We searched MEDLINE and EMBASE databases for papers published between January 1966 and January 2007, using the following keywords: ["Diverticulitis"(MeSH) OR "Diverticulitis, Colonic"(MeSH)] AND ["Radiography"[MeSH] OR "radiography"(Subheading) OR "Radiography, Thoracic"(MeSH) OR "Radiography, Abdominal"(MeSH) OR "Tomography, X-Ray Computed"(MeSH) OR "Tomography Scanners, X-Ray Computed"(MeSH) OR "Tomography, Spiral Computed"(MeSH) OR "Ultrasonography" (MeSH) OR "ultrasonography"(Subheading)].

CINAHL database was also checked for relevant studies with the following keywords: [diverticulitis (MeSH) and (Ultrasonography (MeSH) or Echography (MeSH) or Radiography (MeSH) or Computed tomography (MeSH) or Computer-Assisted Tomography (MeSH)]. The Cochrane database of Systematic Reviews was searched with the following words: Diverticulitis AND (ultrasonography OR computed tomography).

Studies were eligible if they addressed the diagnostic accuracy of US, CT, or both, in patients with suspected ACD. No age, date or language restrictions were applied. If studies were judged potentially eligible, full-text versions of the papers in which they had been reported were retrieved. We crosschecked the references.

Study selection

Two reviewers (WL and AvR) independently evaluated the obtained literature for relevance. Studies were included if they met the following criteria: (1) prospective (data collection) study design; (2) CT and/or US criteria for the presence of diverticulitis were given; (3) graded compression US was performed; (4) reference standard was defined; (5) diverticulitis was located in the large bowel; (6) the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) was reported or could be extracted from the study report.

Study and patient characteristics

Two reviewers independently evaluated the included studies and extracted the data for each included study. Disagreement between the reviewers was solved by discussion among all authors. Data on study design and patient group, technical specifications, and diagnostic accuracy of CT and US were collected using a standardised case record form for each included study.

Study design characteristics The 'Quality assessment for diagnostic accuracy studies' (QUADAS) tool was used for evaluation of study quality [8]. In addition to the inclusion criteria the following characteristics were recorded: (1) department of the first author; (2) design of the study (single- or multicentre); (3) description of patient population, including sample size, age, male-female distribution, the prevalence of ACD and complicated ACD, and study setting (hospitalised patients, outpatients or in-hospital referrals to the radiology department); (4) if US and CT results were independently obtained (head-to-head comparative studies); (5) if US and CT were interpreted independently from clinical information; (6) if a description of US and CT criteria for the presence of acute diverticulitis was given; (7) experience of observers; (8) time interval between US and CT (head-to-head comparative studies); (9) time interval between imaging and reference standard; (10) if interpretation of reference standard was done without information on US and CT findings; (11) if the execution of the reference standard was described. If multiple reference standards had been used, we tried to extract data on the number of patients undergoing each standard and the selection criteria.

US characteristics Recorded were, if available: (1) type of probe; (2) frequency of probe; (3) type of scanning (conventional grey scale, pulsed, colour or power Doppler, graded compression); (4) criteria for the presence of ACD.

CT characteristics The following CT features, if available, were recorded: (1) type of scanner [non-helical, helical

(single- or multislice CT)]; (2) slice thickness or collimation used; (3) use of contrast agents (oral, intravenous and/or rectal contrast) and, if so, the amount; (4) criteria for the presence of ACD.

Data synthesis and analysis

We constructed a 2×2 contingency table for US and CT compared to the reference standard. From these raw data we calculated sensitivity as $TP/(FN + TP)$ and specificity as $TN/(FP + TN)$ for each modality in every included study. Individual study sensitivity and specificity results were plotted in a forest plot and plot in a receiver-operating characteristics (ROC) space to explore inter-study heterogeneity in test performance. The Cochran Q-test and I^2 -test statistics were used to statistically detect and quantify heterogeneity. The Q-test examines the null hypothesis that the results of the investigated studies are homogeneous. A statistically significant result of the Q-test, with a p-value less than 0.05, was assumed to indicate substantial heterogeneity. For quantification of heterogeneity the I^2 -test statistic with 95% confidence intervals was used. The I^2 -test is a measure of inconsistency describing the percentage of total variation across studies that is due to heterogeneity rather than chance. This statistic is a percentage, with larger percentages indicating more heterogeneity [9, 10].

Several statistical models (random, fixed, or mixed effects models) are available when performing a meta-analysis. The Akaike Information Criterion value [11], a global measure of goodness of fit of a statistical model, was used to compare the fit of each available model. It showed that the bivariate random effects model had the best fit and this model was therefore used for meta-analysis. The bivariate random effects model [12, 13] will produce a weighted average of sensitivity and specificity (also called the mean summary estimates of sensitivity and specificity) with corresponding confidence intervals based on the individual study results. In the bivariate random effects model, the logit-transformed sensitivities and logit-transformed specificities are assumed to follow a bivariate normal distribution across studies around a mean logit-sensitivity and mean logit-specificity. The mean logit-sensitivity and mean logit-specificity and the corresponding standard errors were used to obtain the summary estimates of sensitivity and specificity with corresponding confidence intervals after antilogit transformation. The summary estimates of sensitivity and specificity were used to calculate the positive summary likelihood ratio (LR+) as (sensitivity/1-specificity) and the negative summary likelihood ratio (LR-) as (1-sensitivity/specificity). Summary likelihood ratios were calculated with corresponding confidence intervals for each imaging modality. The LR+ is the ratio of the percentage of patients with ACD and a

positive test result and the percentage of patients without ACD with a positive test result. A diagnostic test with a LR+ of 10 and a LR- of 0.01 is generally considered as a test with good diagnostic performance.

Identification of alternative diagnoses

The number of patients with an alternative diagnosis in each study will be recorded. The number of these alternative diagnoses that were detected by US and/or CT will be used to calculate and compare the sensitivity for the identification of alternative diagnosis of both investigations.

Head-to-head comparative studies

Due to the superiority in methodology, we will highlight the results of the head-to-head comparative studies in our results. In these studies data regarding the ability of both investigations to detect the same radiological abnormalities will be extracted. The agreement between US and CT findings, for example for the detection of peri-colic fat inflammation, will be expressed as the kappa statistic. According to Landis and Koch [14] kappa (κ) values can be divided into the following levels of agreement, $\kappa < 0.20$ poor agreement, $\kappa = 0.21-0.40$ fair agreement, $\kappa = 0.41-0.60$ moderate agreement, $\kappa = 0.61-0.80$ good agreement, and $\kappa = 0.81-1.00$ excellent agreement.

Analyses were performed in SAS 9.1 (SAS Institute Inc., Cary, NC) and Microsoft Excel 5.0 (Microsoft Office, Bellevue, WA). The estimates of US and CT were compared by means of a Z-test for unpaired data. In all statistical tests a p-value lower than 0.05 was assumed to indicate statistically significant differences.

