BMJ Open

The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure: A Systematic Review.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-007838
Article Type:	Research
Date Submitted by the Author:	04-Feb-2015
Complete List of Authors:	Hew, Mark; Alfred Hospital, Allergy, Immunology & Respiratory Medicine Corcoran, John; Oxford University Hospitals NHS Trust, Oxford Centre for Respiratory Medicine Harriss, Elinor; University of Oxford, Bodleian Health Care Libraries Rahman, Najib; Oxford University, Oxford Centre for Respiratory Medicine Mallett, Susan; National Institute for Health Research Diagnostic Evidence Co-operative Oxford,
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice, Radiology and imaging
Keywords:	Adult thoracic medicine < THORACIC MEDICINE, Respiratory infections < THORACIC MEDICINE, Ultrasound < RADIOLOGY & IMAGING, Adult intensive & critical care < ANAESTHETICS

The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure: A Systematic Review.

Authors

1. Mark Hew PhD MSc FRACP
Allergy Immunology and Respiratory Medicine, Alfred Hospital, Melbourne, Victoria, Australia

2. John P Corcoran MA MRCP
Oxford Centre for Respiratory Medicine, Oxford University Hospitals NHS Trust, Oxford, UK, Oxford, United Kingdom

3. Elinor Harriss, MA MSc
Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom

4. Najib M Rahman DPhil MSc MRCP
Oxford Respiratory Trials Unit, University of Oxford, Oxford, United Kingdom

5. Susan Mallett BA DPhil
Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author

A/Prof Mark Hew, Head of Allergy, Asthma & Clinical Immunology Service, Alfred Hospital, 55 Commercial Road, Prahran, Victoria 3004 Australia.

m.hew@alfred.org.au Tel 613- 90762934 Fax 613- 90762245

Key words

Pneumonia Respiratory Infection ARDS Assisted ventilation

Abstract word count 299, Manuscript word count 2951, References 24, Tables 3, Figures 7

Structured Abstract.

Objectives: (i) Summarise chest ultrasound accuracy to diagnose radiological consolidation, referenced to chest computer tomography (CT) in patients with acute respiratory failure (ARF). (ii) Directly compared ultrasound to chest X-ray.

Setting: Hospitalised patients.

Participants: Studies were eligible if adult participants in respiratory failure underwent chest ultrasound to diagnose consolidation referenced to CT. Exclusion: (i) Not primary study (ii) Not respiratory failure (iii) Not chest ultrasound (iv) Not consolidation (v) Translation unobtainable (vi) Unable to extract data (vii) Unable to obtain paper. Four studies comprising 224 participants met inclusion.

Outcome measures: As planned, paired Forest plots display 95% confidence intervals of sensitivity and specificity for ultrasound and chest X-ray. Sensitivity and specificity from each study are plotted in receiver operator characteristics (ROC) space. Meta-analysis was planned if studies were sufficiently homogeneous and numerous (≥ 4). Although this numerical requirement was met, meta-analysis was prevented by heterogeneous units of analysis between studies.

Results: All studies were in intensive care, with either a high risk of selection bias or high applicability concerns. Studies had unclear or high risk of bias related to use of ultrasound. Only one study clearly performed ultrasound within 24 hours of respiratory failure diagnosis. Ultrasound sensitivity ranged from 0·91 (95% CI 0·81-0·97) to 1·00 (95% CI 0·95-1·00). Specificity ranged from 0·78 (95% CI 0·52-0·94) to 1·00 (0·99-1·00). In two studies, chest X-ray had lower sensitivity than ultrasound, but there were insufficient patients to compare specificity.

Conclusions: Four small studies suggest ultrasound is highly sensitive and specific for consolidation in acute respiratory failure, but high risk of bias and concerns about applicability in all studies may have inflated diagnostic accuracy. Further robustly-designed studies are needed to define the role of ultrasound in this setting.

Registration: http://www.crd.york.ac.uk/PROSPERO/ (CRD42013006472

Article Summary: Strengths & Limitations of this study

Strengths

- Comparison of sonographic consolidation to a reference of radiological consolidation
- Restricted to studies with reliable gold standard of chest CT
- Examination of the influence of units of analysis on diagnostic accuracy reporting

Limitations

- Small number of eligible studies
- Meta-analysis prevented by heterogeneous units of analysis

INTRODUCTION.

