The American Journal of Surgery (2013) 205, 102–108

The American Journal of Surgery®

Review Article

Surgeon-performed ultrasound at the bedside for the detection of appendicitis and gallstones: systematic review and meta-analysis

Patrick J. Carroll, B.B.L.S., David Gibson, M.B.B.Ch., Osama El-Faedy, M.R.C.S., Colum Dunne, Ph.D., Calvin Coffey, Ph.D., F.R.C.S., Ailish Hannigan, B.Sc., Ph.D., Stewart R. Walsh, M.Sc., M.Ch., F.R.C.S.*

Graduate Entry Medical School, University of Limerick, Limerick, Ireland

KEYWORDS:
Ultrasound;
Surgeon performed;
Appendicitis;
Gallstones;
Systematic review;
Meta-analysis

Abstract

BACKGROUND: We undertook a systematic review and meta-analysis to compare surgeon-performed ultrasound (SPUS) for suspected appendicitis or gallstone disease to the "gold standard" of pathological examination or radiologist-performed ultrasound (RPUS).

DATA SOURCES: MEDLINE, Embase, trial registries, conference proceedings, and article reference lists were searched to identify trials and/or studies comparing SPUS with pathology or RPUS as the reference standard. Data were abstracted from eligible studies to produce 2 × 2 contingency tables, permitting the calculation of pooled sensitivity and specificity values.

RESULTS: Eight studies (1,268 patients) evaluated SPUS for appendicitis. For appendicitis, SPUS had a pooled sensitivity of .92 (95% confidence interval [CI], .887–.939) and a pooled specificity of .96 (95% CI, .946–.974). SPUS for gallstones was evaluated in 8 studies (1,019 patients). The pooled sensitivity was .96 (95% CI, .934–.979), and the specificity was .99 (95% CI .983–.998).

CONCLUSIONS: SPUS achieves acceptable sensitivity and specificity for both gallstones and appendicitis. However, there was some evidence of heterogeneity. Data regarding cost-effectiveness are lacking.

Patients with acute abdominal pain account for up to 10% of emergency department (ED) visits.[1–7] Costs arising from subsequent admission are a significant health economic burden, with the most costly factor being hospital stay.[7,8] The UK National Health Service spends about £100 million per annum managing patients with acute abdominal pain.[7]

Surgeon-performed ultrasound (SPUS) may limit costs by reducing time spent waiting for diagnostic tests.[9] However, to be acceptable, SPUS must compare favorably with radiologist-performed ultrasound (RPUS). SPUS has been the subject of numerous studies. To date, there has been no pooled analysis. We undertook a systematic review and meta-analysis to evaluate the efficacy of SPUS for appendicitis and gallstone disease.

Methods

The systematic review was conducted in accordance with the Preferred Reporting Items in Systematic Reviews and Meta-analysis (PRISMA) guidelines.[10] The MEDLINE da-

* Corresponding author. Tel: +353-86-3625492; fax: +353-61-234850

E-mail address: stewart.walsh@ul.ie

Manuscript received October 6, 2011; revised manuscript February 26, 2012

0002-9610/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2012.02.017

tabase was searched to identify studies reporting results of SPUS. Combinations of the search terms "surgeon," "bedside," and "ultrasound" were used. Two authors (PJC and SW) independently performed electronic searches in February 2011 and June 2011. Conference proceedings from the Association of Surgeons of Great Britain and Ireland and the American College of Surgeons (2001–2011) were searched manually. Article reference lists were scrutinized for further relevant publications. A kappa statistic was calculated to measure the level of agreement between the 2 reviewers. Differences were resolved by consensus.

For appendicitis, SPUS was compared with the pathological examination as the "gold standard" investigation. With respect to gallstone disease, SPUS was compared with the final discharge diagnosis (whether based on RPUS or pathological examination). Studies were eligible for inclusion provided that they were prospective, evaluated SPUS for suspected appendicitis or gallstones, were performed in adults, and provided sufficient data to allow construction of 2×2 contingency tables. Studies undertaken in pediatric populations and those evaluating emergency department–performed ultrasounds were excluded.

