Limited lung ultrasound protocol in elderly patients with breathlessness; agreement between bedside interpretation and stored images as acquired by experienced and inexperienced sonologists

Kylie Baker¹

MBBS, FACEM
Senior Lecturer

Geoffrey Mitchell²

MBBS, FRACGP, PhD

Geoffrey Stieler³

MBBS(Hons)
Master Applied Science
(Medical Ultrasound)

¹Ipswich General
Hospital Emergency
Department
Ipswich
Queensland
Australia

²University of Queensland
Ipswich Campus
Building 12
Ipswich
Queensland
Australia

³Department of Medical
Imaging
Royal Brisbane and
Women's Hospital
Herston
Queensland
Australia

Correspondence to email
kyliebaker888@bigpond.
com

Abstract

Introduction: Bedside lung ultrasound is increasingly performed in the Emergency Department to assess dyspnoeic patients. Quantifying the number of B-lines (a vertical short path reverberation artefact caused by the presence of interstitial fluid) can help clinicians differentiate 'wet lung' – pulmonary oedema, from 'dry lung' – not pulmonary oedema.

The aim of this study was to determine inter-rater agreement for this simple bedside investigation, comparing relative ultrasound novices with more experienced personnel. After completing an introductory ultrasound course the novices had only had four hours of specific lung ultrasound education, followed by 10 proctored scans.

Method: A prospective convenience sample of 217 patients over the age of 60 and presenting with dyspnoea were recruited. Patients were either scanned by an experienced emergency physician sonologist or by relative ultrasound novices. The scans were saved and still images then interpreted by a blinded radiology registrar with sonography training, and agreement calculated using weighted kappa scores.

Results: The experienced physician scanned 155 patients, and the 11 novices 62 between them. Agreement between the experienced sonologist and the blinded expert was excellent (kappa = 0.82, 95% CI 0.72 to 0.92); between novice group and sonographer, agreement was good (kappa = 0.70, 95% CI 0.45 to 0.95).

Conclusion: This study demonstrates reasonable inter-rater agreement firstly between experts, and secondly between novices and experts.

Keywords: accreditation, dyspnoea, lung, ultrasonography.

Introduction

One simple use for lung ultrasound scanning promoted in the European literature is as a screening tool for diagnosing pulmonary oedema.¹⁻⁴ The recommended protocol is not precise, but reportedly better than current clinical diagnostic accuracy, which includes chest xray.^{5,6} In our non tertiary emergency department, overnight radiography call in is increasingly expensive, so we investigated whether lung sonography could be used to select patients who could wait safely for morning radiography. In developing this project, we wanted to check how best to teach the protocol, what errors to expect from inexperienced staff during the learning curve, and whether we could set up a healthy self-audit habit.

Lung sonography is practised in Australia, but a literature search revealed no research that was applicable to credentialling our practitioners. However onerous, credentialing is a medico-legal imperative.⁷⁻¹¹ Well-designed accreditation standards balance risk with need, identifying an acceptable error rate without being so rigorous as to exclude modes of practice that would benefit the patient population.¹¹ Physicians can choose to inform credentialing boards proactively, or to risk wholesale adoption of external or non-contextual guidelines.

Research to date has focussed on the diagnostic accuracy of lung scanning, rather than its implementation. Point of care (POC) scanning standards demonstrate great contrast between

the generic UK, USA and Australian recommendations^{7-9,12} and specific recommendations in the lung ultrasound literature.¹³⁻¹⁶ Time to acceptable competence has been reported as ranging from thirty minutes¹³ to seven months.¹⁶ Recommended experience ranges from five proctored scans¹⁵ to thirty.¹²

Closer examination of these articles show that they refer to differing practitioners, protocols and procedures. Bedetti, *et al.*¹³ needed only 30 minutes to train echocardiographers to demonstrate and interpret wet lungs, while Brooke, *et al.*¹⁴ took 2 days to train ultrasound naïve paramedics to recognise pneumothorax. Liteplo, *et al.*¹⁵ trained medical students to acquire lung scans in 2.5 hours while Tutino, *et al.*¹⁶ took seven months before they graded intensivists as experts in reporting lung ultrasound pathology. Finally the British Thoracic Society requires 20 witnessed procedures, and 30 proctored scans before it allows an unsupervised doctor to drain a pleural effusion.¹² None of these articles directly apply to the teaching of emergency physicians to use an 8-view protocol to detect pulmonary oedema.