Results

Search strategy and study selection

The initial search resulted in 1,689 articles, of which 26 were judged potentially eligible. Full text versions of these articles were used for further selection. Fourteen of the 26 potentially relevant studies had to be excluded because they either did not acquire data prospectively ($n=6$) [15–20], the study report did not allow for the extraction of data to calculate test accuracy ($n=4$) [21–24], they were earlier publications (double publications) on the same clinical trial ($n=2$) [25, 26], they had performed an additional transrectal US, which prohibited extraction of data on the diagnostic accuracy of abdominal US as a separate entity ($n=1$) [27], or they did not perform graded compression US ($n=1$) [28]. Twelve studies met all inclusion criteria (Fig. 1).

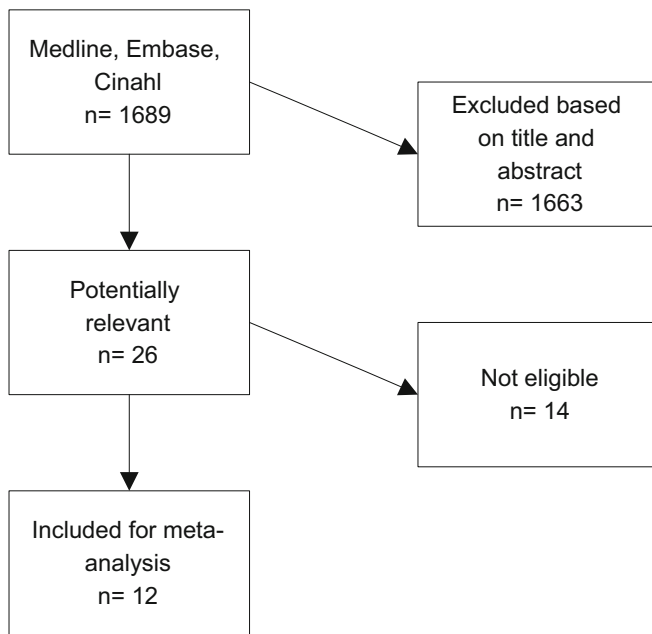


Fig. 1 Flow chart of search strategy and study selection

Study characteristics

Study quality assessment using the QUADAS tool revealed a number of methodological shortcomings. Although all studies investigated patients with suspected ACD, specific inclusion criteria were defined in only 50% of studies (QUADAS question 2). The time interval between execution of the reference standard and the index tests was unclear in nearly all studies (QUADAS question 4). A vague description of the execution of the reference standard was given in slightly more than half of the studies (QUADAS question 8b). Some studies, for example, did not report on the length of follow-up in conservatively treated patients. Three US studies [29–31] defined the

reference standard only for the patients with ACD as reference diagnosis (QUADAS question 5). The methods of two of these three studies report that for the verification of the test results all available clinical data, laboratory and radiological investigations, and operative and histology reports were used. However, in their results they do not provide the type of verification for patients with a final diagnosis other than diverticulitis. In all studies the index test results were incorporated in the reference standard (QUADAS question 9b). Only one study [31] reported its inconclusive test results (QUADAS question 11). Severe abdominal pain, too much bowel gas, and too much abdominal fat were the reasons given for six inconclusive US results. For QUADAS questions 3, 4 and 10 the underreporting of methodological details resulted in a high percentage of “unclear” responses. Figure 2 shows the responses to each question of the QUADAS tool. Based on this assessment we conclude that the overall study quality is moderate (but not poor) and that methodological study details were often underreported. In the [appendix](#) the QUADAS items are reported in detail.

Other study design characteristics

All studies were single-centre studies initiated by the departments of radiology (n=8), surgery (n=3), or internal medicine (n=1). Two studies were methodologically superior to the other studies since they had performed a head-to-head comparative study of US and CT [32, 33], while the other ten studies investigated the diagnostic accuracy of one investigation only (4 US, 6 CT) [28–31, 34–40]. Study characteristics are presented in Table 1.

The mean age of the patients in the included studies was 61 years for US and 63 years for CT. The total number of included patients was 630 for US and 684 for CT. Prevalence of ACD varied in both US and CT studies, ranging from 36% to 68%. The mean prevalence of

Fig. 2 The QUADAS scores of the included studies are summed up per item and presented in a bar chart

