

Accuracy of Surgeon-Performed Ultrasound in Detecting Gallstones: A Validation Study

Camilla Gustafsson¹ · Andrea McNicholas² · Anders Sondén¹ · Staffan Törngren² · Hans Järnbert-Pettersson² · Anna Lindelius¹

Published online: 2 March 2016
© Société Internationale de Chirurgie 2016

Abstract

Background Symptomatic gallstone disease is a common diagnosis in patients with abdominal pain. Ultrasound is considered the gold standard method to identify gallstones. Today the examination may be performed bedside by the treating clinician. Bedside ultrasound could provide a safe and time-saving diagnostic resource for surgeons evaluating patients with suspected symptomatic gallstones; however, large validation studies of the accuracy and reliability are lacking. The aim of this study was to prospectively investigate the accuracy of surgeon-performed ultrasound for the detection of gallstones.

Methods Between October 2011 and November 2012, 179 adult patients, with an acute or elective referral for an abdominal ultrasound examination, were examined with a right upper quadrant ultrasound scan by a radiologist as well as a surgeon. The surgeons had undergone a four-week-long ultrasound education before participating in the study. Ultrasound findings of the surgeon were compared to those of the radiologist, using radiologist-performed ultrasound as reference standard.

Results Surgeon-performed ultrasound agreed with radiologist findings in 169 of 179 patients regarding the detection of gallstones, providing an accuracy of 94 %. The sensitivity was 88 % (67/76), specificity 99 % (102/103), positive predictive value 99 % (67/68), and negative predictive value 92 % (102/111). Agreement between the diagnosis set by the radiologists and the surgeons was high: Cohen's Kappa coefficient = 0.88.

Conclusions Ultrasound-trained surgeons may accurately diagnose gallstones using ultrasound and reach a high level of agreement with radiologists.

Trial registration: The study was registered at clinicaltrials.gov.
Registration number: NCT02469935.

✉ Camilla Gustafsson
camilla.gustafsson@ki.se

Andrea McNicholas
andrea.mcnicholas@gmail.com

Anders Sondén
anders.sonden@sodersjukhuset.se

Staffan Törngren
staffan.torngren@comhem.se

Hans Järnbert-Pettersson
hans.jarnebert-pettersson@sodersjukhuset.se

Anna Lindelius
anna.lindelius@sodersjukhuset.se

¹ Department of Clinical Science and Education, Department of Surgery, Karolinska Institutet, Södersjukhuset (Stockholm South General Hospital), Sjukhusbacken 10, 118 83 Stockholm, Sweden

² Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset (Stockholm South General Hospital), Sjukhusbacken 10, 118 83 Stockholm, Sweden

Introduction

Symptomatic gallstone disease is one of the major causes of acute abdominal pain among adults and ultrasound (US) is considered the gold standard for diagnosis [1, 2].

Radiologist US is not always accessible in the emergency department (ED), especially outside regular office hours, which can lead to unnecessary delay in patient management [3]. Consequently, non-radiologist-performed US (or point-of-care US), at the patient's bedside, has increased during the last two decades [4, 5]. Specialists with a longer experience of systematic US use include cardiologists and obstetricians, but the development of portable, affordable and user-friendly machines has laid ground for a wider use in other specialties as well. Today emergency medicine physicians, anesthesiologists, as well as surgeons use US as a diagnostic tool [5]. A wide range of uses for surgeon-performed US has been reported, including traumatic conditions, diagnostic, and interventional procedures. Surgeons' diagnostic US includes examinations of the breast, thyroid gland, vascular system, and the gastrointestinal tract [4]. In the acute care setting, bedside US has been shown to help surgeons in their decisions concerning patients with abdominal pain [6, 7]. To ensure the quality of surgeon-performed US, there is a need for validation of the examinations. Some studies have previously shown high sensitivity as well as accuracy, but few with a large patient sample [3, 8, 9].

It has been shown that radiologist-performed US is a good method in detecting gallstones reaching high levels of sensitivity [2]. In a review of the literature between 1966 and 1992, Shea et al. found a total sensitivity of 97 % and a specificity of 95 % for ultrasound in finding gallstones [10].

