GASTROINTESTINAL

A comparison of the Accuracy of Ultrasound and Computed Tomography in common diagnoses causing acute abdominal pain

Adrienne van Randen · Wytze Laméris · H. Wouter van Es ·
Hans P. M. van Heesewijk · Bert van Ramshorst · Wim ten Hove · Willem H. Bouma ·
Maarten S. van Leeuwen · Esteban M. van Keulen · Patrick M. Bossuyt ·
Jaap Stoker · Marja A. Boermeester · on behalf of the OPTIMA study group

Abstract

Objectives Head-to-head comparison of ultrasound and CT accuracy in common diagnoses causing acute abdominal pain.
Materials and methods Consecutive patients with abdominal pain for >2 h and <5 days referred for imaging underwent both US and CT by different radiologists/radiological residents. An expert panel assigned a final diagnosis. Ultrasound and CT sensitivity and predictive values were calculated for frequent final diagnoses. Effect of patient characteristics and observer experience on ultrasound sensitivity was studied.

Results Frequent final diagnoses in the 1,021 patients (mean age 47; 55% female) were appendicitis (284; 28%), diverticulitis (118; 12%) and cholecystitis (52; 5%). The sensitivity of CT in detecting appendicitis and diverticulitis was significantly higher than that of ultrasound: 94% versus 76% ($p<0.01$) and 81% versus 61% ($p=0.048$), respectively. For cholecystitis, the sensitivity of both was 73% ($p=1.00$). Positive predictive values did not differ significantly between ultrasound and CT for these conditions. Ultrasound sensitivity in detecting appendicitis and diverticulitis was not significantly negatively affected by patient characteristics or reader experience.

Study group members are listed in the appendix I

A. van Randen (✉) · J. Stoker
Department of Radiology (suite G1-227),
Academic Medical Centre,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands
e-mail: a.vanranden@amc.uva.nl

W. Laméris · M. A. Boermeester
Department of Surgery, Academic Medical Center,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands

H. W. van Es · H. P. M. van Heesewijk
Department of Radiology, St Antonius Hospital,
Koekoekslaan 1,
3435 CM Nieuwegein, The Netherlands

B. van Ramshorst
Department of Surgery, St Antonius Hospital,
Koekoekslaan 1,
3435 CM Nieuwegein, The Netherlands

W. ten Hove
Department of Radiology, Gelre Hospitals,
Albert Schweitzerlaan 31,
7334 DZ Apeldoorn, The Netherlands

W. H. Bouma
Department of Surgery, Gelre Hospitals,
Albert Schweitzerlaan 31,
7334 DZ Apeldoorn, The Netherlands

M. S. van Leeuwen
Department of Radiology, University Medical Centre,
Heidelberglaan 100,
3584 CX Utrecht, The Netherlands

E. M. van Keulen
Department of Radiology, Tergooi Hospitals,
Van Riebeeckweg 212,
1213 XZ Hilversum, The Netherlands

P. M. Bossuyt
Department of Clinical Epidemiology, Biostatistics, and
Bioinformatics, Academic Medical Center,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands

Conclusion CT misses fewer cases than ultrasound, but both ultrasound and CT can reliably detect common diagnoses causing acute abdominal pain. Ultrasound sensitivity was largely not influenced by patient characteristics and reader experience.

Keywords Acute abdominal pain · Computed tomography · Ultrasound · Appendicitis · Emergency Department

Introduction

Of all patients presenting to the Emergency Department (ED), approximately 10% have complaints of acute abdominal pain. Acute abdominal pain can be caused by a wide variety of conditions. Formerly these patients were thought to have a acute abdomen, and surgery was indicated. Nowadays, patients with acute abdominal pain, even if accompanied by abdominal tenderness and rigidity, not all of them will undergo surgery, while others without abdominal rigidity are operated on [1]. Diagnostic imaging is widely used in the work-up of patients with acute abdominal pain. Ultrasound and computed tomography (CT) are both frequently used on top of clinical and laboratory evaluation. The American College of Radiology suggests an abdomen/pelvis CT with contrast medium in patients with acute abdominal pain [2]. Others are in favour of ultrasound as the primary imaging technique mainly because ultrasound is easily accessible and does not expose patients to ionising radiation [3, 4]. Ionising radiation exposure at CT is associated with the risk of radiation-induced cancer. This is a drawback of CT, especially as CT is increasingly being used in the diagnostic work-up of young patients. This may prompt the evaluation of alternative imaging strategies next to CT, such as ultrasound and MRI [5]. However, diagnoses should not be missed or delayed and thus the most accurate imaging technique should be used.

A previous evaluation of diagnostic strategies for unselected patients with acute abdominal pain favoured a conditional CT strategy for the detection of urgent conditions, with ultrasound first and CT after a negative or inconclusive ultrasound [6]. For common diagnoses causing acute abdominal pain, such as appendicitis literature suggests CT in the diagnostic work-up of these patients suspected with appendicitis [7]. Primarily usage of CT in patients suspected with diverticulitis is not supported by literature, as accuracy of US and CT were comparable in a recent published meta-analysis [8]. The fact that ultrasound is observer-dependent is thought to be a major disadvantage. Its accuracy, as reported in the literature, may be overestimated because in a research environment ultrasound is usually performed by highly

experienced observers. Ultrasound accuracy could also be lower in specific patient subgroups, such as in obese patients, women, and in specific age groups, especially women of reproductive age. CT, on the other hand has good inter-observer agreement in general, and even excellent inter-observer agreement for frequent diagnoses causing acute abdominal pain (e.g. appendicitis and diverticulitis) [9].