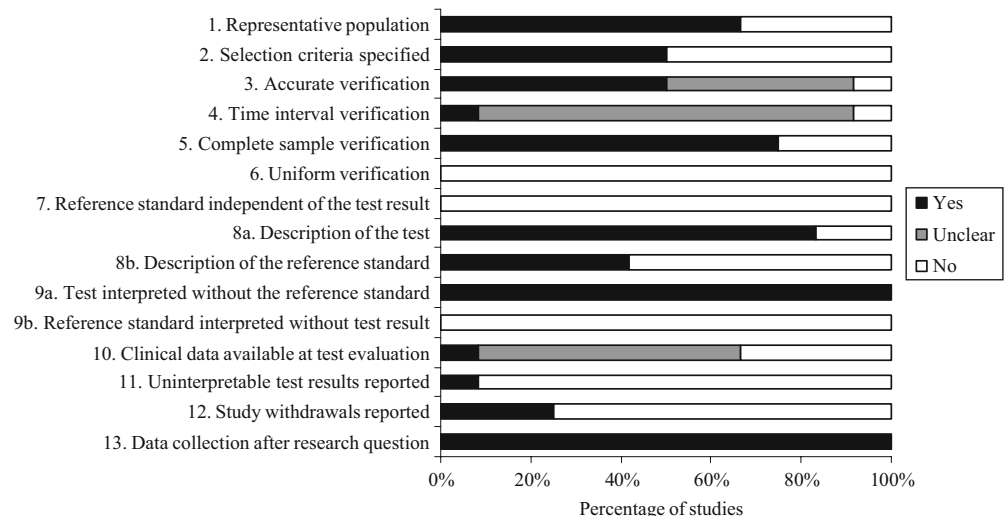


Table 1 Study characteristics

Study/year	Country	No. of patients	Inclusion criteria	Reference standard (no. of patients)	Observer experience
Ultrasound					
Verbanck, 1989 ^a [30]	Belgium	123	Suspected ^d	Surgery ^e (10), endoscopy (5), contrast enema (43)	Not specified
Schwerk, 1993 ^a [29]	Germany	161	Suspected ^d	Surgery ^e (24), clinical follow-up ^f (50)	Experienced ^g
Zielke, 1997 ^a [31]	Germany	143	Suspected ^d	Surgery ^e (32), clinical follow-up ^f (42)	> 3 months ^h
Garcia-Aguayo, 2002 [34]	Spain	76	Pain in LLQ with fever and/or leucocytosis	Surgery ^e (10), clinical follow-up ^f (66)	Experienced ^g
Computed tomography					
Cho, 1990 [35]	USA	56	Pain in LLQ with fever and/or leucocytosis	Surgery ^e (22), clinical follow-up ^f (34)	Not specified
Doring, 1990 [36]	Germany	33	Lower abdominal pain, leucocytosis and an elevated sedimentation rate	Surgery ^e (9), clinical follow-up ^f (24)	Not specified
Stefansson, 1997 [37]	Sweden	88	Pain in LLQ or lower abdomen	Surgery ^e (30), clinical follow-up ^f (58)	Experienced ^l
Rao, 1998 [38]	USA	150	Suspected ^d	Histopathology ^b (41), clinical follow-up ⁱ (109)	Staff member or fellow ^j
Werner, 2003 [39]	Germany	120	Suspected ^d	Surgery ^e (49), clinical follow-up ^k (71)	Experienced ^l
Tack, 2005 [40]	Belgium	110	Pain in LLQ < 2 weeks	Expert panel ^m (110)	20 years body CT
Ultrasound vs. Computed tomography					
Pradel, 1997 [32]	France	64	Suspected ^c	Surgery ^e (18), clinical follow-up ^f (46)	3–5 years
Farag Soliman, 2004 [33]	Germany	63	Suspected ^d	Surgery ^e (34), clinical follow-up ^f (29)	Not specified

^aReference standard specified only for patients with diverticulitis

^bObtained at surgery or colonoscopy

^cReferred for imaging to the department of radiology

^dPatients suspected of diverticulitis not further specified

^eSurgical findings combined with histopathology

^fDuration not specified, combined with endoscopic and radiological findings

^gExperienced radiologists or investigator, not further specified

^hSurgeons in training with a minimum of 3-months rotation at the surgical ultrasonography department

ⁱDuration at least 2 months

^jFrom the department of emergency radiology

^k6 to 12 months clinical follow-up combined with findings of other diagnostic investigations

^lCT separately interpreted by two experienced radiologists, diagnosis was a consensus between the radiologists

^mTwo radiologists and a gastroenterologist assigned the final diagnosis based on pathology results in 74 patients and on all clinical data and CT findings in 36 patients

complicated ACD was not significantly different, 22% in US studies and 24% in CT studies. Other patient population characteristics are presented in Table 2.

Different types of reference standards were applied, including surgery combined with histopathology, clinical follow-up, and other diagnostic investigations, such as barium enema and endoscopy. The number of patients undergoing each type of reference standard is summarised in Table 1. Most patients (n=555) had been treated conservatively, and clinical follow-up was applied as the

reference standard in these patients. A smaller number of patients (n=358) underwent surgery or colonoscopy, either during the acute phase or electively, and had a histopathologically confirmed reference diagnosis. In the study by Verbanck et al. the reference standard was a barium enema in 74% (n=43) of patients with the final diagnosis of ACD.

We found differences in the positivity threshold of US and CT for the presence of ACD between studies. For example, bowel wall thickening with peri-colic fat inflammation was considered diagnostic for ACD in five studies,

Table 2 Patient characteristics

Study/year	No. of patients	Clinical setting	Women (%) ^a	Mean age (range) ^c	Prevalence % (n) ^d	Complicated diverticulitis % ^e
Ultrasound						
Verbanck, 1989 [30]	123	N.A. ^f	N.A.	N.A.	43 (52)	16 (8/52)
Schwerk, 1993 [29]	161	N.A. ^f	52	57 (22–88) ^g	46 (74)	27 (20/74)
Zielke, 1997 [31]	143	Hospitalised	55	56 (20–89)	52 (74)	15 (11/74)
Garcia-Aguayo, 2002 [34]	76	N.A. ^f	47	58 (24–83)	68 (52)	10 (5/52)
Computed tomography						
Cho, 1990 [35]	56	Hospitalised	66	63 (28–98)	48 (27)	60 (16/27)
Doring, 1990 [36]	33	N.A. ^f	60	71 (40–84)	64 (21)	10 (2/21)
Stefansson, 1997 [37]	88	N.A. ^f	72	63 (29–91) ^g	59 (52)	12 (6/52)
Rao, 1998 [38]	150	Emergency dept. ^b	61	59 (19–92)	43 (64)	20 (13/64)
Werner, 2003 [39]	120	N.A. ^f	54	62 (21–88)	56 (67)	16 (11/67)
Tack, 2005 [40]	110	N.A. ^f	64	57 (30–82)	36 (39)	36 (14/39)
Ultrasound vs. Computed tomography						
Pradel, 1997 [32]	64	N.A. ^f	N.A.	64 (38–87)	52 (33)	27 (9/33)
Farag Soliman, 2004 [33]	63	Hospitalised	N.A.	61.8 ^e	68 (43)	47 (20/43)

N.A. not available

^aPercentage of women included^bPatients suspected of diverticulitis at the emergency department^cMean age and range between parenthesis^dPrevalence of diverticulitis as diagnosed by reference standard^ePercentage of complicated diverticulitis on all patients with diverticulitis^fClinical setting not specified^gMedian age instead of mean age

but was judged as a negative test result in four other studies if the presence of diverticula was not additionally visualised. The diagnostic criteria for the presence of ACD as well as other characteristics of US and CT are presented in Tables 3, 4 and 5. Multi-slice helical CT was used in four CT studies, single-slice helical CT in one and conventional CT in three. The use of contrast agents, slice thickness and

interval differed among studies. Intravenous and rectal contrast was administered in the majority of CT studies.

The two head-to-head comparative studies in our meta-analysis performed US and CT blinded to each others result and within 24 h of each other. Similar US and CT criteria for the presence of ACD were used.

Table 3 US features in the included studies

Study/year	Type of probe	Probe frequency (MHz)	Type of scanning	Diagnostic criteria for diverticulitis
Verbanck, 1989 [30]	Linear and curved	3.5–5	Graded compression	Bowel wall thickening ^a
Schwerk, 1993 [29]	Linear and curved	5	Graded compression	Peri-colic fat inflammation with bowel wall thickening ^b
Zielke, 1997 [31]	Linear and curved	3.5–5	Graded compression	Bowel wall thickening ^b with decreased peristalsis
Pradel, 1997 [32]	Linear and curved	5–10 and 2–4	Graded compression	Peri-colic fat inflammation with bowel wall thickening ^b
Garcia-Aguayo, 2002 [34]	Linear and curved	3.7–7.5 and 3.7–7.5	Graded compression	Peri-colic fat inflammation with bowel wall thickening ^b
Farag Soliman, 2004 [33]	Linear and curved	6–10 and 3.5–6	Graded compression	Peri-colic fat inflammation with bowel wall thickening ^b