Acute respiratory failure (ARF) is common and deadly. Published incidence rates[1,2] suggest approximately 50,000 patients each year in the UK at the severest end of the ARF spectrum may require ventilatory support. A quarter of these have underlying pneumonia[2], and face mortality rates as high as 50%[3].

Mortality escalates further when the cause of ARF is misdiagnosed, which occurs in one in five patients[4] due in part to imaging limitations. Patients are difficult to position for chest X-ray[5] resulting in suboptimal films which may miss consolidation[6], the commonest pattern of pneumonic infiltrate. Conversely, chest computer tomography (CT) is highly sensitive but entails risks of transporting critically ill patients[7]. Both shortcomings of traditional imaging may be overcome by chest ultrasound. Unlike X-ray, ultrasound does not require optimal patient positioning. Unlike CT, ultrasound can be brought to the bedside.

Narrative reviews[8], consensus guidelines[9] and systematic reviews[10,11] all advocate the use of ultrasound to diagnose pneumonia but crucially, most studies of ultrasound accuracy have not examined patients in ARF settings. In those that do, the reference standard is often the final clinical diagnosis, risking incorporation bias if ultrasound itself forms part of that standard.

Further confusion arises when ultrasound accuracy studies use 'pneumonia' as the target condition. While ultrasound can diagnose consolidation (an *imaging* finding), only clinicians diagnose pneumonia (a *clinical* diagnosis) by expertly blending imaging findings with available clinical information[12]. However, such clinical incorporation bias distorts estimates of ultrasound accuracy. Instead, the most appropriate target condition for the imaging finding of sonographic consolidation is another imaging finding, in this case radiographic consolidation on chest CT.

To address these key issues, we undertook a systematic review to summarise the accuracy of chest ultrasound to diagnose radiological consolidation, referenced to chest CT, in the specific setting of

hospitalised patients with acute respiratory failure (ARF). We also directly compared ultrasound to chest X-ray, the commonest screening test for consolidation in acute respiratory failure. We excluded paediatric studies because children have a different range of aetiologies for ARF[13].

METHODS.

The protocol was registered at http://www.crd.york.ac.uk/PROSPERO/ (review registration number CRD42013006472) and attached as a supplement; key points are summarised here. The review is reported according to PRISMA guidelines (Preferred Reporting Items for systematic reviews and Meta-Analyses[14]).

Inclusion criteria.

Studies: Cohort, cohort with nested one-gate case-control studies (participants with and without consolidation), randomised controlled trials (ultrasound versus chest X-ray). *Timing:* 24 hours or less between acute respiratory failure (ARF) diagnosis and ultrasound scanning. If insufficient such studies were found, studies where ultrasound was performed more than 24 hours after diagnosis of ARF would be included. *Participants:* Adults (age 18 or greater) admitted to any hospital setting with ARF. Studies excluded if patients discharged home directly from the Emergency Department within 24 hours without ward admission. Acute respiratory failure (ARF) defined as: i) arterial partial pressure of oxygen (PaO_2) < 60 millimetres of mercury, without supplemental oxygen, or; ii) arterial oxygen saturation of < 90% on pulse oximetry, without supplemental oxygen, or; iii) supplemental oxygen required to prevent i) or ii), or; iv) author diagnosis of acute respiratory failure. *Index:* B-mode ultrasound examining lungs and pleura. *Comparator:* Studies comparing chest X-ray to chest ultrasound. Studies evaluating only chest X-ray excluded. *Target condition*: Radiological consolidation. Studies referencing chest ultrasound to a clinical diagnosis of pneumonia excluded.

Reference standard. Chest CT, defined as helical CT to examine the thorax. Studies could be included if only some patients received CT, but only when data could analysis in the CT subgroup.

Exclusion criteria.

The following hierarchy was employed: (i) Not a primary study (ii) Not respiratory failure (iii) Not chest ultrasound (iv) Not consolidation (v) Translation unobtainable (vi) Unable to extract 2 X 2 data (vii) Unable to obtain paper through both Bodleian (University of Oxford) and British Libraries.

Search.

A healthcare librarian assisted with strategy development. Several iterations were trialled using two reference studies. The full search was run on 22 October 2013, in Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present, Ovid Embase 1974 to 2013 October 21, and Web of Knowledge Science Citation Index Expanded (Appendix 1). An update search was run on 6 August 2014 prior to publication. Filters were not used and no language or date restrictions applied. Only published studies were included. Reference lists, citation searches, citing alerts in electronic journals and the 'related articles' feature in PubMed were also used.