Establishing an objective and reproducible technique is a necessary precursor to testing its diagnostic accuracy in the populations of interest. Both practitioner variance and protocol variance should be quantified before the technique can be trialled overnight in our emergency department.

This study tests reproducibility of the most basic lung scanning protocol (LUS)¹⁷ in the hands of an experienced sonologist. Using a rapid and non-invasive technique, a patient is scanned and a diagnosis made at the bedside. The technique is taught to less experienced practitioners, and the interpretation of scans compared with the interpretation of a blinded expert reviewer. By analysing the discordances between expert and experienced or inexperienced sonologists we can suggest ways to improve the teaching of the protocol.

Method

This is a prospective observational study of a convenience sample of patients aged 60 years and over, presenting to the Ipswich Emergency Department (ED) complaining of any breathlessness. Recruiting occurred from 0800 to 2330 hours when a LUS trained medical officer (MO) was available. The project commenced March 2011 and concluded February 2012. Patients were excluded if scanning interfered with active resuscitation, or if a chest x-ray was not performed within one hour of lung scan, or within two hours providing that no active fluid management occurred in the intervening period.

All scans were collected on a GE Logic-*e* portable ultrasound (GE, Schenectady, New York, U.A), using a 2-5MHz curved probe on the 'FAST' preset, with the focus as close to the pleural line as possible. Harmonics and cross-beam were switched off. The protocol was adapted from recent studies, with minor adaptations to suit the resources.^{2,4,15,18}

Ultrasound scan sets were reported as 'wet' if there were three or more simultaneous B lines present in two or more views on each side of the chest. B lines indicate thickening or fluid in the alveolus or interstitium. This most basic form of the protocol assumes that widespread, bilateral B lines indicate pulmonary oedema. Diffuse bilateral inflammatory conditions are considered acceptable false positives. The sonologist could comment on the incidentals field if they thought the protocol was falsely positive or negative, and if

the scan would change current management direction.

All Medical Officers (MOs) performed and saved at least eight longitudinal views per patient. The MO interpreted the protocol as 'wet', 'dry' or 'indeterminate', and documented incidental findings. An MO not directly involved in managing the patient performed lung scans, and the findings were recorded separately to the clinical record. The scanning MO did not inform the treating MO of findings unless a significant incidental finding required emergent management. This was an ethical requirement.

All saved scans were also interpreted as 'wet', 'dry' or 'indeterminate' by a blinded doctor-sonographer. To maintain clinical relevance, we tried to save stills rather than cine loops, as the hospital imaging system does not store ultrasound loops. A set of eight still images required 1.3 megabytes of data, while a single 3 second loop averaged 12 megabytes.

To be allowed to perform LUS independently, the medical officer had to:

- 1 Attend a short course on FAST ultrasound, minimum of one day in preceding two years.
- 2 Attend a pre-set program consisting of a lecture, demonstration scan, and case study (1.5 hours), followed by interpretation quiz (1 hour) and practice on models (1.5 hours).
- 3 Perform ten proctored scans (assistance with labelling and saving of the scans, not interpretation or acquisition). At least four proctored scans were performed on healthy volunteers in the workshop; subsequent scans were performed on patients. The purpose of proctoring was not to alter acquisition or interpretation, but to observe the potential sources of error in the beginner, and ensure the images were saved. Hence when an incorrect interpretation was made, or poor images were saved, the proctor did not correct this. Proctored scans are included to illustrate the nadir of the learning curve, and then the unproctored scan agreement is calculated separately.