N.A. not available

^a>4 mm at least 5 cm long^b>4–5 mm

Table 4 CT features in the included studies

Study/year	Type of scanner	CT protocol, slice thickness/interval	Contrast agents (amount)	Diagnostic criteria for diverticulitis
Cho, 1990 [35]	Non-helical	Lower abdomen: 10×10 mm ^a	I.v. in 50% of patients Standard oral and rectal contrast Air insufflation in 77% of patients	Peri-colic fat inflammation with bowel wall thickening ^b
Doringer, 1990 [36]	Non-helical	N.A.	I.v. and rectal (not specified)	Peri-colic fat inflammation with bowel wall thickening ^b and diverticula
Pradel, 1997 [32]	Non-helical	Upper abdomen: 8×12–16 mm Lower abdomen: 8×8 mm	Intravenous in 88% patients (2 l/kg) Oral in 64% of patients (800 mL) ^c	Bowel wall thickening ^b or diverticula with either peri-colic fat inflammation or abscess
Stefansson, 1997 [37]	Helical, multi-slice	Upper abdomen: 8×12 mm Lower abdomen: 8×8 mm	I.v. (100 ml) Oral (800 ml) Rectal air insufflation	Peri-colic fat inflammation with bowel wall thickening ^b and diverticula
Rao, 1998 [38]	Helical	Abdominal: 5×7.5 mm	I.v. in 2% of patients Rectal (1000 ml)	Peri-colic fat inflammation with bowel wall thickening ^b and diverticula
Werner, 2003 [39]	Helical, multi-slice	Pelvic floor to lower liver edge: 4×2.5 mm	I.v. (1.3 ml/kg) Rectal (1000 ml)	Peri-colic fat inflammation with bowel wall thickening ^b and diverticula
Farag Soliman, 2004 [33]	Helical, multi-slice	Abdominal: 5×7 mm and 5×5 mm	I.v. (100 ml) Rectal (1000 ml)	Peri-colic fat inflammation with bowel wall thickening ^b
Tack, 2005 [40]	Helical, multi-slice	Abdominal: 4×2.5 mm	I.v. (120 ml)	Peri-colic fat inflammation with bowel wall thickening ^b

N.A. not available

^aWithout signs of ACD in the lower abdomen scanning was proceeded to the upper abdomen^b>4–5 mm^cRectal air or contrast was used when standard CT was not satisfactory

Sensitivity and specificity

Inter-study heterogeneity in diagnostic performance is shown in the ROC plot (Fig. 3). Individual sensitivities and specificities with corresponding confidence intervals of US and CT with the results of the Q- and I²-test are presented in Figs. 4 and 5. Mean summary sensitivity estimates of US and CT were not significantly different: 92% (95% CI: 80% to 97%) for US versus 94% (95% CI: 87% to 97%) for CT (p=0.65). Mean summary specificity estimates were 90% (95% CI: 82% to 95%) for US and 99% (95% CI: 90% to 100%) for CT and not significantly different (p=0.07).

Using the Q-test, we found significant heterogeneity in both the sensitivities and the specificities of US and CT studies. The I² for the sensitivities of US was 57% (95% CI: 0% to 83%) and for the specificities 64% (95%CI: 13% to 85%). In Fig. 4 one outlying sensitivity value of CT [37]

and two outlying specificity values [32, 36] responsible for the heterogeneity are easy to identify. The low sensitivity of CT reported by Stefansson was possibly due to a diagnostic laparoscopy rate of 38% in the patients with diverticulitis as these laparoscopies revealed false-negative CT results. These diagnostic laparoscopies were performed routinely as part of the study in the second half of the study period. Removal of the outliers reduced the I² percentage from 76% (95% CI: 52% to 88%) to 0% (95% CI: 0% to 79%) for the sensitivities of the CT studies and from 73% (95% CI: 44% to 87%) to 0% (95% CI: 0% to 79%) for the specificities of the CT studies. Disregarding the outliers resulted in a summary sensitivity estimate for CT of 96% (95%CI: 92% to 98%) and a summary specificity estimate of 99% (95%CI: 97% to 100%). In other words, the observed heterogeneity in CT results had no significant influence on the summary estimates of CT accuracy.

Table 5 Test accuracy of US studies

Study/year	Patients with diverticulitis		Patients without diverticulitis		Sensitivity ^a	Specificity ^a	Positive likelihood ratio ^a	Negative likelihood ratio ^a
	TP	FN	TN	FP				
Ultrasound								
Verbanck, 1989 [30]	44	8	57	14	0.85 (44/52)	0.80 (57/71)	4.3 (0.85/0.20)	0.19 (0.15/0.80)
Schwerk, 1993 [29]	73	1	84	3	0.99 (73/74)	0.97 (84/87)	28.6 (0.99/0.03)	0.01 (0.01/0.97)
Zielke, 1997 [31]	62	12	64	5	0.84 (62/74)	0.93 (64/69)	11.6 (0.84/0.07)	0.17 (0.16/0.93)
Pradel, 1997 [32]	28	5	26	5	0.85 (28/33)	0.84 (26/31)	5.3 (0.85/0.16)	0.18 (0.15/0.84)
Garcia-Aguayo, 2002 [34]	42	10	19	5	0.81 (42/52)	0.79 (42/52)	3.9 (0.81/0.20)	0.24 (0.19/0.79)
Farag Soliman, 2004 [33]	43	0	20	0	1.00 (43/43)	1.00 (43/43)	N.A.	N.A.
Summary estimate (95%CI)^b					0.92 (80–97)	0.90 (82–95)		
Summary likelihood ratio (95%CI)^b							9.6 (5.0–18.6)	0.09 (0.04–0.23)
Computed tomography								
Cho, 1990 [35]	25	2	29	0	0.93 (25/27)	1.00 (29/29)	N.A.	0.07 (0.07/1.00)
Doring, 1990 [36]	20	1	9	3	0.95 (20/21)	0.75 (9/12)	3.8 (0.95/0.25)	0.06 (0.05/0.75)
Pradel, 1997 [32]	30	3	24	7	0.91 (30/33)	0.77 (24/31)	4.0 (0.91/0.23)	0.12 (0.09/0.77)
Stefansson, 1997 [37]	36	16	36	0	0.69 (36/52)	1.00 (36/36)	N.A.	0.31 (0.31/1.00)
Rao, 1998 [38]	62	2	86	0	0.97 (62/64)	1.00 (86/86)	N.A.	0.03 (0.03/1.00)
Werner, 2003 [39]	65	2	52	1	0.97 (65/67)	0.98 (52/53)	51.4 (0.97/0.02)	0.03 (0.03/0.98)
Farag Soliman, 2004 [33]	42	1	20	0	0.98 (42/43)	1.00 (20/20)	N.A.	0.02 (0.02/1.00)
Tack, 2005 [40]	36	3	70	1	0.92 (36/39)	0.99 (70/71)	65.5 (0.92/0.01)	0.08 (0.08/0.99)
Summary estimate (95% CI)^b					0.94 (87–97)	0.99 (90–100)		
Summary likelihood ratio (95% CI)^b							78.4 (8.7–706.6)	0.06 (0.03–0.13)

N.A. not available

^aIndividual study sensitivity, specificity and the likelihood ratios were calculated from the raw data of the individual studies^bMean summary estimates of sensitivity, specificity and the summary likelihood ratios were calculated from the results of the bivariate random effects analysis

Although the Verbanck study (ref) used barium enema as a reference standard, it did not result in an outlying sensitivity and specificity of US. Excluding this study from the meta-analysis would not significantly change the summary estimates and would result in a summary sensitivity estimate for US of 93% (95% CI: 79% to 98%) and summary specificity for US of 92% (95% CI: 88% to 95%).