Study Selection.

Titles and abstracts were screened according to inclusion and exclusion by two reviewers independently of each other, and results pooled. Full texts were assessed for inclusion independently by two reviewers. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Data Collection.

Pre-specified data extraction forms (Appendix 2) were developed [15], trialled in one study and modified. Data items included participants, index, comparator, reference, flow and diagnostic

performance. Data was independently extracted by two reviewers. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Quality assessment.

QUADAS2 (the Revised Tool for Quality Assessment of Diagnostic Accuracy Studies) [16] was tailored for this review (Appendix 3). Rating guidelines were developed, piloted in one included study, and applied to remaining studies by both reviewers independently. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Analysis Plan.

Most studies were expected to analyse patients, although we anticipated beforehand that studies might report results for each lung. We did not anticipate reporting by lung region. We considered analysis of any units other than patients as biased, since one lung (or lung region) of a patient is not independent of another.

Paired Forest plots were used to display 95% confidence intervals of sensitivity and specificity. Sensitivity and specificity from each study were plotted in receiver operator characteristics (ROC) space. Meta-analysis was planned if studies were sufficiently homogeneous and numerous (≥ 4). Although this numerical requirement was met, meta-analysis was prevented by heterogeneous units of analysis between studies. Details of the planned meta-analysis are available at the registered protocol. Where ultrasound was compared to chest X-ray, sensitivity and specificity for both tests were plotted in ROC space using Revman 5[17]. Tests of unpaired proportions for large samples were used to compare tests within same study, as insufficient data was available for paired comparison.

RESULTS.

Study selection.

Figure 1 shows the PRISMA flowchart. Totals from both (original and update) searches are combined. Four studies met inclusion criteria[18-21]. 2 X 2 contingency tables could be extracted from all included studies.

We were concerned regarding possible duplication or overlap of cohorts because two included papers had the same first author and year of publication (Lichtenstein, 2004a[18], Lichtenstein, 2004b[19]). However, we found key differences between the studies which rendered this unlikely (Table 1).

Study characteristics.

Table 2 summarises settings and patient characteristics of studies. All four studies were intensive care cohorts, but each reported different severity measures making comparison of acute respiratory failure (ARF) severity difficult.

Table 3 summarises ultrasound methods and units of analyses. Only one study[18] met our criterion for preferred studies, with ultrasound undertaken within 24 hours of ICU admission (and thus probably of ARF diagnosis). In the other studies, timing of ultrasound in relation to ARF diagnosis was not stated. Scanning protocols were similar across all studies; each lung was divided into six regions; anterior, lateral, and posterior; with upper and lower divisions.

Risk of bias and applicability concerns.

Figures 2 and 3 summarise the quality assessment of individual primary studies. Applicability was considered in relation to our review question, which examined the diagnostic accuracy of chest ultrasound for CT-detected radiographic consolidation in adults with acute respiratory failure (ARF).

Lichtenstein, 2004a[18].

Selection. There was a low risk of selection bias due to consecutive recruitment of patients with acute respiratory distress syndrome. However, since acute respiratory distress syndrome represents the highest acuity of ARF, concerns regarding applicability were high.

Index Test. The risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information, (although blinded to CT). Applicability concerns were low.

Reference Test. CT interpretation was blinded to ultrasound results, but clinical data may have been available, so risk of bias was unclear. Applicability concerns were low.

Flow & timing. CT was performed within six hours after ultrasound. Risk of bias was low.

***Lichtenstein, 2004b*[19].**

Selection. The risk of selection bias was high because subjects were recruited on the clinical need for CT. Most patients were likely in ARF based on their need for intubation and specific diagnoses; however, ARF was not specifically stated so applicability concerns were unclear.

Index Test. The risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information, (although blinded to CT). Applicability concerns were low.

Reference Test. CT interpretation was blinded to ultrasound results, but clinical data may have been available, so risk of bias was unclear. Applicability concerns were low.

Flow & timing. CT was performed within six hours after ultrasound. The risk of bias was low.

***Xirouchaki, 2011*[20].**

Selection. Risk of selection bias was high as subjects were recruited on their clinical need for CT. Again, most patients were likely in ARF based on their need for intubation and specific diagnoses; however, ARF was not specifically stated so applicability concerns were unclear.

Index Test. Risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information (although blinded to CT). Applicability concerns were low.