Ultrasound will only be an acceptable alternative for CT if its diagnostic accuracy is comparable, i.e. if it can be reliably used for the detection of frequent causes of abdominal pain in unselected patients presenting at the ED. In this paper we report a head-to-head comparison of the accuracy of ultrasound and CT in detecting common causes of acute abdominal pain, such as appendicitis and diverticulitis, in patients presenting at the ED with acute abdominal pain. We also evaluated to what extent the accuracy of ultrasound was affected by patient characteristics and observer experience.

Materials and methods

Patients

Details of the study protocol have been published elsewhere [6, 10]. We identified consecutive patients presenting with acute abdominal pain for more than 2 h and less than 5 days at the emergency department (ED) of two university and four (large) teaching hospitals. Patients discharged from the ED by the treating physician without any diagnostic imaging (ultrasound, CT or plain radiographs), patients under 18 years, pregnant women, patients with a blunt or penetrating trauma, patients with distinctive flank pain, suspected with renal colic, as well as patients in haemorrhagic shock caused by a gastrointestinal bleeding or acute abdominal aneurysm were not invited. Two of the teaching hospitals included patients from Monday to Friday between 9 am and 5 pm. In all other hospitals, patients were included 7 days a week from 8 am until 11 pm.

Eligible patients were invited to the study after being informed orally about the study by the treating physician. An information brochure was provided to them. Consenting patients were included in the study. This study had been approved by the Institutional Review Boards of participating hospitals before its initiation.

All included patients were clinically evaluated at the ED by the treating physician, usually a surgical or emergency medicine resident, after which the patients underwent a full diagnostic protocol. The treating physician prospectively recorded patients' characteristics and the findings of clinical history and examination in a case record form.

Observers

After clinical assessment at the ED, all consenting patients underwent ultrasound and computed tomography (CT) within a few hours of presentation to the ED. Ultrasound and CT were independently evaluated by two different blinded observers. Between 5 pm and 11 pm, when often only one attending radiologist or radiological resident was present, both ultrasound and CT were evaluated by the same observer. The ultrasound examination was performed and evaluated by the observers: the attending radiologist or radiological resident, not by a sonographer. To guarantee a blinded evaluation for study purposes, ultrasound was performed first and documented in the case record form. CT was only evaluated after finalising the ultrasound part of the case record form.

The CT findings with immediate treatment consequences were communicated to the treating physician. In cases presenting after hours, CT examinations were re-evaluated by an abdominal radiologist the next morning and these findings were documented in the case record form. This radiologist was blinded to the ultrasound evaluation and had access to the same details on clinical findings as the person evaluating the ultrasound examination. This second reading was used for this comparative study, so all CT examinations were read or supervised by a radiologist. Contrary to ultrasound examinations, which were performed by radiological residents alone after hours. To evaluate the effects of experience, all observers were asked to record the number of abdominal ultrasounds they had performed (<100, 100–500, 500–1,000, 1.000–5.000, 5.000–10.000 or >10.000 examinations).

Ultrasound

To standardise the ultrasound examination, a general survey of the abdomen was performed and findings were recorded on a digital case record form. In this case record form, the following general image characteristics and specific radiological features were recorded: image quality, visualisation of the painful quadrant (quadrant of interest), infiltration of mesenteric fat (hyperechoic tissue), free fluid, abscess, free intra-peritoneal air and fistulas. Image characteristics were assessed per organ: gallbladder, bile duct, liver, pancreas, appendix, gastro-intestinal tract, lymph nodes, vascular system, kidneys, and if appropriate, the female reproductive system. In the case of abnormalities further specification on the observed abnormality was warranted. All observers recorded an ultrasound diagnosis. Observers assigned their diagnoses based on the imaging findings in combination with the clinical information provided by

the treating physician, no specific set of criteria was provided per diagnosis, reflecting daily practice. Ultrasound cases in which the quadrant of interest could not be visualised, were considered examinations with low quality.

Computed tomography

Different types of CT were used in the participating centres, varying from 4- to 16-slice or more CT (Table 1). All patients received intravenous contrast medium; no oral or rectal contrast agents were used. In 16 (1.6%) patients an unenhanced CT was performed because of known renal failure ($n=14$); Or known previous reaction to contrast agents ($n=2$).

The CT was evaluated in the same standardised way as the ultrasound examinations. Approximately the same general image findings and specific radiological features as at ultrasound were assessed for CT and recorded on a digital case record form: image quality, fat infiltration, free fluid, abscess, free intraperitoneal air and fistulas. Image assessment per organ: gallbladder, bile duct, liver, pancreas, appendix, gastrointestinal tract, lymph nodes, vascular system, kidneys, and if appropriate, female genitalia. If no abnormalities were recorded, no specification was asked, but in the case of abnormalities further specification on the observed abnormality was warranted, a CT diagnosis was recorded. Comparable to ultrasound, no specific set of criteria was provided per diagnosis to assist observers in assigning their diagnosis.

Reference standard

A final diagnosis was assigned after 6 months by an independent expert panel, consisting of two experienced gastrointestinal surgeons and an experienced abdominal radiologist (Appendix II) [6, 10]. Members of this panel individually evaluated all available data for each patient, including initial clinical, laboratory and imaging findings, as well as additional clinical, laboratory, imaging findings and if applicable, surgical and histopathological findings, and in and out-patient follow-up for at least 6 months. This information was provided to the expert panel in a standardised way. In case of disagreement, consensus was reached in a group discussion.