The experienced sonologist had three certificates of clinician performed ultrasound, ten years of point of care scanning, and one year trialling lung ultrasound protocols and attending specific workshops. The blinded reporting was by a registrar training in radiology, with a Master of Applied Science (Medical Ultrasound) and echocardiography credentialing, with 14 years as a professional sonographer.

To allow comparison with prior studies, we extracted age, sex, length of stay, triage category, delayed expert radiology report and the ED diagnosis code for the patient presentation from the hospital information systems.

Ethical considerations

Waiver of consent was obtained from the Queensland Civil and Administrative Tribunal before commencement of study, and ethics approval was granted by the West Moreton Health Service District Human Research Ethics Committee.

Statistics

This study is a substudy of the diagnostic accuracy of LUS against a radiologist reported chest x-ray and specialist cardiologist diagnosis after chart review. The proposed sample size was 200. The full details will be reported elsewhere.

Other includes pulmonary fibrosis, anaemia, anxiety and non respiratory infections. (CCF – congestive cardiac failure, LRTI – lower respiratory tract infection, COPD – chronic obstructive pulmonary disease)

Figure 1: Range of diagnoses in patients over sixty years describing some breathlessness at presentation to the emergency department.

Agreement between bedside and blinded interpretation of the protocol was calculated with linear weighted kappa using Vassarstat online calculator.¹⁹ Ninety-five per cent confidence intervals were calculated.

Results

Of 230 eligible patients 13 were excluded initially. Of the excluded patients, three had scans reported at the bedside but not saved, two had inadequate saved scans and two were not scanned at all (reason unknown).

A further six patients were incorrectly enrolled (wrong age (1), did not report breathlessness (1) no chest x-ray (2) or intervening fluid management (2)).

The participants ($n = 217$) had had a median age of 76 years (IQR 70 to 83 years), and 45% ($n = 98$) were female. Positive LUS scans estimated the incidence of pulmonary oedema at 23%. The spread of diagnoses is illustrated in Figure 1.

One experienced emergency physician performed 155 studies, while eleven inexperienced sonologists contributed 62 studies (28 proctored, 34 unproctored). The inexperienced sonologists had worked in emergency medicine between 7 and 25 years. Four were emergency physicians (EMP), four were senior medical officers (SMO) and three were senior registrars (Figure 2). There was no seniority pattern predicting uptake of LUS. Those who performed the most scans either used bedside ultrasound for other purposes (EMP number 6, SMO 10 and 11) or had just recently attended FAST training (REG 7 and 9, SMO 8).

Level of agreement between participants

Objectivity of the protocol was demonstrated by the 'excellent agreement' between the experienced practitioner at the bedside and blinded reviewer. ($\kappa = 0.82$, 95% CI 0.72 to 0.92). Interpretation differed in 12 of 155 scans, of which 3 were designated 'indeterminate' by the bedside sonologist.

Agreement between inexperienced practitioner and blinded

Figure 2: Number of scans performed by each inexperienced clinician.

reviewer was 'good' ($\kappa = 0.70$, 95% CI 0.45 to 0.95). This is subdivided into proctored agreement ($\kappa = 0.65$, 95%CI 0.019 to 1) and unproctored agreement ($\kappa = 0.77$, 95%CI 0.52 to 1). The wide confidence interval in the proctored group was due to the very low number of 'wet' scans (1).

Post hoc review showed that two forms of disagreements predominated among the discordant reports. Borderline calls occurred when B lines were indistinct or barely satisfying the protocol, whereas dual pathology caused many B lines associated with pleural line abnormalities.

Discussion

When considering a new test, physicians must ask two questions – "How does the test perform?" and "How well do clinicians perform the test?" The first evaluates the test inaccuracy, the second evaluates clinician limitations. Then, when implementing a point of care (POC) test, physicians must pose the questions, "Can we audit this test?" and "How much teaching is enough?"