Reference Test. CT was interpreted blinded to both clinical information and ultrasound results, giving a low risk of bias. Applicability concerns were low.

Flow & timing. CT was performed no longer than six hours after ultrasound. Risk of bias was low.

***Refaat, 2013*[21].**

Selection. Recruitment was consecutive and exclusions were reasonable, giving a low risk of selection bias. The aetiological diagnoses in this patient group were restricted to a subgroup of respiratory failure aetiologies, causing high applicability concerns.

Index Test. The risk of bias was high as the sonographer had access to clinical information. There were low applicability concerns.

Reference Test. CT was interpreted blind to ultrasound results, giving a low risk of bias. Applicability concerns were low.

Flow & timing. CT was performed within 24 hours of ultrasound; risk of bias was low.

Analysis.

Ultrasound.

Sensitivity for diagnosing CT-detected consolidation ranged from 0·91 (95% CI 0·81-0·97) to 1·00 (95% CI 0·95-1·.00, Figure 4). Specificity ranged from 0·78 (95% CI 0·52-0·94) to 1·.00 (0·99-1·00).

Ultrasound compared to chest X-ray.

Two studies, both of which only included ventilated patients[18,20] (Figures 4, 5 & 6) evaluated both ultrasound and chest X-ray in the same patient populations, the best study design to compare tests[22]. In both studies the sensitivity of ultrasound was significantly greater than that of chest X-ray; 0·24 higher (95% CI 0·15 to 0·34, p<0·.0001; Figures 4, 6) in the first study[18] and 0·62 (95% CI 0·50 to 0·74, p<0·0001) in the second[20] (Figures 4, 6). When compared using 12 lung regions per patient, specificity was higher for ultrasound (0·049, 95% CI 0·023 to 0·075, p=0·0003) in the first study[18], but lower in in the second (-0·049, 95% CI -0·23 to -0·075, p=0·0003)[20]. Specificity compared at 2 lung regions per patient lacked sufficient power to detect a difference (Figure 4).

Impact of unit of analysis.

Three different units of analyses were reported across four studies (Table 3). Only one study used the patient as a unit of analysis[21]. A second study[19] used only the lung (two per patient), and a third[18] used only lung regions (12 per patient).

The fourth study[20] reported its results using the lung as the unit of analysis, but provided additional data using lung regions in an electronic supplement. This provided the opportunity to study the impact of different units of analysis on test characteristics within same dataset (Figure 7).

Importantly, for both ultrasound and chest X-ray, changing the unit of analysis from lung to lung region reduced sensitivity but enhanced specificity, and gave more precise estimates of accuracy (narrower confidence intervals). It also inflated the prevalence of consolidation.

Reported sources of ultrasound and chest X-ray error.

In one study[19], five of six false negative ultrasounds were in patients with posteriorly placed consolidation. This study evaluated patients only in the supine position, which may hinder the detection of posterior consolidation.

In another study[20], all four false positive ultrasounds detected only small areas of consolidation. This study only used a tissue-like pattern to diagnose consolidation which may have reduced specificity.

None of the studies proposed reasons for false positive or false negative chest X-ray results in their discussion.

Synthesis of results.

Meta-analysis was not performed due to heterogeneous units of analysis across studies.

Additional analyses.

Heterogeneity could not be explored due to the small number of studies, apart from comparisons between different units of analysis.

DISCUSSION.

Summary of evidence.

In four small studies, the reported sensitivity and specificity of ultrasound for CT-diagnosed consolidation was high among hospitalised patients with acute respiratory failure (ARF). Ultrasound sensitivity was greater than for chest X-ray, in two studies directly comparing both methods in the same patient populations. However, paired comparisons in individual patients which are the best evidence for comparing tests were not available.

This review identified four quality issues that impact the reported test accuracy of ultrasound in included studies. Firstly, patient selection in every eligible study was either at high risk of bias, or had concerns about applicability to our systematic review. These concerns included recruitment of participants in ICU at the severest acuity of ARF (spectrum bias), restriction to limited ARF aetiologies, and non-consecutive recruitment. The sensitivity of ultrasound for consolidation may

thus be markedly poorer in unselected populations with less severe ARF (and lower burdens of consolidation) or a wider range of ARF aetiologies.

Secondly, in no study were sonographers clearly blinded to clinical data. This is pertinent because sonographers (who in three studies were actually clinicians) could have integrated bedside clinical data with ultrasound evaluation, artificially inflating ultrasound sensitivity.