Analysis

The primary analysis was focused on a comparison of the accuracy of ultrasound and CT in detecting common diagnoses in patients with acute abdominal pain at the ED, using the final diagnosis as the reference standard. The sensitivity, specificity, positive and negative predictive

Table 1 Imaging characteristics

N	Computed tomography				Ultrasound	
	Type of system	Slice thickness	i.v. contrast (ml)	Imaging dose	Convex Mhz	Linear Mhz
279	MDCT	3 mm	125	120 Kv, 165 mAs	4-5	7-8
32	MDCT	1.5	100	140 Kv, 200 mAs	5-2	12.5
285	MDCT	6.5	120	120 Kv, 165 mAs	8-5 en 5-2	12-5
180	MDCT	3	100	120 Kv, 165 mAs	5-2	12-5
108	MDCT	3	120	120 Kv, 80–140 mAs	5-2	12-5
137	MDCT	5 mm, 4 mm[a]	120	120 Kv, 200–250 mAs[b]	5-2	4-7 and 5-12

[a] Slice thickness was 5 or 4 mm at the PACS, and 1 mm at the CT workstation

[b] Dose adaptation was used

values for ultrasound and CT were calculated. Differences in sensitivity and specificity between ultrasound and CT were evaluated with McNemar's test statistic. Differences between ultrasound and CT with regard to predictive values were evaluated with the Chi-squared test statistic.

The percentage of diagnoses missed at ultrasound in patients in whom image quality was sufficient (patients in whom the quadrant of interest was visualised) was compared with the percentage of missed cases with insufficient image quality. The Chi-squared test statistic for unpaired data was used to test differences for statistical significance. The percentage of diagnoses missed was calculated as the number of false-negatives relative to the number of patients with the corresponding diagnosis as the final diagnosis (1-sensitivity).

As patient characteristics could influence the accuracy of ultrasound, potential differences in sensitivity between patient groups were evaluated. Patient subgroups were defined by sex, age, body mass index and duration of symptoms. In addition, sensitivity and predictive values of ultrasound in attending radiologists including supervised residents were compared with those of unsupervised residents. Unsupervised residents who had performed and evaluated less than 500 ultrasound examinations were compared with unsupervised residents who had performed and evaluated more than 500 ultrasound examinations. Subgroup differences were evaluated with Chi-squared test statistics.

For all comparisons p values less than 0.05 were taken to indicate statistically significant differences. All analyses were performed in SPSS 15.0.1 (SPSS Inc. Chicago, IL, USA)

Results

Patients

Between March 2005 and November 2006, 1,101 patients were included. Case record forms were incomplete for 80

patients (7.3%); these were excluded from the analysis. The remaining 1,021 patients had a mean age of 47 years (range 19–94); 484 (47%) were younger than 45 years, 258 (25%) were older than 65 years, 565 (55%) were female, 157 (15.4%) had a body mass index over 30, 320 (31%) had prolonged 'acute' abdominal pain for (more than 2 days but still less than 5 days), and 705 (69%) a body temperature exceeding 38°C.

Consensus on the final diagnosis was reached after individual evaluation in 76% of the patients; in 24% (244) the expert panel needed a group discussion to reach consensus. A list of the final diagnoses in the study group is provided in Appendix III. The most frequent final diagnoses were acute appendicitis, acute diverticulitis, bowel obstruction and acute cholecystitis. Urgent gynaecological disorders (n=27) consisted of pelvic inflammatory disease (13), ovarian torsion (9), rupture or bleeding ovarian cyst (5).

Sensitivity

The sensitivity in detecting acute appendicitis and acute diverticulitis differed significantly between ultrasound and CT (both $p<0.01$): ultrasound sensitivity in detecting acute appendicitis was 76% versus 94% for CT. Ultrasound sensitivity for acute diverticulitis was 61% versus 81% on CT (Table 2). For urgent gynaecological disorders the sensitivity was also significantly higher for CT than for ultrasound: 67% versus 37% (p=0.04). Likewise, the sensitivity in detecting inflammatory bowel disorders was higher for CT than for ultrasound (p=0.05). For acute cholecystitis and bowel obstruction sensitivity did not differ significantly between ultrasound and CT (p=1.00 and 0.57, respectively (Table 2).

Predictive values

Positive predictive values did not differ significantly in detecting acute appendicitis and acute diverticulitis between

Table 2 Sensitivity, specificity, positive and negative predictive values for US and CT in patients with acute abdominal pain at the emergency department

Diagnoses	N	Sensitivity US (%)	Sensitivity CT(%)	p values	Specificity US (%)	Specificity CT (%)	p value*
Appendicitis	284	76 (71–81)	94 (92–97)	<0.01*	95 (94–97)	95 (94–97)	1.00
Diverticulitis	118	61 (52–70)	81 (74–88)	<0.01*	99 (99–100)	99 (98–99)	0.42
Bowel obstruction	68	63 (52–75)	69 (58–80)	0.57	99 (99–100)	99 (99–100)	1.00
Gastrointestinal non-urgent[a]	56	27 (15–38)	36 (23–48)	0.38	99 (98–100)	99 (98–100)	0.36
Cholecystitis	52	73 (61–85)	73 (61–85)	1.00	97 (96–98)	98 (97–99)	0.73
Hepatic-pancreatic-biliary disease[b]	43	65 (51–79)	47 (32–61)	0.08	98 (97–99)	98 (97–99)	0.28
Inflammatory bowel disorder[c]	30	37 (19–54)	67 (50–79)	0.05	97 (96–98)	98 (98–99)	0.07
Pancreatitis	28	39 (21–57)	68 (51–85)	0.08	100 (99–100)	100 (99–100)	1.00
Gynaecological urgent[d]	27	41 (23–50)	70 (54–86)	0.04*	98 (98–99)	98 (97–99)	0.31