Likelihood ratio

Calculated summary LR+ were 9.63 (95% CI: 4.98 to 18.62) for US and 78.41 (95% CI: 8.70 to 706.58) for CT (p=0.07). Calculated summary LR- were 0.09 (95% CI: 0.04 to 0.23) for US and 0.06 (95% CI: 0.03 to 0.13) for CT (p=0.53).

Identification of alternative diseases

Eight of the 12 studies reported on the sensitivity for the identification of alternative diseases. This sensitivity

ranged between 33% and 78% for the US studies and between 50% and 100% for the CT studies. Table 6 presents the sensitivities for alternative diseases for the US and CT studies.

Head-to-head comparative studies

Although the head-to-head comparative studies did not report a significant difference between the accuracy of US and CT, there was a difference in their individual accuracy results. Farag Soliman et al. [33] reported higher sensitivities (100% for US and 98% for CT) and specificities (100% for both US and CT) compared to the sensitivities (85% for US and 91% for CT) and specificities (84% for US and 77% for CT) of Pradel et al. [32]. The study by Farag Soliman et al. merely included hospitalised patients, in contrast to the study by Pradel et al. in which all patients with suspected ACD referred for US or CT were included. The percentage of complicated ACD was 47% in the study by Farag Soliman et al. compared to 27% in the study by Pradel et al. The difference in clinical setting and spectrum of disease could be the cause of differences in

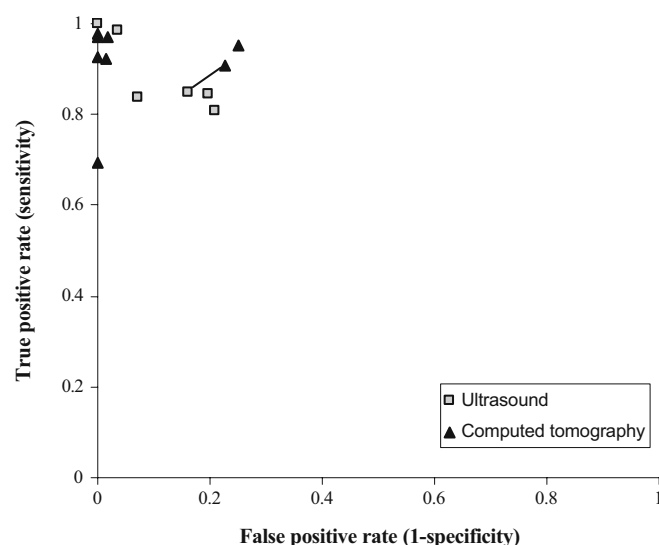


Fig. 3 Sensitivity and 1-specificity for US results (open square) and for CT results (closed triangle) per included study are shown in a ROC plot. For the two head-to-head comparative studies the open squares and closed triangles represent the individual results of the head-to-head comparative studies and are connected by a line. This is visible for only one pair; the other paired data points are located in the upper left corner in the ROC plot

reported accuracy values. Pradel et al. report good kappa agreement between US and CT findings. Kappa agreement was good for depicting peri-colic fat inflammation ($\kappa=0.78$), good for depicting bowel wall thickening ($\kappa=0.69$), and good for depicting peri-colic abscesses ($\kappa=0.69$).

Discussion

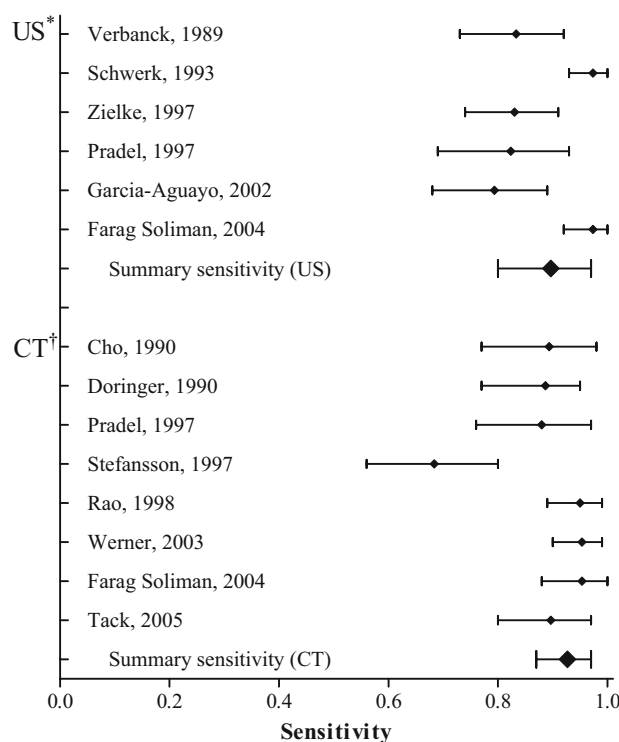
In this systematic review we found that diagnostic studies of US and CT in patients suspected of ACD are of moderate quality. No significant differences in the diagnostic accuracy of US and CT in diagnosing ACD were found. Calculated sensitivities, specificities, positive and negative likelihood ratios were all higher for CT, but none of these differences were significant. The range of the sensitivities for the identification of alternative diagnoses was higher for CT than US, suggesting that CT is more accurate for detecting alternative diagnoses.

Although the two head-to-head comparative studies found different accuracy values for US and CT, they both concluded that the accuracy of US and CT was not significantly different. These two studies with the best methodological design, providing approximately 20% of our study population, support the result of our overall meta-analysis [32, 33].

Heterogeneous results, reported by studies investigating the same effect, can lead to inaccurate and

irrelevant summary point estimates when pooled for meta-analysis. For this reason we explored heterogeneity in the US and CT study results. In the US studies the heterogeneity was slightly above 50%, which can be considered as moderate heterogeneity [9], allowing the pooling of these results [9]. Although we found heterogeneity in CT study results [32, 36, 37], resulting in high I^2 values, exploring this heterogeneity showed that it did not influence the summary estimates of CT significantly.