Random-effects modeling was used to calculate pooled sensitivities and specificities using the contingency tables. Summary receiver operating characteristic (ROC) curves were plotted. These curves have been recommended as a way of representing the performance of a diagnostic test using data from a meta-analysis. The area under the curve, a global reflection of diagnostic accuracy with a theoretic range of 0 to 1, was calculated together with its standard error. The area under the curve has been shown to be a useful summary of the ROC curve and reasonably robust to heterogeneity.[11] Heterogeneity was tested by means of the I^2 test. This initial statistical analysis was performed using Meta-DiSc version 1.4 (Universidad Complutense Madrid and Hospital Ramon y Cajal, Madrid, Spain).[12]

The quality of each eligible study was evaluated using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. The QUADAS tool is structured around 14 questions (Table 1) that have proven their validity in the evaluation of diagnostic studies.[13] The questions include whether the spectrum of patients in the study was representative of all patients who will be given the test, whether the selection criteria were clearly described, and detailed questions regarding the reference standard used in the study. The maximum score that can be obtained on QUADAS is 14, indicating the study met of all the necessary criteria.

Results

The results of the systematic review are summarized in the flow chart (Fig. 1). The review identified 12 relevant studies (2,287 patients). Most of the 4,207 identified citations were excluded by title alone because they referred to pathology other than gallstones or appendicitis. Of the 166

Table 1 QUADAS score components of individual studies included in the meta-analysis

		Spectrum composition	Selection criteria	Reference standard	Disease progression Bias	Partial verification Bias	Differential verification	Incorporation Bias	Index test execution	Reference standard execution	Test review Bias	Reference standard review Bias	Clinical review Bias	Uninterrpretable test result	Withdrawals	Total
Alleman et al,[14]	1999	Y	Y	Y	Y	N	N	Y	Y	Y	Y	U	Y	Y	Y	11/14
Chen et al,[18]	1998	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	12/14
Burford et al,[9]	2011	Y	Y	Y	U	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	12/14
Amgwerd et al,[16]	1994	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	U	Y	Y	Y	12/14
Davies et al,[19]	1991	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	13/14
Kang et al,[20]	1989	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	13/14
Chen et al,[17]	2000	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	11/14
Fang et al,[24]	1999	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	14/14
Ahmad et al,[22]	2005	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	13/14
Kell et al,[23]	2002	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	13/14
Williams et al,[15]	1994	Y	Y	Y	Y	N	U	Y	Y	Y	Y	U	Y	Y	Y	11/14
Eiberg et al,[21]	2008	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	14/14