In a systematic review from 2013, Carroll et al. made an attempt at pooling the numbers from several studies evaluating surgeon-performed US of the right upper quadrant (RUQ) [11]. However, there was a significant heterogeneity among existing validation studies regarding inclusion criteria, diagnostic criteria, definition of reference standard, and number of participating surgeons. Diagnostic criteria in the included studies ranged from the presence of gallstones or cholecystitis to any biliary tract disease, the latter often without further specification. Nevertheless, the pooled results suggested that surgeons become clinically capable of performing a RUQ scan after a short education in US.

Since 2004 Stockholm South General Hospital (Södersjukhuset) provides a 4-week-long training program in abdominal US for surgeons. In a large randomized study conducted at the same hospital, Lindelius et al. showed in 2008 that the US-trained surgeons reached a higher level of overall diagnostic accuracy in the ED, when using US as a part of their clinical examination [12]. A question that remained unanswered was how accurate the US

examinations performed by surgeons were. The purpose of this study was to validate surgeon-performed abdominal US compared with radiologist-performed abdominal US.

Materials and methods

Enrollment of patients

Three hundred patients, with an acute or elective referral to the radiology department at Stockholm South General Hospital, for any diagnostic abdominal US examination, were prospectively enrolled between October 2011 and November 2012. Eligible patients were identified in the radiology department by a study surgeon, including both patients admitted to in-hospital care and outpatients, and informed consent was obtained. Six US-educated surgeons participated in the enrollment of patients. Exclusion criteria were age <18 years or inability to communicate with the examiner. Referrals concerning metastases of the liver or contrast-enhanced examinations were considered not suitable for the study and were also excluded. The patients were enrolled consecutively if time allowed.

Data collection

Enrolled patients received one US examination by the study surgeon as well as the standard US examination by the on-duty radiologist. In a majority of cases, the two examinations were performed consecutively and the time interval between the surgeon-performed US and radiologist-performed US never exceeded 24 h. The surgeon's examination took place either before or right after the radiologist's examination. The examining surgeon and radiologist were blinded to each other's findings. The surgeon's US examination followed a standardized protocol, which included a full abdominal scan, regardless of the nature of the referral. The presence of gallstones was marked as a 'yes' (positive finding, regardless of number or size) or 'no' (negative finding) by the surgeon. In cases where a full abdominal scan could not be performed, due to urgent patient management, a focused examination based on the referral as well as a right upper quadrant (RUQ) scan was advised. The on-duty radiologist performed a standard care US focusing on the individual referrals. The radiologist's statement was collected from the patient's medical record and transferred to the study protocol by a separate radiologist, who was also blinded to the surgeon's examination. Among the radiologists, the major part of the scans was done by US-specialized radiologists with several years of training (73 % of the scans were performed by specialists in radiology and the remaining 27 % by radiologists in specialist training).

The surgeons used a portable US machine of the model LOGIQ e with a convex (1.6–4.6 MHz) or linear (5–13 MHz) transducer, GE Healthcare, WuXi, China. All the surgeons' scans were saved on a separate hard drive, which was kept together with the study protocol. The radiologists used Philips iU22 with a convex C5-1 or a linear L12-5 transducer.

US training of surgeons participating in the study

Six study surgeons, five in their final years of specialist training and one specialist in surgery, with limited or no previous US training, attended a 1-week course, comprising US physics, technique, anatomy, and hands-on training, led by specialists in US. After attending the course, the surgeons received three weeks of training in the radiology department under the guidance of an US specialist. The surgeons were expected to perform a minimum of 50 supervised scans, which were obtained in all cases. The training focused on detecting gallbladder stones, dilated bile ducts, thickened wall of the gallbladder, lesions in the liver parenchyma, hydronephrosis, abdominal aortic aneurysms, free abdominal fluid, and appendicitis. After the training was completed, each surgeon spent a minimum of 2 weeks enrolling and scanning patients during office hours in the hospital's radiology department.

Ethics

The patients received oral and written information from the study surgeon and were included after informed consent. The Ethical Review Board, at Karolinska Institutet, Stockholm, Sweden, approved the study.