Thirdly, only one study specified that ultrasound was performed within 24 hours of ICU admission (and presumably, of ARF diagnosis). The more time elapses before ultrasound is performed, the more likely lung consolidation would progress to a detectable extent, but the less likely the test result would improve patient outcome. This would boost reported ultrasound sensitivity but overstate its utility as an initial test.

Fourthly, the two studies comparing ultrasound to chest X-ray were undertaken wholly among ventilated patients. This spectrum bias would augment ultrasound sensitivity since patients would be more likely to have extensive (and more easily detectable) consolidation. The necessarily supine chest X-rays would render films less sensitive for consolidation, again exaggerating the benefit of ultrasound.

The variable units of analyses employed across studies also introduce additional concerns. Different units of analyses had an evident effect on test accuracy. The use of lung regions for analysis (as opposed to lungs) diminished sensitivity, inflated specificity, and gave the misleading appearance of greater precision. Another drawback of different units of analyses across studies is that meta-analysis of results would be misleading because studies using lung regions would have undue numerical weight. For these reasons, we highly recommend future studies be conducted and reported always including patients as the unit of analysis.

In addition, we strongly recommend future studies should compare different tests in the same patients and present results as 2 by 2 tables of paired results separately in disease-positive and

disease negative patients. This is important to understand whether false-positive and false-negative results occur in the same or different patients, and to design subsequent studies.

Compared to previous systematic reviews[10,11], the distinguishing features of our review were: i) the emphasis on a single clinical presentation ie ARF; in this 'high stakes' patient group, the additional resources required to perform ultrasound are better justified than in less severe clinical presentations; ii) the requirement for a CT reference; providing greater confidence in estimates of diagnostic accuracy; iii) the focus on a single radiographic abnormality *i.e.* consolidation rather than the clinical diagnosis of pneumonia, removing the risk of bias of incorporating clinical information into the target condition, and; iv) the pre-registered systematic review protocol.

Limitations.

This review is limited by the small number of studies meeting inclusion. Four studies were performed by three investigator groups, limiting generalizability to other clinical environments. Where more than one ultrasound sign was used to diagnose consolidation, test characteristics of individual signs for consolidation were not assessed. The small number of studies and different units of analyses prevented meta-analysis, exploration of clinical and methodological heterogeneity, and pooled comparisons between ultrasound and chest X-ray.

Conclusion.

Based on a small body of evidence at high risk of selection bias and index test bias, ultrasound is both sensitive and specific for CT-detected consolidation in acute respiratory failure. Heterogeneous units of analyses between studies limited comparisons between studies.

While ultrasound may have a role as an add-on test in ARF when the chest X-ray is negative for consolidation, this possibility is tempered by the narrow evidence base available associated with substantial risks of bias and applicability concerns. We conclude there is insufficient evidence to

support the widespread introduction of ultrasound to detect pneumonia in hospitalised patients diagnosed with acute respiratory failure.

Robustly designed studies are needed, controlling for the fundamental biases discussed above. They should aim to determine if an add-on, or replacement, test strategy is truly beneficial and identify clinical determinants of test accuracy. Ultrasound should be compared to current methods and also to emerging diagnostic alternatives using biomarkers[23] and other novel imaging[24]. The feasibility of implementing ultrasound should also be studied, coupled with clinical and cost-effectiveness modelling.

Author contributions.

MH- Original idea, first draft of protocol, first reviewer, first draft of manuscript. SM- Supervised protocol development and overall review. JPC- Second reviewer. NMR- Third reviewer. EKH- Search design and execution. All the authors analysed the data and wrote the manuscript.

Conflict of Interests

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

1. Lewandowski K, Metz J, Deutschmann C, et al. Incidence, severity, and mortality of acute respiratory failure in Berlin, Germany. Am J Respir Crit Care Med. 1995 Apr;151(4):1121-5.

2. Luhr OR, Antonsen K, Karlsson M, et al. Incidence and mortality after acute respiratory failure and acute respiratory distress syndrome in Sweden, Denmark, and Iceland. The ARF Study Group. Am J Respir Crit Care Med. 1999 Jun;159(6):1849-61.

3. Woodhead M, Welch CA, Harrison DA, Bellingan G, Ayres JG. Community-acquired pneumonia on the intensive care unit: secondary analysis of 17,869 cases in the ICNARC Case Mix Programme Database. Crit Care. 2006;10 Suppl 2:S1.