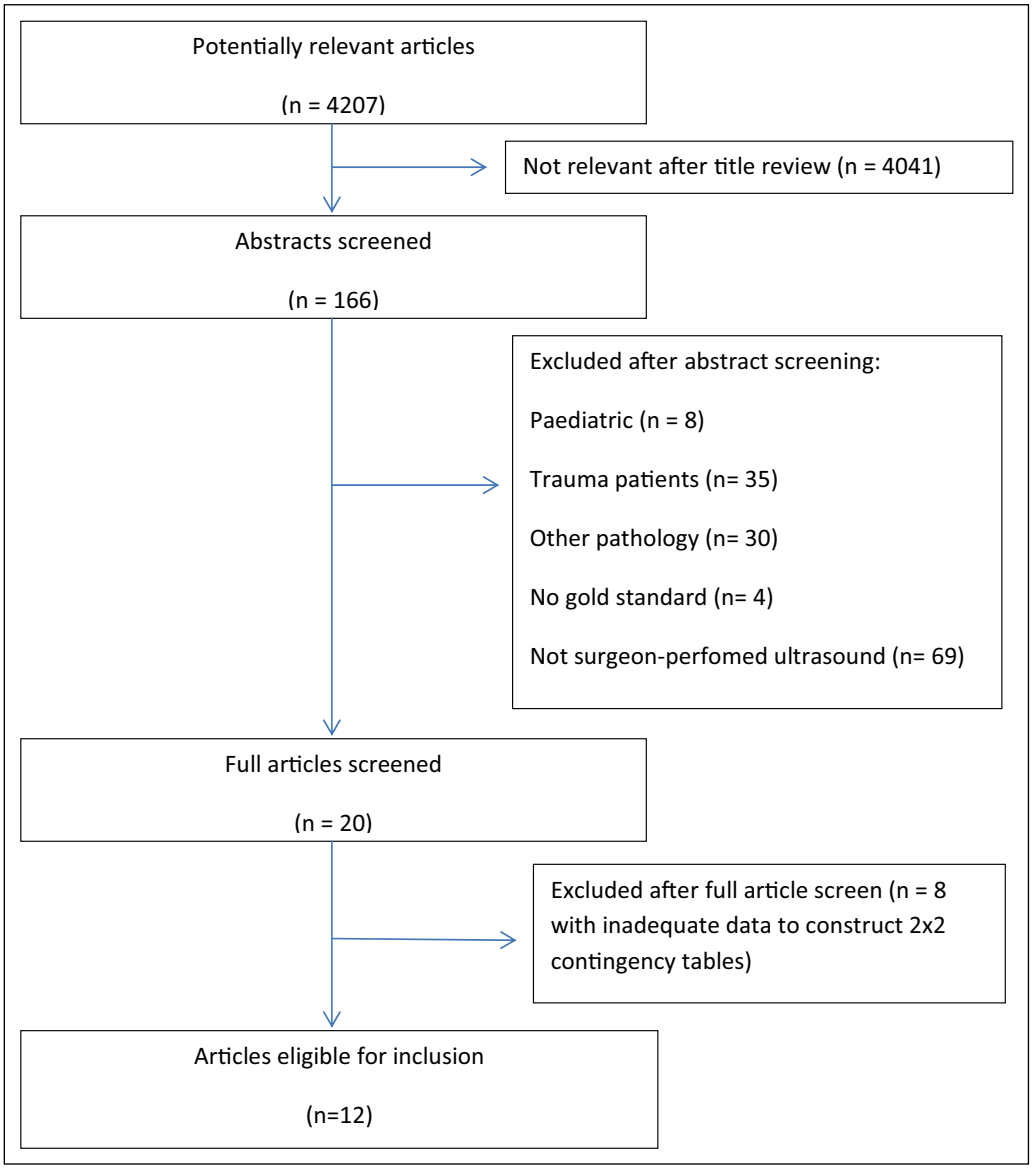

Figure 1 The PRISMA flowchart.

screened abstracts, the most common reason for exclusion was that the study did not evaluate SPUS (n = 69). Studies reporting evaluations in trauma, other pathology, or pediatric patients accounted for most of the remaining exclusions. A kappa statistic of .90 indicated a high level of agreement between the 2 independent reviewers on the studies to be included. Overall, the QUADAS scores ranged from 11 to 14, with a median score of 12.5 (Table 1); therefore, individually the quality of the studies included was high.

Appendicitis

Eight studies[9,14–20] (1,268 patients) provided sufficient data to evaluate SPUS for suspected appendicitis (Table 2). There was evidence of significant heterogeneity with respect to both sensitivity and specificity (I^2 = .84 and .89, respectively; $P < .001$). A random-effects model was used to estimate pooled sensitivity and specificity. The pooled sensitivity was .92 (95% confidence interval [CI], .887–.939), and the pooled specificity was .96 (95% CI, .946–.974). The summary ROC curve (Fig. 2) had an area under the curve of .9839 (standard error = .0067).

Gallstones

Eight studies[14,15,17,19,21–24] (1,019 patients) provided sufficient data for inclusion in this analysis (Table 3). From the meta-analysis, the pooled sensitivity was .96 (95% CI, .934–.979). The pooled specificity was .99 (95% CI, .983–.998) (Table 3). The summary ROC curve (Fig. 3) had an area under the curve of .9888 (standard error = .0055). There was no evidence of heterogeneity for pooled sensitivity (I^2 = .46, P = .073). However, there was evidence of heterogeneity for pooled specificity (I^2 = .74, $P < .001$).

Table 2 Meta-analysis for appendicitis

Study name	Alleman et al[14]	Chen et al[17]	Burford et al[9]	Amgwerd et al[16]	Davies et al[19]	Kang et al[20]	Chen et al[18]	Williams et al[15]
Sensitivity	.94	.92	.79	.92	.91	.86	.99	.55
Specificity	1.00	.99	.96	.95	.94	1.00	.68	.50
True-positives	89	11	23	109	21	36	143	11
False-positives	2	1	1	10	1	0	15	0
True-negatives	399	89	24	173	17	20	32	0
False-negatives	6	1	6	10	2	6	1	9

Comments

SPUS confers several potential benefits in patients with suspected appendicitis or gallstones. These may include a reduced hospital stay by avoiding waiting periods for radiologic investigation, reduced bed use costs, and possibly increased patient satisfaction. Conversely, there are risks of missed pathology. To be acceptable, SPUS must not be inferior to the current standard of care. Although numerous studies have reported the results of SPUS, there has been no attempt at a systematic evaluation of the technique's efficacy.