Sample size

McNemar's test of paired proportions was used to detect a systematic difference between the radiologist and the surgeon postulated as 2 versus 8 % (gallstones identified only by the surgeon vs. only by the radiologist). We assumed this to be the smallest clinically relevant difference. A sample size of 190 patients being scanned for gallstones was calculated using SamplePower 2.0 and was set to detect this difference with a power of 80 % and at a 5 % significance level (two-tailed). In consultation with the hospital's radiology department, it was estimated that two-thirds of all patients being referred to the radiology department for an abdominal scan would be examined for the occurrence of gallstones. Enrollment was therefore aimed at 300 patients in pursuit of 190 included patients with a RUQ scan.

Statistical analysis

We calculated accuracy, sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV) for surgeon-performed US in detecting gallstones, as well as Cohen's Kappa coefficient, with radiologist-performed US as reference. We used the efficient-score method to calculate the 95 % confidence intervals (CI) of the above, due to Wilson [13, 14].

A p value <0.05 (two-tailed) was considered statistically significant. Analyses were done in IBM SPSS Statistics, versions 20–22.

Results

Patients

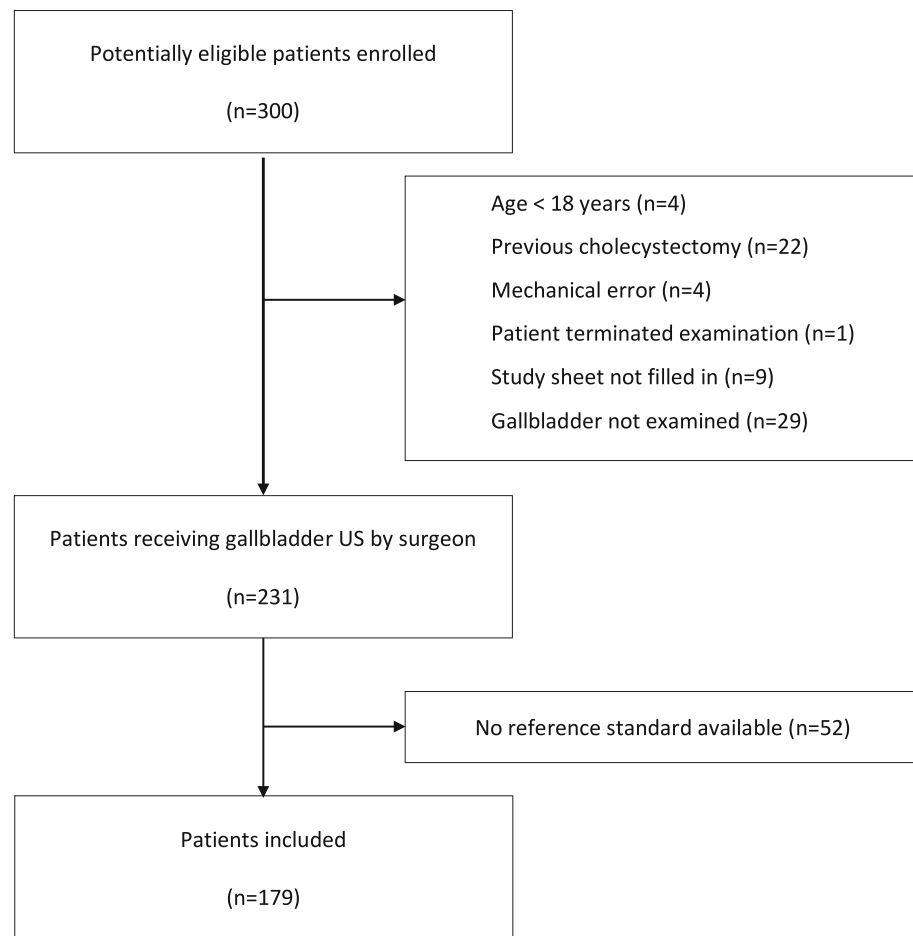
Of the 300 patients enrolled, 179 received a scan of the RUQ, including the gallbladder, from both radiologist and surgeon (Fig. 1). Baseline characteristics of the patients are shown in Table 1.