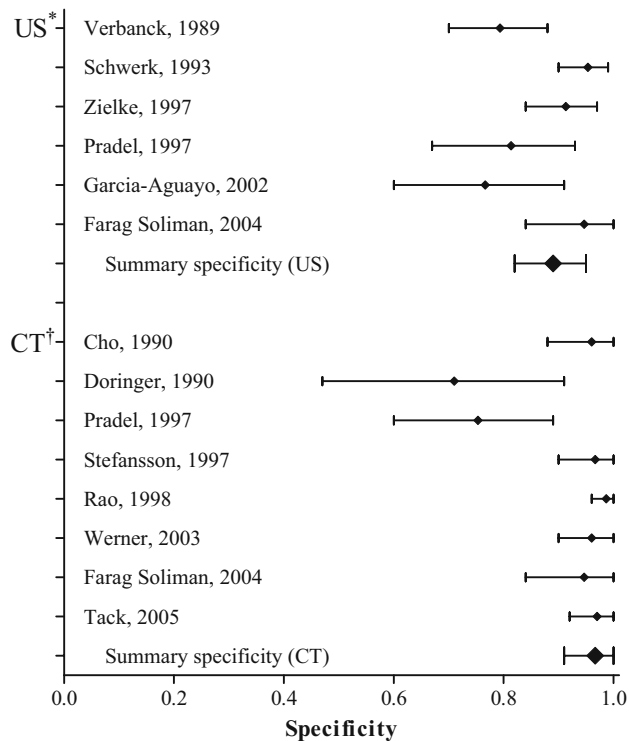
The studies investigating the diagnostic accuracy of US and CT in ACD were susceptible to bias because they applied differential verification. With US and CT results being part of the reference standards, an incorporation bias could lead to over-estimation of the diagnostic accuracy. Unlike histopathology, clinical follow-up is a reference standard open to subjective interpretation and is less likely to identify the correct reference diagnosis. Using a reference standard that is open to subjective interpretation can enhance the effect of over-estimation [41]. For example, a patient with a



* Q-test of US sensitivities: 11.7 ($p=0.04$, $Df=5$), $I^2=57\%$ (95%CI: 0%-83%)

† Q-test of CT sensitivities: 29 ($p=0.001$, $Df=7$), $I^2=76\%$ (95%CI: 52%-82%)

Fig. 4 Individual sensitivities and the summary sensitivity estimates of US and CT studies are shown with their corresponding 95% confidence intervals. Heterogeneity between study results is presented in the footnotes



* Q-test of US specificities: 13.9 ($p=0.02$, $Df=5$), $I^2=64\%$ (95%CI: 13%-85%)

† Q-test of CT specificities: 25.6 ($p<0.001$, $Df=7$), $I^2=73\%$ (95%CI: 44%-87%)

Fig. 5 Individual specificities and the summary specificity estimates of US and CT studies are shown with their corresponding 95% confidence intervals. Heterogeneity between study results is presented in the footnotes

positive US or CT for ACD who is successfully treated conservatively will most likely receive the reference diagnosis acute diverticulitis. However, the underlying illness may not be ACD, since other illnesses may present with similar symptoms and also resolve completely when treated conservatively. This can lead to an underestimation of false-positive test results and therefore give an over-estimation of the accuracy. Since this is a problem related to the reference standard, it will likely have a similar effect on US and CT.

The diagnostic accuracy of US is thought to be more dependent on observer experience than CT. This is sometimes used as an argument to discourage the usage of US. Surprisingly, the study by Zielke [31] shows that 11 surgeons in training with at least 3 months of US experience achieved equal diagnostic accuracy compared to studies using experienced observers. This does not prove that the accuracy of US is not observer dependent for abdominal pathology, but an intensive training of several months in abdominal US could be sufficient to make an accurate diagnosis of ACD on US.

A limitation of our study is that the meta-analysis is mostly based on unpaired data. Ideally the diagnostic accuracy of two competing tests is investigated in the same patient population. Meta-analysis of such head-to-head studies will be able to estimate and compare the diagnostic accuracy of tests even with greater validity and precision. Only two head-to-head comparative

Table 6 The sensitivity for the detection of alternative disease for the US and CT studies

Study/year	No. of patients	No. of ACD	No. of NSAP*	No. of alternative diseases	Sensitivity for alternative diagnosis (no. detected by US or CT)
Ultrasound					
Verbanck, 1989 [30]	123	52	N.A.	N.A.	N.A.
Schwerk, 1993 [29]	161	74	50	37	70% (26)
Pradel, 1997 [32]	64	33	7	24	33% (8)
Zielke, 1997 [31]	143	74	34	35	46% (16)
Garcia-Aguayo, 2002 [34]	76	52	N.A.	N.A.	N.A.
Farag Soliman, 2004 [33]	63	43	11	9	78% (7)
Computed tomography					
Cho, 1990 [35]	56	27	6	23	87% (20)
Doring, 1990 [36]	33	21	N.A.	N.A.	N.A.
Pradel, 1997 [32]	64	33	7	24	50% (12)
Stefansson, 1997 [37]	88	52	N.A.	N.A.	N.A.
Rao, 1998 [38]	150	64	22	64	78% (50)
Werner, 2003 [39]	120	67	22	31	71% (22)
Farag Soliman, 2004 [33]	63	43	11	9	89% (8)
Tack, 2005 [40]	110	39	49	22	100% (22)

*Non-specific abdominal pain: NSAP

studies were performed. In the unpaired data between-study heterogeneity, i.e., between US and CT studies, is not to be avoided. In our results the between-study heterogeneity and methodological shortcomings in regard to patient selection, reference standard, experience of observer, imaging technique, and test interpretation are clearly presented. Facing heterogeneity and methodological shortcomings is almost inevitable when performing a meta-analysis of diagnostic accuracy studies. We reported and explored this in our results. For example, we reported on heterogeneity in patient selection, but exploration showed that the prevalence of ACD and complicated ACD nevertheless was comparable between US and CT populations. Complete reporting of all study characteristics and methodological shortcomings facilitates valid interpretation of the result of our meta-analysis.

Reporting proper methodological details in diagnostic studies is a known problem. Without these details on methodology, results of studies reporting on diagnostic performance are hard to interpret. This error in reporting was also detected in some of the included studies during quality assessment using the QUADAS tool. Attempts are made to improve methodological reporting of diagnostic test accuracy studies with the Standards for Reporting of Diagnostic Accuracy (STARD) [42]. The STARD initiative provides a checklist with items that should be included in the report of a study of diagnostic accuracy.

We tried to minimise bias in our meta-analysis by using two independent reviewers for data extraction, using specified inclusion criteria, and exploring heterogeneity between and within studies.

Two surveys conducted under surgeons from the UK and the USA [3, 4] showed that the daily use of diagnostics in patients suspected of ACD varied significantly. Of the questioned surgeons from the UK who found imaging necessary at initial assessment, 42% favoured CT and 33% favoured US. In contrast, two-thirds of the questioned surgeons from the USA favoured CT and less than 7% favoured US. So where US is used as a competitive initial diagnostic test in the UK, it seems that in the USA less value is rewarded to the diagnostic opportunities of US. This is illustrated by the appropriateness criteria for imaging in patients with left lower quadrant pain of the American College of Radiology [5], which state that CT is more appropriate than US, especially in older patients with a typical presentation of ACD. This preference for CT in the USA is reflected in this meta-analysis. All US studies in this meta-analysis concerned European studies, while CT studies originated both from the USA and Europe. In countries with a high prevalence of obesity physicians will favour CT, since the use of US is practically inappropriate in obese patients. With US being less frequently used in the USA, the performance of US by radiologists from the USA for the diagnosis of ACD might be lower. The preference for CT of many

physicians is also based on the fact that CT is often regarded as a more credible test than US for the exclusion and identification of alternative diagnosis. The range of sensitivity for the identification of alternative diagnosis for CT and US shows that this is probably true. Unfortunately, the included study did not provide data that made it possible to compare the ability of US and CT to exclude alternative diagnoses.