The point estimates for pooled sensitivity and specificity from this meta-analysis suggest that SPUS achieves acceptable sensitivities and specificities for the diagnosis of gallstones and appendicitis. These results must be interpreted with some caution. There is significant heterogeneity among the studies for both appendicitis and gallstones. This partly arises from inherent inter- and intraobserver variation in studies of diagnostic imaging techniques. Ultrasound in particular is very operator dependent, such that significant variation between individual operator sensitivities is likely to result in statistical heterogeneity. It may also arise from variation in instrument quality. Some studies included in the meta-analyses used portable, handheld ultrasound scanners, while others used full-scale machines (Table 4). The resolution of portable bedside ultrasound machines does not match that of high-resolution units available in radiology departments. Thus, the sensitivity and specificity of SPUS would be expected to vary depending on which type of machine was used.

There are further limitations. Although the overall study quality was good, there is some potential for bias. Surgical findings were used as the "gold standard" comparison in some studies. In 2[17,18] studies, the operating surgeon operating was not blinded to the results of the SPUS, and this information was unclear in 3[14–16] studies, introducing potential observer bias. It is important to note that the current pooled analyses are restricted solely to surgeons' ability to

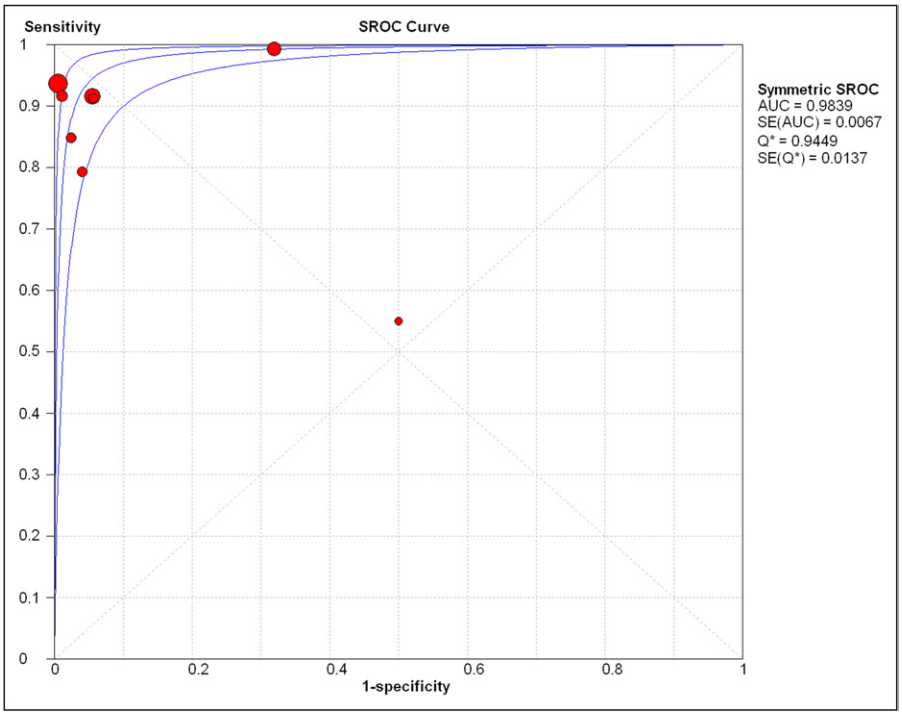

Figure 2 The summary ROC curve for appendicitis.

Table 3 Meta-analysis for gallstones

Study name	Alleman et al[14]	Fang et al[24]	Chen et al[17]	Ahmad et al[22]	Kell et al[23]	Williams et al[15]	Eiberg et al[21]	Davies et al[19]
Sensitivity	.91	.99	1	.99	.95	.89	1	1
Specificity	1	.82	1	1	1	0	.96	1
True-positives	49	74	3	100	40	33	13	5
False-positives	1	2	0	0	0	1	1	0
True-negatives	441	9	99	41	11	0	22	61
False-negatives	5	1	0	1	2	4	0	0

identify gallstones or appendicitis using ultrasound. The results cannot be generalized to all patients presenting with acute abdominal pain. To do so risks missing significant pathology. Missed pathology is always a significant concern when nonradiologists undertake radiologic examinations. Accurate evaluation of missed pathology rates requires prolonged follow-up, which none of the series included here undertook. Thus, the missed pathology rate for SPUS remains unknown. Patients with acute abdominal pain in whom an SPUS rules out appendicitis or gallstones require further evaluation.