Surgeon-performed US performance

Seventy-six patients had confirmed gallstones by the radiologist. Surgeon-performed US agreed with radiologist-performed US in 169 of 179 patients, reaching an overall accuracy of 94.4 % (95 % CI 90.0–96.9). The sensitivity was 88.1 % (79.0–93.6 %) and the specificity was 99.0 % (94.7–99.8 %). The agreement of gallstones detected between surgeon and radiologist was high, Cohen's Kappa coefficient = 0.88. There were 67 true-positive and one false-positive diagnoses, resulting in a PPV of 98.5 % (92.1–99.7 %). One hundred and two true-negative and nine false-negative diagnoses provided a NPV of 91.9 % (85.3–95.7 %). There was a systematic difference (p value = 0.021) between false-positive 0.6 % (1/179) versus false-negative 5.0 % (9/179) diagnoses, which indicates that the surgeon more often missed to diagnose gallstones, compared with how often they set a false-positive diagnosis (Fig. 2).

False-positive and false-negative cases

In the only false-positive case, there were no noted positive biliary findings from the radiologist, where the surgeon simply noted gallstones, with no further comment. There was no registered data concerning this patient's weight or BMI. Information about fasting was missing. In the nine cases where the surgeon did not find gallstones (false negatives), the radiologist mentioned that the patient was difficult to examine in two cases. The gallbladder was

Fig. 1 Flow chart. Included patients**Table 1** Patient characteristics (total $n = 179$)

Patient characteristics	N (%)	Mean (range)
Sex		
Male	85 (47.5)	
Female	94 (52.5)	
Age (years)		52.1 (19–91)
Height (m)		1.70 (1.45–2.01)
Weight (kg)		76 (43–125)
BMI (kg/m^2)		26.1 (16.1–39.7)
Admitted	103 (57.5)	
Way of referral ^a		
ED	86 (48.0)	
Surgery Dpt ^b	52 (29.1)	
Other	33 (18.4)	
RUQ pain	80 (44.7)	
RUQ tenderness	56 (31.3)	
Gallstone-specific referral	133 (74.3)	

N total number of patients, BMI body mass index, ED emergency department, RUQ right upper quadrant

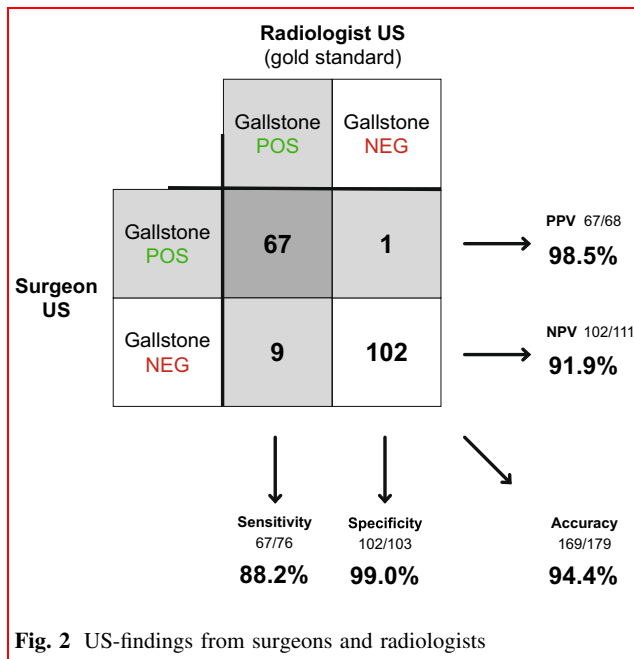
^a Information not available in 8 patients

^b Surgery ward (48) outpatient clinic (4)

either hard to find (“what is considered to be the gallbladder...”) or difficult to evaluate (“the gallbladder is difficult to evaluate, collapsed.”) The radiologist furthermore noted millimeter-sized stones in the gallbladder in three patients, single stones wedged in the neck of the gallbladder in four patients (in one case: “One two millimeter-sized stone is *believed* to be seen in the neck of the gallbladder”), and one case of multiple gallbladder stones. In six of the nine cases, the patient was either not fasting at the time of the scanning, or information about fasting was missing. The false negatives are presented in Table 2 where some of these possible predictive factors are listed. A gallstone wedged in the neck of the gallbladder (missed by the surgeon) is shown in Figs. 3 and 4.