International research has been published in depth on the first question^{1-4,15,18} but far smaller samples inform the subsequent credentialing issues.¹³⁻¹⁶ Our study addresses the latter questions.

Our study is one of the few that incorporates inexperienced sonologists both acquiring and interpreting their scans. It becomes one of the largest data sets if the proctored scans are included. This study shows that for a dichotomous question, the LUS protocol is objective, as demonstrated by the excellent agreement of experienced practitioner with the blinded reviewer.

The protocol is compatible with storage and audit of still images. There is acceptable variance estimated for the review of still images. The protocol is simple, as supported by the good agreement with inexperienced practitioners after only 4 hours teaching. Post hoc analysis reveals how the agreement can be improved.

Due to the small numbers in the subgroups, the confidence intervals are wide although calculated levels of agreement are acceptable. This suggests that larger studies with inexperienced practitioners should be safe, particularly if the teaching incorporates proposed improvements.

We performed post hoc analysis of discordant scans. Two main sources of disagreement were identified 'borderline calls' (7) and those with dual pathology (6). 'Borderline calls' occur when practitioner and reviewer disagree over the exact

Figure 3a: Three distinct B lines radiating from smooth pleura to the base of the screen, indicating a 'wet' view.

Figure 3b: Edges of indistinct B lines may be more obvious on a cine loop. The focus is as superficial as allowed by the preset. Use of a standoff may also clarify B lines.

number of B lines, a discrepancy confined to patients with low triage acuity. (Compare Figures 3a and 3a) From observing the proctored scans, it appeared that the inexperienced practitioners found it easier to identify B lines on the moving screen, but then tended to freeze the image to demonstrate that the region was scanned rather than to capture an image of the simultaneous B lines. Inexperienced practitioners were unused to scrolling cine loops, registering, labelling and saving scans. This implies that self-audit practices are underdeveloped.

Two considerations diminish the impact of this discrepancy. The clinical consideration is that borderline counts are easily recognised by the practitioner who then has time to arrange further investigations in the undistressed patient. The academic consideration is more complex, and involves the error inherent in the test protocol. Liteplo, *et al.*¹⁵ used ROC curves to demonstrate that there is no discrete number of B lines that can always differentiate cardiac from pulmonary conditions. An exact count will not distinguish the two conditions because they manifest

Figure 4: An irregular pleural surface suggests underlying lung pathology. Pulmonary oedema may coexist with pleural line abnormalities, but increased interstitial resistance makes this less likely. The most basic protocol does not recognise pleural line abnormalities. This patient had diffuse pneumonitis.

Figure 5: This area of bright confluent B lines occupies approximately 50% of the space between the rib shadow, therefore would be considered the equivalent of 5 B lines, hence a 'wet' view.

with overlapping findings. This means that the accuracy of the test is dependent on the disease prevalence in the population.²⁰ As the population broadens to include a wider spectrum of disease, test accuracy may decline. The issue of 'borderline calls' implies that it is the dichotomous form of the test that limits its accuracy, rather than the experience of the practitioner. Further research already includes cardiac or inferior vena cava views.²¹⁻²⁴

The second major source of disagreement was among the sicker patients with multiple B lines as well as other signs of

lung pathology. There was inadequate provision made to define dual pathology prior to the study, such that the experienced practitioner was influenced by pleural line abnormalities (Figure 4) whereas the blinded reviewer was not. Furthermore, the method for quantifying confluent B lines was not published until early 2012¹⁷ after recruitment. It is now suggested that we estimate the percentage of rib space occupied by the 'white' or 'shining lung' then divide by ten to give the number of B lines represented by the confluence (Figure 5).

Limitations

Overnight recruitment of the sickest patients was prevented by lack of supervision and low staffing levels. This deficit was foreseen, but considered acceptable for two reasons. First, exclusively high acuity populations have already been examined by experts^{1,3,18} and our aim was to subject the method to a broader population, rather than duplicate studies in an acutely ill population. Second, inter-rater agreement should not be dependent on disease prevalence.