Diagnoses	N	PPV US	PPV CT	p value	NPV US	NPV CT	p value*
Appendicitis	284	86 (81–90)	89 (85–92)	0.35	91 (89–93)	98 (97–99)	<0.01*
Diverticulitis	118	90 (83–97)	89 (83–95)	0.81	95 (94–97)	98 (97–99)	<0.01*
Bowel obstruction	68	86 (76–96)	86 (76–95)	0.94	97 (96–98)	98 (97–99)	0.56
Gastrointestinal non-urgent[a]	56	81 (70–92)	78 (66–89)	0.69	98 (98–9)	99 (98–99)	0.72
Cholecystitis	52	37 (22–51)	51 (36–67)	0.19	96 (95–97)	96 (95–98)	0.56
Hepatic-pancreatic-biliary disease[b]	43	54 (40–67)	54 (38–70)	0.99	99 (98–99)	98 (97–99)	0.21
Inflammatory bowel disorder[c]	30	30 (15–45)	57 (41–74)	0.02*	98 (97–100)	99 (98–100)	0.09
Pancreatitis	28	73 (51–96)	83 (67–98)	0.69	98 (98–99)	99 (99–100)	0.12
Gynaecological urgent[d]	27	37 (19–55)	51 (36–67)	0.57	98 (97–99)	99 (98–100)	0.27

* p values <0.05 were considered significant

[a] Gastrointestinal disorder non-urgent (n=56), consisted of gastroenteritis (n=27), constipation (n=12), epiploic appendagitis/omental infarction (n=11), gastritis (n=5), ulcus ventriculi/duodeni (n=1)

[b] HPB (n=43) consisted of, cholecystolithiasis (n=33), choledocholithiasis (n=5), hepatitis (n=3), liver metastases (n=1), chronic pancreatitis (n=1)

[c] Inflammatory bowel disorder consisted of: non-specified inflammatory bowel disorder (n=16); infectious (n=11), Crohn's disease (n=1), ulcerative colitis (n=2)

[d] Urgent gynaecological disorder (n=27) consisted of Pelvic Inflammatory Disease (PID) (n=13), adnexal torsion (n=9), bleeding/rupture ovarian cyst (n=5)

ultrasound and CT (Table 2). Positive predictive values for a final diagnosis of inflammatory bowel disorder were significantly higher with CT ($p=0.02$). The negative predictive values for acute appendicitis and acute diverticulitis were significantly higher for CT (both $p<0.01$).

Insufficient ultrasound image quality

Significantly fewer cases of acute appendicitis and of acute diverticulitis were missed in patients in whom the radiologist stated that image quality was sufficient compared with cases in which image quality was insufficient (Table 3). For all other diagnoses, the percentage of diagnoses missed with ultrasound was not significantly lower in patients with sufficient image quality compared with those with insufficient image quality (Table 3).

Patient characteristics and missed diagnoses

The percentage of acute appendicitis and acute diverticulitis cases missed by ultrasound did not differ significantly in patient subgroups defined by sex, body mass index, duration of pain, or age (Table 4).

Observers

In the six participating hospitals, ultrasound was evaluated by 107 different observers and CT was evaluated by 88 different observers, ranging from first-year radiology residents to a radiologist with more than 30 years of experience. Residents evaluated 582 (57%) of the ultrasound examinations, of which 282 were read after hours (28%), the latter not being supervised by radiologists. Of these non-supervised ultrasound examinations, 187 were performed by residents who had evaluated and performed more than 500 abdominal ultrasound examinations, and 95 were performed by residents who had evaluated and performed less than 500 abdominal ultrasound examinations. Radiologists evaluated 439 (43%) of the ultrasound examinations. CT were evaluated by supervised residents in 299 patients (29%); in 722 patients (71%) CT were evaluated by radiologists.

The sensitivity of ultrasound for acute appendicitis and acute cholecystitis was somewhat lower—with no significant difference—for unsupervised residents compared with attending radiologists including supervised residents: 73% versus 78% ($p=0.33$) and 60% versus 62% ($p=0.43$), respectively (Fig. 1).