Discussion

This study shows that surgeons can accurately detect gallstones with US and reach a high level of agreement when compared to radiologists.



Our study is, to our knowledge, the largest prospective validation study so far in the area [3, 8, 9], and the setting is clinically relevant. Patients included were all referred to the radiology department for an abdominal scan, but not all presented with RUQ pain (80/179) or were referred with the specific question of gallstones (133/179). The calculated number of patients needed to reach the intended power made the study feasible by including patients in this manner. This also left the examining surgeon with some differential diagnoses in mind, focusing not only on gallstones, at the time of the scanning. We believe that this setting contributes to a less selected patient population and that it might mimic the true clinical situation. For the same reason, a portable US machine, and not a high-end US machine, was used for surgeon-performed US in our study.

We chose not to include any differential diagnoses, or complications to gallstones, in our analyses, since this would have demanded a different study setting and the opinion of the radiologist could not be considered gold standard reference in the same way as for gallstones.

We demonstrate a lower sensitivity for detecting gallstones compared to some previous studies where sensitivities in the range of 95–100 % have been described [3, 8, 9, 15]. These studies had a higher prevalence of gallstones in the study population, which together with clinically suspected biliary disease for the patients included could have led to selection bias, and an overestimation of the sensitivity. Results from larger studies, performed in a more acute setting, are similar to ours, including level of sensitivity. In the study by Alleman et al. [6], including 496 patients who presented with acute abdominal pain at the ED, the surgeons' sensitivity for biliary tract disease (not further specified) ($n = 54$) was shown to be 91 %. When Scruggs et al. [16] studied 575 examinations retrospectively and evaluated the accuracy of ED bedside US (performed by emergency medicine doctors), sensitivity was 88 % and specificity was 87 % in detecting gallstones.

The systematic difference in detecting gallstones between surgeons and radiologists implies that surgeons have more difficulties with excluding the presence of gallstones among patients that actually have the diagnosis, compared with finding gallstones among patients with the diagnosis. Thus, when an US-trained surgeon finds stones, it is most likely that radiologist-performed US would confirm this and we can trust the surgeon's positive examination to a high degree. On the other hand, we cannot use the negative examination to exclude gallstones.

The high PPV (99 %) in our study further supports this. It indicates that patients with typical signs of symptomatic gallstones, and a positive surgeon-performed US scan, could be considered for surgery, and do not need further

Table 2 False-negative results from surgeon-performed ultrasound (total $n = 9$)

Patient number	Weight (kg)	Age (years)	Fasting	BMI (kg/m ²)	Sex	Difficulties/uncertainty according to radiologist	Stone in neck of gallbladder	Small stones (1–3 mm)	Gallbladder completely filled with stones	BMI >30	Not fasting (<4 h)
29	70	38	(<4 h)	22.6	M	–	–	–	+	–	+
62	90	88	>12 h	29.7	M	+	–	–	–	–	–
114	MD	30	MD	MD	F	–	–	+	–	MD	MD
116	51	73	MD	16.1	M	–	+	–	–	–	MD
182	78	52	MD	23.3	M	+	–	–	+	–	MD
192	105	52	(<4 h)	30.7	M	–	+	–	–	+	+
219	105	68	4–6 h	31.7	M	–	–	+	–	+	–
287	102	59	4–6 h	32.6	M	–	+	–	–	+	–
288	60	46	(<4 h)	21.3	F	+	+	+	–	–	+