This research has not taken aptitude into account. Radiologists, sonographers and POC sonologists tend to self-select.¹¹ Our teaching was minimal, but only those interested continued to collect cases, leading to a potential selection bias.

Further research

The goal of the entire project is to define how, when and where LUS can safely substitute for chest x-ray. Lung ultrasound is unlikely to replace chest x-ray as it does not image the hilum. However, it may inform emergent treatment while awaiting the arrival of a radiographer, or postponement of radiography until daylight hours in non-tertiary centres.

The next step is to calculate a valid diagnostic accuracy for LUS in all patients for whom a chest x-ray is ordered to exclude pulmonary oedema. LUS is unlikely to replace chest x-ray, but in hospitals where expert reporting is delayed, interim LUS may prove more reliable than an inexpertly read chest x-ray.

Conclusion

This research is directed at emergency departments interested in teaching lung ultrasound. It tests emergency practitioners who have a good knowledge of pathophysiology, but rudimentary ultrasound skills. It shows that a very basic protocol can be taught in four hours, saved and re-reported with good reproducibility.

Knowing the error inherent in any new technique is necessary but not sufficient when determining how to introduce the technique into clinical practice. Physicians also must know their own limitations, and their learning curve.

This LUS protocol is a very small step toward expert lung sonography. Perhaps the greatest benefit of the basic protocol is not in its precision, but in its utility as 'training wheels'.

Acknowledgements

This research was made possible by a grant from the Queensland Emergency Medicine Research Fund.

The authors would like to express their gratitude to Assoc Prof James Rippey for his expert review and suggestions toward this publication.

The authors would like to thank Ms Jennifer Cochrane, Ms Elisabeth Cash and Ms Lyn Burrows for research assistance and library support.

The authors would like to acknowledge Ms Sue Davies and the Australian Institute of Ultrasound for facilitating the novice training sessions.

The authors thank the patients for their understanding and encouragement, and Doctors Ian Brandon, Shane Brun, Daniel Bitmead, Duncan Murray, Peter Ivermee, Quentin Shaw, Bilesha Jayawardena, Tammra Warby, Courtney Pennuto, Karen Richards and Khalid Yousif.