Ultrasound sensitivity in detecting acute appendicitis and acute diverticulitis

There were no significant differences between unsupervised residents who had evaluated (and performed) more than 500 ultrasound examinations and those who had evaluated less than 500 ultrasound examinations for these two diagnoses (Table 5). Unsupervised residents had a higher sensitivity than attending radiologists, including supervised

Table 3 Sensitivity of ultrasound with sufficient image quality versus insufficient image quality

Diagnoses	N	Missed diagnoses sufficient image quality[a] (%)	N	Missed diagnoses insufficient image quality[a] (%)	p value
Appendicitis	241	16 (11–20)	43	67 (53–81)	<0.01
Diverticulitis	96	30 (21–39)	22	77 (57–90)	<0.01
Bowel obstruction	37	32 (17–48)	31	42 (26–59)	0.46
Gastrointestinal Non-Urgent[b]	38	71 (57–85)	18	78 (55–91)	0.75
Cholecystitis	45	22 (10–34)	7	57 (25–84)	0.08
Hepatic-pancreatic-biliary disease[c]	31	29 (13–45)	12	50 (25–75)	0.29
Inflammatory bowel disorder[d]	21	52 (31–74)	9	89 (56–98)	0.10
Pancreatitis	11	45 (16–75)	17	71 (47–87)	0.25
Gynaecological urgent[e]	30	53 (30–75)	8	88 (53–98)	0.19

[a] Insufficient image quality is defined as ultrasound examinations in which the region of interest could not be visualised

[b] gastrointestinal disorder non-urgent ($n=56$), gastroenteritis ($n=27$), constipation ($n=12$), epiploic appendagitis/omental infarction ($n=11$), gastritis ($n=5$), ulcus ventriculi/duodeni ($n=1$)

[c] HPB ($n=43$) consisted of; cholecystolithiasis ($n=33$), choledocholithiasis ($n=5$), hepatitis ($n=3$), liver metastases ($n=1$), chronic pancreatitis ($n=1$)

[d] Inflammatory bowel disorder consisted of: non-specified inflammatory bowel disorder ($n=16$); infectious ($n=11$), Crohn's disease ($n=1$), ulcerative colitis ($n=2$)

[e] Urgent gynaecological disorder ($n=27$) consisted of Pelvic Inflammatory Disease (PID) ($n=13$), adnexal torsion ($n=9$), bleeding/rupture ovarian cyst ($n=5$)

Table 4 Missed diagnoses of appendicitis and diverticulitis at ultrasound

Patient characteristics	Appendicitis			Diverticulitis		
	N	Missed (%)	p value	N	Missed (%)	p value
Female	121	27	0.21	65	43	0.31
Male	163	21		53	34	
BMI >30	29	21	0.70	19	26	0.22
BMI <30	255	24		99	41	
BMI >30 female	14	29	0.39	7	43	0.31
BMI >30 male	15	13		12	17	
Duration pain >2 days	214	22	0.42	39	33	0.38
Duration pain <2 days	70	27		79	42	
Age <45	111	22	0.53		n.a.	
Age >45	173	25			n.a.	
Age <60		n.a.		73	40	0.32
Age >60		n.a.		45	38	

n.a. not applicable

residents for the diagnosis of diverticulitis with ultrasound, 83% versus 57% (*p*=0.04). Here, the sensitivity was significantly higher for more experienced unsupervised residents (Table 5).
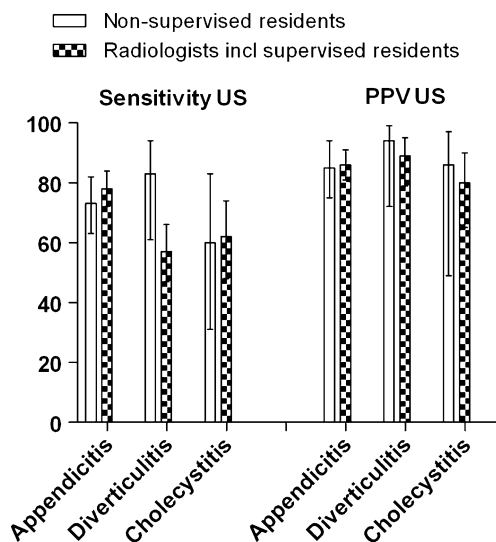
Positive predictive values for common diagnoses such as acute appendicitis, acute diverticulitis and acute cholecystitis were comparable for non-supervised residents and attending radiologists, including supervised residents (Fig. 1).

Discussion

In this study we found that the sensitivity of CT was significantly higher than that of ultrasound in detecting appendicitis and diverticulitis. Fewer cases of acute

Fig. 1 Comparison of sensitivity and positive predictive value (PPV) for subgroups of observers

appendicitis and acute diverticulitis were missed by CT, but positive predictive values of ultrasound and CT were comparable. For acute cholecystitis and bowel obstruction there were no significant differences in accuracy between ultrasound and CT. No subgroup differences in ultrasound sensitivity in detecting acute appendicitis and acute diverticulitis were found for any of the evaluated patient characteristics: BMI, age and duration of pain. There were no statistically significant differences between obese women and men. The sensitivity of ultrasound performed by non-supervised radiological residents was not significantly lower than that of ultrasound performed by attending radiologists, including supervised residents. The percentage of missed acute appendicitis and acute diverticulitis cases was lower if the observer was able to visualise the region of interest compared with the percentage of missed cases of acute appendicitis or diverticulitis with insufficient image quality. For all other diagnoses, such a reduction in the number of missed diagnoses was not found.