The use of magnetic resonance colonography for the diagnosis of ACD was investigated by Ajaj et al. [43]. This feasibility study reported a promising sensitivity of 86% and specificity of 92%. Although magnetic resonance colonography is not yet routinely applied in the acute setting in patients suspected of ACD in daily practice, the accuracy results seem promising and feasibility of this modality in the diagnostic work-up of these patients deserves attention.

The practice parameters from 2006 by the American Society of Colon and Rectal Surgery [44] also advocate the use of CT in diagnosing ACD. They state that US can sometimes be useful to differentiate between a phlegmon and an abscess in ACD, but that US findings are often obscured by overlying bowel loops. Our study recorded only few inconclusive US investigations. Graded compression US possibly reduces the number of inconclusive findings due to overlying bowel loops. The two head-to-head comparative studies we included both used the same US and CT criteria for the presence of ACD. Their results show that next to differentiating between a phlegmon and an abscess US can accurately measure bowel wall thickness, show peri-colic fat inflammation and detect complications. Kappa agreement between US and CT findings was good for the above mentioned imaging features [32].

This analysis of 16 years of published literature comprehends roughly the same amount of data for US and CT and provides detailed information on between-study heterogeneity in both US and CT studies. In conclusion, diagnostic accuracy studies of US and CT in patients with suspected ACD are of moderate quality and there is a need for new methodologically solid studies. Our meta-analysis found no significant difference between the diagnostic accuracy of US and CT in diagnosing ACD. The best available evidence shows that both US and CT can be used as an initial diagnostic tool in the assessment of patients suspected of having ACD. However, in severely ill patients presenting with abdominal pain the use of CT is probably more suitable as CT images are more able than US to assist in planning of a radiological or surgical intervention, and CT images in contrast to US can be re-read at any time by any specialist involved in the treatment of severely ill patients. Moreover, reviewed data indicate that CT is more accurate for detecting alternative diagnoses than us.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Databases used

- MEDLINE
- EMBASE
- CINAHL
- Cochrane

Databases NOT used

- DARE (Cochrane)
- SUMSEARCH/MEDSCAPE (also in Pubmed and Cochrane)

Studies marked not relevant, exclusion based on title and abstract

- Animal, laboratory studies
- Other imaging: PET/Scintigraphy/Transvaginal US/Endoscopic ultrasonography/MRI/urography/CT colonography/enterocyclis/Barium enema/intraoperative ultrasonography
- Comments/letters
- Treatment, fine-needle aspiration
- Case reports (<10 patients)
- Reviews

Complete database: 1689

Relevant papers (26)

PUBMED/EMBASE

Population:

[Diverticulitis (MeSH) OR Colonic diverticulitis (MeSH)]
AND

Diagnostic

["Radiography"(MeSH) OR "radiography"(Subheading)] OR
["Radiography, Thoracic"(MeSH) OR "Radiography, Abdominal"
(MeSH)] OR ["Tomography, X-Ray Computed"(MeSH) OR
"Tomography Scanners, X-Ray Computed"(MeSH) OR "Tomogra-
phy, Spiral Computed"(MeSH)] OR ["Ultrasonography"(MeSH) OR
(ultrasonography"(Subheading))]

AND

Methodological filters

No methodological filter was applied
human

Limit:

Pubmed: Search strategy yielded 901 articles

EMBASE (1980–2005): Search strategy yielded 734 articles

CINAHL

Population

[Diverticulitis (MeSH) OR Colonic diverticulitis (MeSH)]
AND

Diagnostic

[Ultrasonography OR Echography (MeSH) OR Radiography OR
RADIOGRAPHY (MESH) OR Computed tomography OR Com-
puter-Assisted Tomography]

Methodological filter

No methodological filter was applied

Limit

Human

CINAHL: Search strategy yielded 24 articles

COCHRANE

[Searches Cochrane Database or Systematic Reviews (Cochrane Reviews), Database of Abstracts of Reviews of Effects (DARE), The Cochrane Central Register of Controlled Trials (CENTRAL), The Cochrane Database of Methodology Reviews (Methodology Reviews), The Cochrane Methodology Register (CMR), Health Technology Assessment Database (HTA), NHS Economic Evaluation Database (NHS EED)]

AND

“Acute diverticulitis”

AND

Ultrasonography OR Echography OR Radiography OR Computed tomography OR Computer-Assisted Tomography

Population

Diagnostic

Search strategy Cochrane: No systematic reviews or clinical trials investigating the diagnostic accuracy of US and CT in diverticulitis were identified

QUADAS questions*:

- (1) Was the spectrum of patients representative of the patients who will receive the test in practice?
- (2) Were selection criteria clearly described?
- (3) Is the reference standard likely to correctly classify the target condition?
- (4) Is the time period between surgery (histopathology) and index test short enough to be reasonably sure that the target condition did not change between the two tests?
- (5) Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?
- (6) Did patients receive the same reference standard regardless of the index test result?
- (7) Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
- (8a) Was the execution of the index test described in sufficient detail to permit replication of the test?
- (8b) Was the execution of the reference standard described in sufficient detail to permit its replication?
- (9a) Were the index test results interpreted without knowledge of the results of the reference standard?
- (9b) Were the reference standard results interpreted without knowledge of the results of the index test?
- (10) Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
- (11) Were uninterpretable/intermediate test results reported?
- (12) Were withdrawals from the study explained?
- (13) Were the data collected after the research question was defined?