MD missing data, BMI body mass index, F female, M male

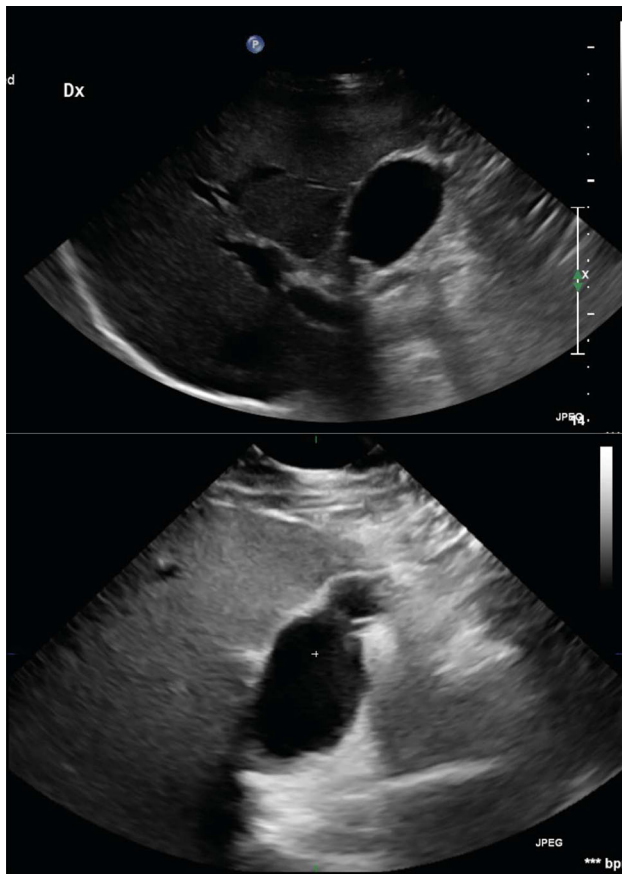


Fig. 3 Patient 287. Ultrasound performed by radiologist. Centimeter-sized stone in the neck of the gallbladder

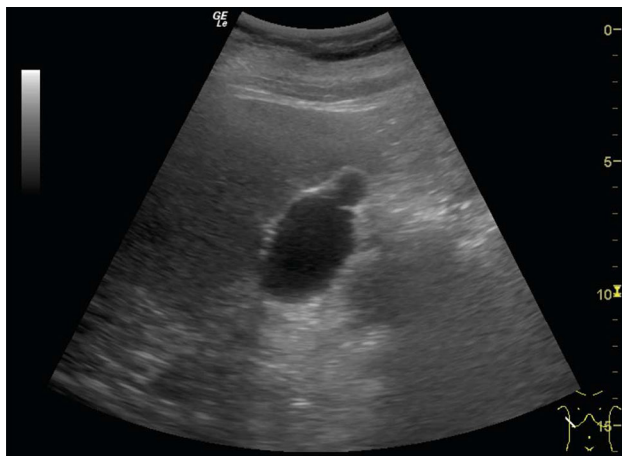


Fig. 4 Patient 287. Ultrasound performed by surgeon. Missing the stone

examination by a radiologist. In case of typical symptoms but a negative surgeon-performed US scan, further investigation at the radiology department (with US or MRI) should be advised. In our study, the NPV was 92 %, but in

a group of patients with a higher prevalence of gallstones, the NPV might have been lower.

Since patient enrollment required surgeon availability at presentation enrolled patients were not consecutive, hence there is a risk of selection bias. However, in all patients where the surgeon did not perform a gallbladder scan ($n = 29$), the referral concerned other abdominal organs. It is possible that other factors could have contributed to a RUQ scan not being performed in these cases, such as the stress level of the patient or perceived examining difficulties by the surgeon. One could argue that the surgeon should not have been aware about the reason for performing an US for each patient, to avoid selection bias, although in our study the surgeon and the radiologist both had information about the patient's condition and the reason for referral. There was also a possibility of patients overhearing findings and revealing the result of the previous examination, thus influencing the latter examiner's investigation (observer bias).

Using multiple radiologists and thus multiple individuals with various experiences as a reference standard might have had an influence on our results, as compared to using one US specialist as an expert examiner. However, using several radiologists might reflect a more actual clinical practice where the US examination would be performed by the available radiologist on duty.

The growing use of surgeon-performed ultrasound has increased the need of a standardized US training. Current recommendations on US training for surgeons are based on expert society recommendations rather than study evidence, hence the need of validation studies. The role of surgeon-performed US should not be to replace formal radiological assessment but to complement physical examination [5, 17].

US training as well as investment in equipment is associated with costs, hence the importance of defining the amount of initial and continuous training needed in order to reach and maintain an adequate level of US competence. Further studies aiming to validate how to maintain US skills would add valuable information to this question. The presence of a learning curve for novices performing US of the RUQ has previously been studied in emergency physicians [18], where the authors found that full agreement with the expert examiner was generally reached after performing 25 scans, suggesting that this amount might suffice as practice in a US training program to perform accurate RUQ scans.