A number of potential limitations of this analysis should be acknowledged. One could object that the sensitivity of US was underestimated, because ultrasound was partly performed and interpreted by unsupervised radiological residents. Unsupervised residents did not have a significantly lower sensitivity in detecting disease in this study compared with attending radiologists. In a previous study, the overall sensitivity of ultrasound performed by unsupervised residents for detecting urgent diagnoses was significantly lower than that of ultrasound performed by attending radiologists, without a significant difference in positive predictive value [6], indicating that residents more often missed an urgent diagnosis. Whenever an urgent diagnosis was assigned, however, this was most likely correct. In a study by Hertzberg et al. training in ultrasound was evaluated and a significant improvement

Table 5 Comparison of ultrasound accuracy per diagnosis for observers with different ultrasound experience

Ultrasound experience per diagnosis	Sensitivity (CI)[a]	p value sensitivity*	PPV (CI)[a]	p value PPV*
Appendicitis				
<500 US experience	0.64 (0.44–0.84)	0.27	0.82 (0.64–1.00)	0.70
>500 US experience	0.76 (0.65–0.87)		0.86 (0.77–0.96)	
Diverticulitis				
<500 US experience	0.50 (0.18–0.81)	0.03	1.00 (0.44–1.00)	1.00
>500 US experience	1.00 (0.76–1.00)		0.92 (0.67–0.99)	
Cholecystitis				
<500 US experience	1.00 (0.34–1.00)	0.47	1.00 (0.34–1.00)	1.00
>500 US experience	0.50 (0.22–0.79)		0.80 (0.38–0.96)	

* p values <0.05 were considered significant

[a] CI: confidence interval

was found at between 50 and 200 cases [11]. In the present study 23% of the observers had performed fewer than 500 abdominal ultrasound examinations, but only 4% had performed fewer than 100 ultrasound examinations.

Comparisons of CT accuracy between residents and radiologists or between CT reading after hours and during daytime were not considered meaningful, because residents were always supervised by a radiologist during daytime. The diagnosis recorded on the case record form by the supervised resident, for both CT and ultrasound, can be considered as a consensus diagnosis. CT scans of patients evaluated after hours were always re-evaluated the next day by a radiologist. For radiologists inter-observer agreement for abdominal CT is known to be good [9].

This study was aimed at evaluating ultrasound and CT in daily practice in six institutions. A considerable number of observers contributed, with a wide variety of experience. Although one could object that this may have negatively influenced accuracy, our study probably reflects daily practice better than studies where all patients were evaluated by one or two very experienced observers. It is a well known phenomenon that the diagnostic accuracy reported in the literature can be higher than that in an average hospital, not only because tests in research settings are often evaluated by experienced observers, but also because standardised record forms are used in studies to minimise the number of indeterminate findings [12].

With this study no specific set of criteria was provided to the observers from which a diagnosis was supposed to be made. Instead the observers assigned their ultrasound or CT diagnoses based on imaging findings in combination with the clinical information provided by the treating physician. This way of evaluating imaging examinations reflects daily practice.

We relied on an expert panel to assign the final diagnoses. This clinical reference standard may imply a form of incorporation bias, as the experts had access to all available information, including imaging findings. In this study population, with a wide variety of possible diagnoses, it is impossible to use a single reference standard, and the use of a panel is an appropriate alternative in a setting with multiple possible underlying diseases [13]. Our experts had access to extensive clinical information, including follow-up. A final diagnosis of acute appendicitis was based on histopathology in 95% of the cases, while the remaining 5% had undergone conservative therapy or percutaneous drainage of peri-appendiceal abscess.

In discordance with previous studies [6, 14], we did not find a significantly lower accuracy for residents compared with radiologists. One of the previous studies also demonstrated a significantly lower sensitivity of ultrasound in female patients compared with males with suspected appendicitis [14]. In our study, we did not see such a difference in sensitivity. Nor did we detect a significant difference between obese and non-obese patients in acute appendicitis cases and acute diverticulitis cases missed with ultrasound, although the number was markedly higher in obese women. It is a known limitation of ultrasound that it has difficulty in penetrating fat. Because ultrasound is a real-time examination not all obese patients are a priori unsuitable for ultrasound examination. In patients with a large proportion of extra-mesenteric fat ultrasound images can more often be interpreted adequately.

All patients underwent the same CT protocol for better evaluation of the accuracy of CT in patients with acute abdominal pain. If CT protocols had been tailored to the clinically suspected diagnosis [6], bias would have been introduced and a valid comparison of CT and

ultrasound would not have been possible. Recent research has shown that usage of oral contrast agent does not increase the accuracy of diagnosing appendicitis with CT [15, 16]. For the evaluation of acute diverticulitis a wide variety of CT protocols is described in the literature, ranging from solely intravenous contrast to a combination of oral, rectal and intravenous contrast. The CT protocol solely using iv contrast agent did not achieve lower accuracy values compared with studies with extended contrast agent usage [8].

We observed a low prevalence in our study group of a number of important disorders, such as perforated viscus or bowel ischaemia and other common diagnoses causing acute abdominal pain such as pancreatitis and urinary tract calculus (patients with distinctive flank pain, suspected with renal colic, were not eligible for this study). This low prevalence limited any comparison of CT or ultrasound accuracy for the full range of diagnoses in patients presenting with acute abdominal pain.

The study reported here was not designed to separately evaluate the sensitivity and specificity of specific complications of any of the diagnoses causing acute abdominal pain. We only aimed to study the accuracy of ultrasound and CT in assigning the correct diagnosis.

A meta-analysis did not show any significant difference in accuracy between ultrasound and CT in detecting diverticulitis, although CT is more likely to detect complications of acute diverticulitis [8]. We did not find a significant difference in the accuracy of detecting bowel obstruction between ultrasound and CT; the aetiology of the obstruction is better evaluated with CT than with US. Likewise, a better accuracy for CT has been described in detecting complicated bowel obstruction [17–21], although the accuracy of CT in the detection of bowel ischaemia is at best mediocre [22].