The prospective series included in this systematic review could not provide hospital stay or cost-effectiveness data because they did not include comparisons with groups of patients receiving standard care. A large single-center trial comparing SPUS with routine care, which randomized 800 patients, has recently been reported.[25] The length of hospital stay and time spent in the emergency department were evaluated. The mean length of hospital stay was 4.3 days with SPUS versus 5.4 days in controls ($P = .964$). The overall delay in the emergency department was 4.4 hours with SPUS versus 4.6 hours without ($P = .545$). Although it is difficult to draw robust conclusions from 1, single-center trial, SPUS does not appear to confer a large benefit in terms of the length of stay. However, the authors did report that the SPUS group was more likely to proceed straight to theater, with no additional delay once admitted.

Although pathology and RPUS remain the "gold standard," the costs associated with hospital stay and the inefficient use of human resources remains of utmost importance and concern to hospitals. However, to replace the current "gold standard" for appendicitis and gallstone detection, any alternative imaging modality must accurately identify appendicitis and gallstones with high sensitivity and specificity. Our meta-analysis suggests that SPUS may have a role within care pathways for patients presenting with acute abdominal pain. A number of questions remain. The efficacy of SPUS compared with emergency depart-

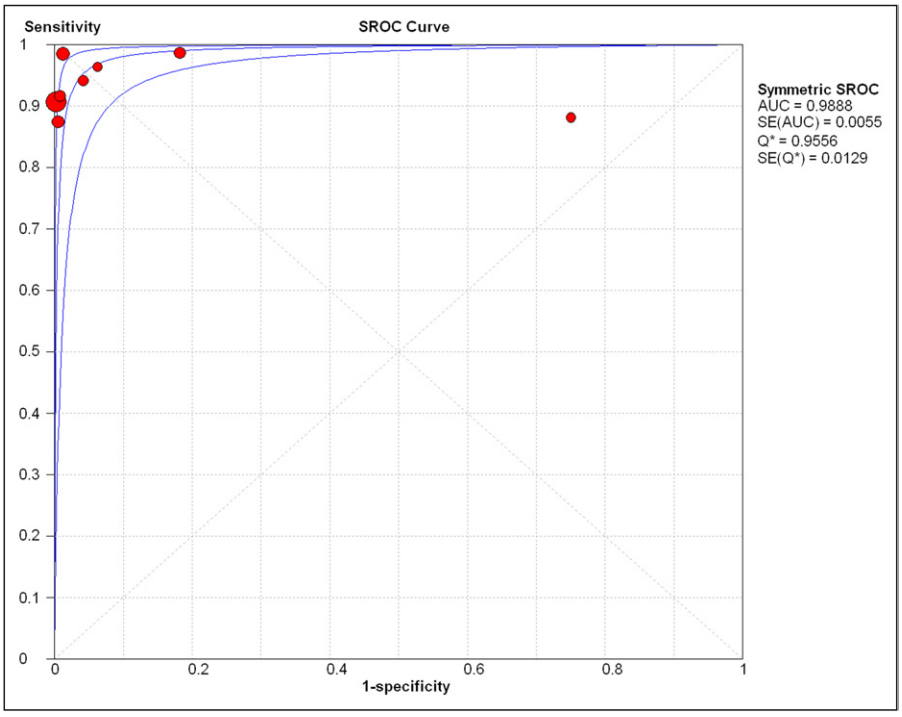

Figure 3 The summary ROC curve for gallstones.

Table 4 Characteristics of the studies used in systematic review and meta-analysis

Study name	Alleman et al[14]	Chen et al[18]	Burford et al[9]	Amgwerd et al[16]	Davies et al[19]	Kang et al[20]	Chen et al[17]	Fang et al[24]	Ahmad et al[22]	Kell et al[23]	Williams et al[15]	Eiberg et al[21]
Year	1999	1998	2011	1994	1991	1989	2000	1999	2005	2002	1994	2008
Total patients	496	158	54	203	152	62	102	77	142	53	139	30
Gold standard	Pathology as last resort	Pathology	Final discharge	Pathology	Final discharge	Pathology	Pathology	Pathology	RPUS	Final diagnosis	Final diagnosis	RPUS
Portable machine	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No

ment physician–performed ultrasounds has not been evaluated. Missed pathology rates and readmission rates are unclear. Data regarding cost-effectiveness, patient satisfaction, and hospital stay are lacking. A focused prospective analysis, by way of a randomized controlled trial, comparing SPUS with standard care is required to further examine the role and usefulness of SPUS.