4. Ray P, Birolleau S, Lefort Y, et al. Acute respiratory failure in the elderly: etiology, emergency diagnosis and prognosis. Crit Care. 2006;10(3):R82.

5. Ovenfors C, Hedgecock MW. Intensive care unit radiology: problems of interpretation. Radiol Clin North Am 1978:16:407-439

6. Khan AN, Al-Jahdali H, Al-Ghanem S, Gouda A. Reading chest radiographs in the critically ill (Part I): Normal chest radiographic appearance, instrumentation and complications from instrumentation. Ann Thorac Med. 2009 Apr;4(2):75-87.

7. Bouhemad B, Zhang M, Lu Q, Rouby JJ. Clinical review: Bedside lung ultrasound in critical care practice. Crit Care. 2007;11(1):205. Review.

8. Hew M, Heinze S. Chest Ultrasound in Practice: a review of utility in the clinical setting. Intern Med J. 2012 Aug;42(8):856-65.

9. Volpicelli G, Elbarbary M, Blaivas M, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med. 2012 Apr;38(4):577-91.

10. Hu QJ, Shen YC, Jia LQ, et al. Diagnostic performance of lung ultrasound in the diagnosis of pneumonia: a bivariate meta-analysis. Int J Clin Exp Med. 2014 Jan 15;7(1):115-21

11. Chavez MA, Shams N, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. Respir Res. 2014 Apr 23;15:50.

12. British Thoracic Society (BTS) Community Acquired Pneumonia in Adults Guideline Group. Guidelines for the management of community acquired pneumonia in adults: update 2009.October 2009 Vol 64 Supplement III.

13. Schneider J, Sweberg T. Acute respiratory failure. Crit Care Clin. 2013 Apr;29(2):167-83.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097.

15. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. AJR Am J Roentgenol. 2003 Jul;181(1):51-5. Review.

16. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.

17. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

18. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology. 2004 Jan;100(1):9-15.

19. Lichtenstein DA, Lascols N, Mezière G, Gepner A. Ultrasound diagnosis of alveolar consolidation in the critically ill. Intensive Care Med. 2004 Feb;30(2):276-81.

20. Xirouchaki N, Magkanas E, Vaporidi K, et al. Lung ultrasound in critically ill patients: comparison with bedside chest radiography. Intensive Care Med. 2011 Sep;37(9):1488-93.

21. Refaat R, Abdurrahman L. The diagnostic performance of chest ultrasonography in the up-to-date work-up of the critical care setting. The Egyptian Journal of Radiology and Nuclear Medicine (2013) 44, 779–789.

22. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med. 2013 Apr 2;158(7):544-54

23 Bozinovski S, Hutchinson A, Thompson M, et al. Serum amyloid a is a biomarker of acute exacerbations of chronic obstructive pulmonary disease.Am J Respir Crit Care Med. 2008 Feb 1;177(3):269-78.

24. Masaryk T, Kolonick R, Painter T, Weinreb DB. The economic and clinical benefits of portable head/neck CT imaging in the intensive care unit. Radiol Manage. 2008 Mar-Apr;30(2):50-4.

TABLES

Table 1. Differences in the two studies by Lichtenstein.

	Lichtenstein 2004 a	Lichtenstein 2004 b
Institution	Pitié-Salpétrière Hospital (stated in text)	Hopital Ambroise-Pare (implied by; author affiliation; acknowledgement of the ICU department head; acknowledgement of the Radiology department head where scans took place)
Type of ICU	Surgical	Medical
CT scanner used	Tomoscan SR 7000 (Philips, Eindhoven, The Netherlands)	CT Twin Flash (Elscint Limited, Haifa, Israel)
Reason for CT	Research study protocol	Clinical decision
Recruitment Period	Unstated, but inferred as 1993-1997 (from another paper arising from the same CT ARDS study, Puybasset et al, 2000)	Unstated

Table 2. Included studies: patient characteristics.