References

- 1 Agricola E, Bove T, Oppizzi M, Marino G, Zangrillo A, Margonato A, *et al.* 'Ultrasound comet-tail images': a marker of pulmonary edema. A comparative study with wedge pressure and extravascular lung water. *Chest* 2005; 127 (5): 356–363.
- 2 Cibinel G, Casoli G, Elia F, Pandoan M, Pivetta E, Lupia E. Diagnostic accuracy and reproducibility of pleural and lung ultrasound in discrimination cardiogenic causes of acute dyspnoea in the Emergency Department. *Intern Emerg Med* 2012; 7: 65–70.
- 3 Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby J. Comparative Diagnostic Performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. *Anesthesiology* 2004; 100 (1): 9–15.
- 4 Volpicelli G, Mussa A, Garofalo G, Cardinale L, Casoli G, Perotto F, *et al.* Bedside lung ultrasound in the assessment of alveolar-interstitial syndrome. *Am J Emerg Med* 2006; 24: 689–96.
- 5 McCullough P, Nowak R, McCord J, Hollander J, Herrmann H, Steg Pea. B-type natriuretic peptide and clinical judgement in emergency diagnosis of heart failure. *Circulation* 2002; 106: 416–22.
- 6 Vandenbos F, Mejdoubi N, Pinier Y, Epuron J, Grelier M. BNP tests in the emergency department to diagnose congestive heart failure. *Presse Medicale* (French, English Abstract) 2006; 35 (11 pt 1): 1625–31.
- 7 ACEM. Policies and Guidelines 2012. Available at <http://www.acem.org.au/infocentre.aspx?docId=59>. Accessed 2 December 2012.
- 8 ACEP Policy Statement: Emergency Ultrasound guidelines. *Ann Emerg Med* 2009; 53: 550–70.
- 9 ASUM. Policy and standards 2012. Available at <http://www.asum.com.au/newsite/Resources.php?p=Policy>. Accessed 25 November 2012.
- 10 Huang S, McLean A. Do we need a critical care ultrasound certification program? Implications from an Australian Medico-legal perspective. *Crit Care Med* 2010; 14 (3): 313.
- 11 Langlois S. Focused ultrasound training for physicians. *Crit Care Med* 2007; 35 Supplement: S138–43.
- 12 Havelock T, Teoh R, Laws D, Gleeson F. Pleural procedures and thoracic ultrasound: British Thoracic Society pleural disease guideline. *Thorax* 2010; 65 Supplement 2: ii61–76.
- 13 Bedetti G, Gargani L, Corbisiero A, Frassi F, Poggianti E, Mottola G. Evaluation of ultrasound lung comets by hand-held echocardiography. *Cardiovasc Ultrasound* 2006; 4: 34.
- 14 Brooke M, Walton J, Scutt D, Connolly J, Jarman B. Acquisition and interpretation of focused diagnostic ultrasound images by ultrasound-naïve advanced paramedics: trialling a PHUS education programme. *Emerg Med J* 2012; 29: 322–26.
- 15 Liteplo A, Marill K, Villen T, Miller R, Murray A, Croft P, *et al.* Emergency thoracic ultrasound in the differentiation of the etiology of shortness of breath (ETUDES): Sonographic B-lines and N-terminal Pro-brain-type Natriuretic Peptide in diagnosing congestive heart failure. *Acad Emerg Med* 2009; 16: 201–10.
- 16 Tutino L, Cianchi G, Barbani F, Batacchi S, Cammelli R, Peris A. Time needed to achieve completeness and accuracy in bedside lung ultrasound reporting in Intensive Care Unit. *Scand J Trauma Resuscitation and Emergency Medicine* 2010; 18: 44.
- 17 Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein D, Mathis G, Kirkpatrick A. International evidence-based recommendations for point of care lung ultrasound. International consensus conference on lung ultrasound; 2012: Intensive Care Medicine.
- 18 Prosen G, Klemen P, Strnad M, Grmec S. Combination of lung ultrasound (a comet-tail sign) and N-terminal pro-brain natriuretic peptide in differentiating acute heart failure from chronic obstructive pulmonary disease and asthma as cause of acute dyspnea in prehospital emergency setting. *Crit Care* 2011; 15: R114.

- 19 Lowry R. VassarStats: Web Site for Statistical Computation; Clinical calculators for sensitivity, specificity, and confidence intervals for difference in proportion. 1998–2013. Available at <http://faculty.vassar.edu/lowry/VassarStats.html%5D>. Accessed 17 April 2013.
 - 20 Simundic AM. Measures of diagnostic accuracy: basic definitions. *eJIFCC* 2008; 19 (4): e pub.
 - 21 Gargani L, Pang P, Davis E, Schumacher A, Sicari R, Picano E. Lung ultrasound to rule out congestive heart failure in the emergency department. *Eur Heart J* 2012; 33. Abstract supplement: 249–50.
 - 22 Kimura B, Yogo N, O'Connell C, Phan J, Showalter B, Wolfson T. Cardiopulmonary limited ultrasound examination for 'quick-look' bedside application. *Am J Cardiol* 2011; 108: 586–90.
 - 23 Mantha K, Berry B, Anderson T, Bastani A, Bahl A. The PIC Protocol: A Novel Emergency Department Ultrasound Protocol for Patients Presenting With Dyspnea. William Beaumont Hospital, Royal Oak, MI. *Ann Emerg Med* 2011; S229.
 - 24 Stewart VM, Bjornsson H, Ravi B, Byars D, Clinton M. B-RIPPED ultrasonography scan for evaluation of emergency department patients with shortness of breath. *Ann Emerg Med* 2012; 60 (4). Supplement 1: S18.
-