Conclusions

Although RPUS and pathology are the "gold standard" in diagnosing appendicitis and gallstones, evidence suggests that SPUS has a role in its diagnosis. This study confirms that SPUS offers promise as a sensitive and specific modality for the detection of appendicitis and gallstones, potentially obviating the need for RPUS or pathological diagnosis. However, a cautious approach is required because of the high degree of heterogeneity for appendicitis. Because of the available series of data and acceptable sensitivity and specificity, there is the potential to streamline care by minimizing diagnostic delay via SPUS. Further research by way of the suggested randomized control trial is required for SPUS to be used as the sole imaging modality for appendicitis and gallstones.

References

1. Brewer BJ, Golden GT, Hitch DC, et al. Abdominal pain. An analysis of 1,000 consecutive cases in a University Hospital emergency room. Am J Surg 1976;131:219–23.
2. Powers RD, Guertler AT. Abdominal pain in the ED: stability and change over 20 years. Am J Emerg Med 1995;13:301–3.
3. Sanson TG, O'Keefe KP. Evaluation of abdominal pain in the elderly. Emerg Med Clin North Am 1996;14:615–27.
4. Kamin RA, Nowicki TA, Courtney DS, et al. Pearls and pitfalls in the emergency department evaluation of abdominal pain. Emerg Med Clin North Am 2003;21:61–72.
5. Hawthorn IE. Abdominal pain as a cause of acute admission to hospital. J R Coll Surg Edinb 1992;37:389–93.
6. Irvin TT. Abdominal pain: a surgical audit of 1190 emergency admissions. Br J Surg 1989;76:1121–5.
7. Sheridan WG, White AT, Havard T, et al. Non-specific abdominal pain: the resource implications. Ann R Coll Surg Engl 1992;74:181–5.
8. Morino M, Pellegrino L, Castagna E, et al. Acute nonspecific abdominal pain: A randomized, controlled trial comparing early laparoscopy versus clinical observation. Ann Surg 2006;244:881–6.
9. Burford JM, Dassinger MS, Smith SD. Surgeon-performed ultrasound as a diagnostic tool in appendicitis. J Pediatr Surg 2011;46:1115–20.
10. Liberati A, Altman DG, Tetzlaff J, et al. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009;151:W65–94.
11. Walter SD. Properties of the summary receiver operating characteristic (SROC) for diagnostic test data. Stat Med 2002;21:1237–56.
12. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006; 6:31.

13. Whiting PF, Weswood ME, Rutjes AWS, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006;6:9.
14. Allemann F, Cassina P, Röthlin M, et al. Ultrasound scans done by surgeons for patients with acute abdominal pain: a prospective study. Eur J Surg 1999;165:966–70.
15. Williams RJ, Windsor AC, Rosin RD, et al. Ultrasound scanning of the acture abdomen by surgeons in training. Ann R Coll Surg Engl 1994; 76:228–33.
16. Amgwerd M, Rothlin M, Candinas D, et al. Ultrasound diagnosis of appendicitis bysurgeons-a matter of experience? A prospective study. Langenbecks Arch Surg 1994;379:335–40.
17. Chen SC, Lin FY, Hsieh YS, et al. Accuracy of ultrasonography in the diagnosis of peritonitis compared with the clinical impression of the surgeon. Arch Surg 2000;135:170–3.
18. Chen SC, Chen KM, Wang SM, et al. Abdominal sonography screening of clinically diagnosed or suspected appendicitis before surgery. World J Surg 1998;22:449–52.
19. Davies AH, Mastorakou I, Cobb R, et al. Ultrasonography in the acute abdomen. Br J Surg 1991;78:1178–80.
20. Kang WM, Lee CH, Chou YH, et al. A clinical evaluation of ultrasonography in the diagnosis of acute appendicitis. Surgery 1989;105:154–9.
21. Eiberg JP, Grantcharov TP, Eriksen JR, et al. Ultrasound of the acute abdomen performed by surgeons in training. Minerva Chir 2008;63:17–22.
22. Ahmad S, Zafar A, Ahmad M, et al. Accuracy of surgeon-performed ultrasound for gallstones. J Ayub Med Coll Abbottabad 2005;17:70–1.
23. Kell MR, Aherne NJ, Coffey C, et al. Emergency surgeon-performed hepatobiliary ultrasonography. Curr Probl Surg 2002;38:141–212.
24. Fang R, Pilcher JA, Putnam AT, et al. Accuracy of surgeon-performed gallbladder ultrasound. Am J Surg 1999;178:475–9.
25. Lindelius A, Törngren S, Pettersson H, et al. Role of surgeon-performed ultrasound on further management of patients with acute abdominal pain: a randomised controlled clinical trial. Emerg Med J 2009;26:561–6.