Conclusion

Our results support that adequately trained surgeons can accurately detect gallstones using US and reach a high level of agreement with radiologists.

We therefore recommend that patients with a clinical history of suspected gallstones and a positive scan performed by the US-trained surgeon could be considered for surgery without further radiology. A negative surgeon-performed scan on a patient with typical history of gallstones should, however, be referred to the radiology department for further examination.

Acknowledgments The authors would like to thank Marie Beer-mann and Odd Runeborg in the Department of Radiology, Stockholm South General Hospital, for their expertise and highly skilled training in ultrasound. We would also like to thank Anna-Klara Nordblad-Sasnauskas for the help with interpreting radiologist statements. We send our gratitude to all participating surgeons in the Department of Surgery, Stockholm South General Hospital: Åsa Hallqvist, Karin Lind, Linda Nigard, Martin Nordberg, and Jenny Oddsberg. We thank Martin Dahlberg, Department of Surgery, Stockholm South General Hospital, for useful help and invaluable support in manuscript writing.

Funding This study was supported by a grant from the Swedish Society of Medicine, SEK180.000 (=USD 20.000).

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

1. Powers RD, Guertler AT (1995) Abdominal pain in the ED: stability and change over 20 years. *Am J Emerg Med* 13:301–303
2. Cooperberg PL, Burhenne HJ (1980) Real-time ultrasonography. Diagnostic technique of choice in calculous gallbladder disease. *N Engl J Med* 302:1277–1279
3. Kell MR, Aherne NJ, Coffey C et al (2002) Emergency surgeon-performed hepatobiliary ultrasonography. *Br J Surg* 89:1402–1404
4. Rozycki GS (1998) Surgeon-performed ultrasound: its use in clinical practice. *Ann Surg* 228:16–28
5. Moore CL, Copel JA (2011) Point-of-care ultrasonography. *N Engl J Med* 364:749–757
6. Allemann F, Cassina P, Rothlin M et al (1999) Ultrasound scans done by surgeons for patients with acute abdominal pain: a prospective study. *Eur J Surg* 165:966–970
7. Lindelius A, Torngren S, Pettersson H et al (2009) Role of surgeon-performed ultrasound on further management of patients with acute abdominal pain: a randomised controlled clinical trial. *Emerg Med J* 26:561–566
8. Fang R, Pilcher JA, Putnam AT et al (1999) Accuracy of surgeon-performed gallbladder ultrasound. *Am J Surg* 178:475–479
9. Ahmad S, Zafar A, Ahmad M et al (2005) Accuracy of surgeon-performed abdominal ultrasound for gallstones. *J Ayub Med Coll Abbottabad* 17:70–71
10. Shea JA, Berlin JA, Escarce JJ et al (1994) Revised estimates of diagnostic test sensitivity and specificity in suspected biliary tract disease. *Arch Intern Med* 154:2573–2581
11. Carroll PJ, Gibson D, El-Faedy O et al (2013) Surgeon-performed ultrasound at the bedside for the detection of appendicitis and gallstones: systematic review and meta-analysis. *Am J Surg* 205:102–108
12. Lindelius A, Torngren S, Sonden A et al (2008) Impact of surgeon-performed ultrasound on diagnosis of abdominal pain. *Emerg Med J* 25:486–491
13. Newcombe RG (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 17:857–872
14. Dean AG, Sullivan KM, Soe MM. Open source epidemiologic statistics for public health, Version 3.03a
15. Irkorucu O, Reyhan E, Erdem H et al (2013) Accuracy of surgeon-performed gallbladder ultrasound in identification of acute cholecystitis. *J Invest Surg* 26:85–88
16. Scruggs W, Fox JC, Potts B et al (2008) Accuracy of ED bedside ultrasound for identification of gallstones: retrospective analysis of 575 studies. *West J Emerg Med* 9:1–5
17. Shepherd AE, Gogalniceanu P, Kashef E et al (2012) Surgeon-performed ultrasound—a call for consensus and standardization. *J Surg Educ* 69:132–133
18. Gaspari RJ, Dickman E, Blehar D (2009) Learning curve of bedside ultrasound of the gallbladder. *J Emerg Med* 37:51–56