Some of the accuracy estimates for ultrasound in this study are lower than those reported elsewhere in the literature. The reported sensitivities for ultrasound in experienced hands in detecting appendicitis have been as high as 90% [23]. In recent meta-analyses of diagnostic imaging in acute appendicitis, ultrasound sensitivity varied between 86% [24] and 78% [7], which is comparable to the estimates in the present study. The accuracy in detecting acute diverticulitis is lower than in the aforementioned recent meta-analysis. Summary sensitivity of 92% for ultrasound was reported, which is much higher than the sensitivity of 68% [8]. The most likely explanation for this difference might be that we included unselected patients with acute abdominal pain, whereas the studies included in the meta-analysis more often had recruited selected patients with a clinically suspected acute diverticulitis. A higher pre-test likelihood of disease is known to result in a higher accuracy [25].

We observed the significantly higher sensitivity of CT compared with ultrasound with regard to urgent gynaecological disorders. This result may be counterintuitive to some as ultrasound is the imaging technique of choice in these patients [26]. Our findings may be explained by the fact that we used abdominal ultrasound performed by radiologists, not trans-vaginal ultrasound performed by the gynaecologist. Gynaecologists can be expected to be more experienced in the evaluation of gynaecological disorders; they can probably achieve a higher sensitivity with trans-vaginal ultrasound than radiologists can with transabdominal ultrasound. Unfortunately patients directly referred to gynaecologists are not routed through the emergency department and therefore not included in this study.

In summary, we observed that CT sensitivity is higher than that of ultrasound in detecting appendicitis and diverticulitis in unselected patients presenting with acute abdominal pain, but positive predictive values are comparable. Accuracy of bowel obstruction and acute cholecystitis were not significantly different. The percentage of cases missed on ultrasound was not influenced by patient characteristics and observer experience at large with regard to common diagnoses. The proportion of missed acute appendicitis and acute diverticulitis was significantly lower in the subgroup of patients in whom the radiologist could adequately visualise the region of interest. These results indicate that ultrasound is a good first-line technique.

Appendix I

Members of the OPTIMA study group:
  Academic Medical Center, Amsterdam
  A. van Randen, Department of Radiology
  W. Laméris, Department of Surgery
  J. Stoker, Department of Radiology
  M.A. Boermeester, Department of Surgery
  P.M.M. Bossuyt, Department of Clinical epidemiology, Biostatistics, and Bioinformatics
  St Antonius Hospital Nieuwegein
  B. van Ramshorst, Department of Surgery
  J.P.M. van Heesewijk, Department of Radiology
  M.P. Gorzeman, Department of Emergency Medicine
  Gelre Hospitals, Apeldoorn
  W.H. Bouma, Department of Surgery

W. ten Hove, Department of Radiology

J. Winkelhagen, Department of Surgery

University Medical Center Utrecht

H.G. Gooszen, Department of Surgery,

M.S. van Leeuwen, Department of Radiology

D.E.J.G.J. Dolmans, Department of Surgery

Tergooi Hospitals, Hilversum

E. van Keulen, Department of Radiology

J.W. Juttmann, Department of Surgery

M.J. van der Laan, Department of Surgery

Onze Lieve Vrouwe Gasthuis, Amsterdam

S.C. Donkervoort, Department of Surgery

V.P.M. van der Hulst, Department of Radiology

Appendix II

OPTIMA trial expert panel members:

Academic Medical Center

O.R.C. Busch, Department of Surgery

T.M. van Gulik, Department of Surgery

O.D. Henneman, Department of Radiology, Bronovo Hospital, Den Haag

Tergooi Hospitals

A.A.W. van Geloven, Department of Surgery, Tergooi Hospitals, Hilversum

J.W. Juttmann, Department of Surgery, Tergooi Hospitals, Hilversum

E. van Keulen, Department of Radiology, Tergooi Hospitals, Hilversum

Onze Lieve Vrouwe Gasthuis

S.C. Donkervoort, Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

M.P. Simons, Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

J. Peringa, Department of Radiology, Onze Lieve Vrouwe Gasthuis, Amsterdam

St Antonius Hospital Nieuwegein

H.W. van Es, Department of Radiology, *St* Antonius Hospital, Nieuwegein

P.M.N.Y.H Go, Department of Surgery, *St* Antonius Hospital, Nieuwegein

M.J. Wiezer, Department of Surgery, *St* Antonius Hospital, Nieuwegein

Gelre Hospitals

W.H. Bouma, Department of Surgery, Gelre Hospitals, Apeldoorn

E.J. Hesselink, Department of Surgery, Gelre Hospitals, Apeldoorn

W. ten Hove, Department of Radiology, Gelre Hospitals, Apeldoorn

Appendix III [6]

Table 6 Final diagnoses in 1021 patients assigned by the expert panel

Diagnoses	N	%
Acute appendicitis	284	27.8
Non-specific abdominal pain[a]	183	17.9
Acute diverticulitis	118	11.6
Bowel obstruction	68	6.7
Gastro-intestinal disorder non-urgent	56	5.5
Acute cholecystitis	52	5.1
HPB[b]	43	4.2
Inflammatory bowel disorder	30	2.9
Acute pancreatitis	28	2.7
Gynaecological disorder; urgent	27	2.6
Urinary tract disorder; urgent	22	2.2
Urinary tract disorder	20	0.2
Abscess	14	1.4
Perforated viscus	13	1.3
Bowel ischaemia	12	1.2
Pneumonia	11	1.1
Gynaecological disorder; non-urgent	9	0.9
Retro-peritoneal or abdominal wall bleeding	9	0.9
Malignancy	5	0.5
Acute peritonitis[c]	3	0.3
Herniation[d]	2	0.2
Other[e]	12	1.2
Total	1,021	100