*The answer can be yes, no, or unclear

References

1. Stollman N, Raskin JB (2004) Diverticular disease of the colon. *Lancet* 363:631–639
2. Ferzoco LB, Raptopoulos V, Silen W (1998) Acute diverticulitis. *N Engl J Med* 338:1521–1526
3. Munikrishnan V, Helmy A, Elkhider H, Omer AA (2006) Management of Acute Diverticulitis in the East Anglian Region: Results of a United Kingdom Regional Survey. *Dis Colon Rectum* 49:1332–1340
4. Schechter S, Mulvey J, Eisenstat TE (1999) Management of Uncomplicated Acute Diverticulitis: Results of a survey. *Dis Colon Rectum* 42:470–476
5. Balfe DM, Levine MS, Ralls PW (2000) Evaluation of left lower quadrant pain. American College of Radiology. *ACR Appropriateness Criteria. Radiology* 215(suppl):167–171
6. Bruel JM (2003) Acute colonic diverticulitis: CT or ultrasound? *Eur Radiol* 13:2557–2559
7. Rhea JT (2000) CT evaluation of appendicitis and diverticulitis. Part II: Diverticulitis. *Emergency Radiology* 7:237–244
8. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 3:25
9. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560
10. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558

11. Akaike H (1974) A new look at Statistical Model Identification. *IEEE Transactions on Automatic Control* 19:716–723
12. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58:982–990
13. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8:239–251
14. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
15. Ernst S, Wypior HJ, Stark V, Rath M (1996) [The computed tomography of acute sigmoid diverticulitis]. *Rofo* 164:102–107
16. Hulnick DH, Megibow AJ, Balthazar EJ (1984) Computed tomography in the evaluation of diverticulitis. *Radiology* 152
17. Kircher MF, Rhea JT, Kihiczak D, Novelline RA (2002) Frequency, sensitivity, and specificity of individual signs of diverticulitis on thin-section helical CT with colonic contrast material: Experience with 312 cases. *Am J Roentgenol* 178
18. Ripolles T, Agramunt M, Martinez MJ, Costa S, Gomez-Abril SA, Richart J (2003) The role of ultrasound in the diagnosis, management and evolutive prognosis of acute left-sided colonic diverticulitis: a review of 208 patients. *Eur Radiol* 13:2587–2595
19. Wilson SR, Toi A (1990) The value of sonography in the diagnosis of acute diverticulitis of the colon. *AJR Am J Roentgenol* 154:1199–1202
20. Smith TR, Cho KC, Morehouse HT, Kratka PS (1990) Comparison of computed tomography and contrast enema evaluation of diverticulitis. *Dis Colon Rectum* 33
21. Ambrosetti P, Jenny A, Becker C, Terrier F, Morel P (2000) Acute left colonic diverticulitis - Compared performance of computed tomography and water-soluble contrast enema: Prospective evaluation of 420 patients. *Dis Colon Rectum* 43
22. Ambrosetti P, Becker C, Terrier F (2002) Colonic diverticulitis: Impact of imaging on surgical management - A prospective study of 542 patients. *Eur Radiol* 12:01
23. Eggesbo HB, Jacobsen T, Kolmannskog F, Bay D, Nygaard K (1998) Diagnosis of acute left-sided colonic diverticulitis by three radiological modalities. *Acta Radiol* 39:315–321
24. Hachigian MP, Honickman S, Eisenstat TE, Rubin RJ, Salvati EP (1992) Computed tomography in the initial management of acute left-sided diverticulitis. *Dis Colon Rectum* 35
25. Schwerek WB, Schwarz S, Rothmund M (1992) Sonography in acute colonic diverticulitis: A prospective study. *Dis Colon Rectum* 35
26. Zielke A, Hasse C, Bandorski T, Sitter H, Wachsmuth P, Grobholz R, Rothmund M (1997) Diagnostic ultrasound of acute colonic diverticulitis by surgical residents. *Surg Endosc* 11:1194–1197
27. Hollerweger A, Macheiner P, Rettenbacher T, Brunner W, Gritzmann N (2001) Colonic diverticulitis: Diagnostic value and appearance of inflamed diverticula-sonographic evaluation. *Eur Radiol* 11
28. Moll R, Mittelkötter U, Reith H-B, Schindler G, Thiede A (2002) Which imaging in case of sigmoid diverticulitis? The value of ultrasound (conventional B-mode in combination with hydrocolonosonography and colour flow doppler) in comparison to the well-established modalities like contrast enema and helical computertomography. *Zent bl Chir* 127
29. Schwerek WB, Schwarz S, Rothmund M, Arnold R (1993) [Colon diverticulitis: imaging diagnosis with ultrasound—a prospective study]. *Z Gastroenterol* 31:294–300
30. Verbanck J, Lambrecht S, Rutgeerts L, Ghillebert G, Buyse T, Naesens M, Tytgat H (1989) Can sonography diagnose acute colonic diverticulitis in patients with acute intestinal inflammation? A prospective study. *J Clin Ultrasound* 17
31. Zielke A, Hasse C, Nies C, Kisker O, Voss M, Sitter H, Rothmund M (1997) Prospective evaluation of ultrasonography in acute colonic diverticulitis. *Br J Surg* 84
32. Pradel JA, Adell J-F, Taourel P, Djafari M, Monnin-Delhom E, Bruel J (1997) Acute colonic diverticulitis: Prospective comparative evaluation with US and CT. *Radiology* 205
33. Farag SM, Wustner M, Sturm J, Werner A, Diehl SJ, Duber C, Post S (2004) Primary diagnostics of acute diverticulitis of the sigmoid. Sonography versus computed tomography: A prospective study. *Ultraschall in der Med* 25
34. Garcia-Aguayo FJ, Gil PM (2002) Acute diverticulitis of the sigmoid colon: Value of ultrasound as an initial diagnostic test. *Radiologia* 44
35. Cho KC, Morehouse HT, Alterman DD, Thornhill BA (1990) Sigmoid diverticulitis: Diagnostic role of CT - Comparison with barium enema studies. *Radiology* 176
36. Doringe E, Ferner R (1990) CT of colonic diverticulitis. *Rofo Fortschritte auf dem Gebiete der Röntgenstrahlen und der Neuen Bildgebenden Verfahren* 152
37. Stefansson T, Nyman R, Nilsson S, Ekblom A, Pahlman L (1997) Diverticulitis of the sigmoid colon. A comparison of CT, colonic enema and laparoscopy. *Acta Radiol* 38:313–319
38. Rao PM, Rhea JT, Novelline RA, Dobbins JM, Lawrason JN, Sacknoff R, Stuk JL (1998) Helical CT with only colonic contrast material for diagnosing diverticulitis: prospective evaluation of 150 patients. *AJR Am J Roentgenol* 170:1445–1449
39. Werner A, Diehl SJ, Farag-Soliman M, Duber C (2003) Multi-slice spiral CT in routine diagnosis of suspected acute left-sided colonic diverticulitis: A prospective study of 120 patients. *Eur Radiol* 13
40. Tack D, Bohy P, Perlot I, De M, V, Alkellani O, Sourtzis S, Gevenois PA (2005) Suspected acute colon diverticulitis: Imaging with low-dose unenhanced multi-detector row CT. *Radiology* 237
41. Rutjes AW, Reitsma JB, Di NM, Smidt N, van Rijn JC, Bossuyt PM (2006) Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 174:469–476
42. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Radiology* 226:24–28
43. Ajaj W, Ruehm SG, Lauenstein T, Goehde S, Kuehle C, Herborn CU, Langhorst J, Zopf T, Gerken G, Goyen M (2005) Dark-lumen magnetic resonance colonography in patients with suspected sigmoid diverticulitis: a feasibility study. *Eur Radiol* 15:2316–2322
44. Rafferty J, Shellito P, Hyman NH, Buie WD (2006) Practice parameters for sigmoid diverticulitis. *Dis Colon Rectum* 49:939–944