[a] non-specific abdominal pain was abbreviated as NSAP, which is not truly a diagnosis but merely a negative patient, without disease

[b] 33 cholecystolithiasis, 5 common bile duct stones, 3 hepatitis, 1 liver metastasis

[c] Peritonitis not caused by perforation

[d] Hernia without strangulation, otherwise it would have been classified as bowel ischaemia

[e] Other diagnoses were abdominal wall infiltration, oesophagitis (2), renal infarction (2), gastric band problem (2), SLE, mesenteric lymphadenitis, post-procedural pain, uterine haemorrhage and a testicular torsion

References

1. Stoker J, van Randen A, Laméris W, Boermeester MA (2009) Imaging patients with acute abdominal pain. Radiology 253:31–46

2. Shuman WP, Ralls PW, Balfe DM et al (2000) Imaging evaluation of patients with acute abdominal pain and fever. American College of Radiology. ACR Appropriateness Criteria. Radiology 215(Suppl):209–212

3. Puylaert JB (2003) Ultrasonography of the acute abdomen: gastrointestinal conditions. Radiol Clin North Am 41:1227–1242, vii

4. The 2007 Recommendations of the International Commission on Radiological Protection (2007) ICRP publication 103. Ann ICRP 37:1–332

5. Stoker J (2008) Magnetic resonance imaging and the acute abdomen. Br J Surg 95:1193–1194

6. Laméris W, van Randen A, van Es HW et al (2009) Imaging strategies for detection of urgent conditions in patients with acute abdominal pain: diagnostic accuracy study. BMJ 338:b2431

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

8. Laméris W, van Randen A, Bipat S, Bossuyt PM, Boermeester MA, Stoker J (2008) Graded compression ultrasonography and computed tomography in acute colonic diverticulitis: meta-analysis of test accuracy. Eur Radiol 18:2498–2511

9. van Randen A, Laméris W, Nio CY et al (2009) Inter-observer agreement for abdominal CT in unselected patients with acute abdominal pain. Eur Radiol 19:1394–1407

10. Laméris W, van Randen A, Dijkgraaf MG, Bossuyt PM, Stoker J, Boermeester MA (2007) Optimization of diagnostic imaging use in patients with acute abdominal pain (OPTIMA): design and rationale. BMC Emerg Med 7:9

11. Hertzberg BS, Kliewer MA, Bowie JD, Carroll BA, DeLong DH, Gray L et al (2000) Physician training requirements in sonography: how many cases are needed for competence? AJR Am J Roentgenol 174:1221–1227

12. Cuschieri J, Florence M, Flum DR et al (2008) Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. Ann Surg 248:557–563

13. Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM (2007) Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess 11(50)

14. Gaitini D, Beck-Razi N, Mor-Yosef D et al (2008) Diagnosing acute appendicitis in adults: accuracy of color Doppler sonography and MDCT compared with surgery and clinical follow-up. AJR Am J Roentgenol 190:1300–1306

15. Anderson SW, Soto JA, Lucey BC et al (2009) Abdominal 64-MDCT for suspected appendicitis: the use of oral and IV contrast material versus IV contrast material only. AJR Am J Roentgenol 193:1282–1288

16. Gurusamy K, Samraj K, Gluud C, Wilson E, Davidson BR (2010) Meta-analysis of randomized controlled trials on the safety and effectiveness of early versus delayed laparoscopic cholecystectomy for acute cholecystitis. Br J Surg 97:141–150

17. Hainaux B, Agneessens E, Bertinotti R et al (2006) Accuracy of MDCT in predicting site of gastrointestinal tract perforation. AJR Am J Roentgenol 187:1179–1183

18. Lazarus DE, Slywotsky C, Bennett GL, Megibow AJ, Macari M (2004) Frequency and relevance of the "small-bowel feces" sign on CT in patients with small-bowel obstruction. AJR Am J Roentgenol 183:1361–1366

19. Maglinte DD, Howard TJ, Lillemoe KD et al (2008) Small-bowel obstruction: state-of-the-art imaging and its role in clinical management. Clin Gastroenterol Hepatol 6:130–139

20. Schmutz GR, Benko A, Fournier L, Peron JM, Morel E, Chiche L (1997) Small bowel obstruction: role and contribution of sonography. Eur Radiol 7:1054–1058

21. Silva AC, Pimenta M, Guimarães LS (2009) Small bowel obstruction: what to look for. Radiographics 29:423–439

22. Sheedy SP, Earnest F, Fletcher JG, Fidler JL, Hoskin TL (2006) CT of small-bowel ischemia associated with obstruction in emergency department patients: diagnostic performance evaluation. Radiology 241:729–736

23. Puylaert JB, Rutgers PH, Lalisang RI et al (1987) A prospective study of ultrasonography in the diagnosis of appendicitis. N Engl J Med 317:666–669

24. Terasawa T, Blackmore CC, Bent S, Kohlwes RJ (2004) Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med 141:537–546

25. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH (2008) Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 54:729–737

26. Potter AW, Chandrasekhar CA (2008) US and CT evaluation of acute pelvic pain of gynaecologic origin in nonpregnant premenopausal patients. Radiographics 28:1645–1659, Review