Author Country	Study type/ Period	Demographics	Setting	Inclusion	Illness severity	Mechanical ventilation
Lichtenstein 2004a France	Cohort, likely 1993-1997	*n*=32, *Age* 58 +/-15 (SD), *M:F* Not stated	Surgical ICU	ARDS, (pneumonia 18, pulmonary contusion 4, aspiration pneumonia 4, fat embolism 1, septic shock 3, cardiopulmonary bypass 2)	Lung injury severity score 2.6 +/- 0.8(SD), *(i.e. severe),* ARDS severity score of 11 +/- 6 (SD), Mortality 42%	All
Lichtenstein 2004b, France	Cohort, Period not stated	*n*=60, *Age* 53 (range 20-84), *M:F* 37:23	Medical ICU	Patients with critical illness requiring chest CT	Not stated	30/60
Xirouchaki 2011, Greece	Cohort, Period not stated	*n*=42, *Age* 57.1 +/-21.5 (SD), *M:F* 34:8	Mixed ICU	Patients with critical illness requiring chest CT (sepsis/multiorgan failure 18, trauma 11, Airways disease 7, pulmonary oedema 2, post-operative respiratory failure 2)	APACHE2 16.5 +/-6.5 (SD)	All
Refaat 2013 Egypt	Cohort, 2012-13	*n*=90, *Age* 50 (45-65), *M:F* 55:35	Chest ICU	Respiratory failure (pneumonic consolidation 16, lung cancer 7, lung metastases 7, pleural effusion 36, pneumothorax 12, hydropneumothorax 6, mesothelioma 7)	Not stated	Not stated

n: number, M: male, F: female, ICU: intensive care unit, ARDS: acute respiratory distress syndrome, CT: computer tomography, APACHE2: Acute Physiology and Chronic Health Evaluation II.

Table 3. Included studies: Ultrasound technique, signs of consolidation and units of analysis

Study	Ultrasound timing	Sono-grapher	Probe/ scanner	Scan position	Scan protocol	US signs of consolidation	Unit of analysis	Consolidation prevalence
Lichtenstein 2004 a	Within 24 hours of ICU admission (approximated to ARF diagnosis)	1 intensivist (of 2), experience not quantified	Micro-convex, Portable (Hitachi 405)	Supine	12 lung regions	Tissue-like pattern, no change in dimensions with respiration. Air bronchograms not mandatory	Lung region (12/ patient)	31% of lung regions
Lichtenstein 2004 b	Unstated	2 intensivists (kappa coefficient 0.89) experience not quantified	Micro-convex, Portable (Hitachi 405)	Supine	12 lung regions	Tissue-like pattern, arising from the pleural line, irregular deep border (regular if lobar), no change in dimensions with respiration. Air bronchograms not used	Lung (2/ patient)	56% of lungs
Xirouchaki 2011	Unstated	1 intensivist, 4 years' experience	Micro-convex, Portable (Hitachi 8500)	Supine & lateral	12 Lung regions	Tissue-like pattern +/- power doppler. Irregular deep border not used	Lung (2/ patient) & Lung region (12/ patient)	24% of lungs, but 79% of lung regions

Refaat 2013	Unstated	1 radiologist, > 7 years' experience	Linear & convex, Portable Shenzhen mindray DP-1100 Plus)	Supine & lateral Erect when possible	12 lung regions	Hypoechoic pattern, non-homogenous echo-texture, irregular shape, serrated margin, air and fluid bronchograms	Patient	18% of patients

FIGURE LEGENDS

Figure 1. PRISMA flow diagram of study identification and selection.

Figure 2. QUADAS2 risk of bias and applicability assessment of individual studies.

Figure 3. QUADAS2 risk of bias and applicability assessment across primary studies.

Figure 4. Sensitivity and specificity of ultrasound and chest X-ray for included studies. Note different units of analyses which preclude pooling of studies: *lung regions* (12 per patient) in Lichtenstein 2004a; *lungs* (2 per patient) in Lichtenstein 2004b and Xirouchaki 2011; and *individual patients* in Refaat 2013.

Figure 5. Sensitivity and specificity of ultrasound and chest X-ray for consolidation.

Figure 6. Direct comparisons for ultrasound and chest X-ray in two individual studies. Units of analysis are lungs (2 per patient) for Xirouchaki 2011, and lung regions for Lichtenstein 2004a (12 per patient).

Figure 7. Impact of units of analyses on sensitivity and specificity of ultrasound and chest X-ray for consolidation. Data from Xirouchaki et al 2011 are stratified according to two units of analysis; 'lung' (2/patient) and 'lung region' (12/patient). In comparison to lung analysis, *lung region* analysis reduces sensitivity but inflates specificity. It also increases total study numbers, giving the appearance of tighter estimates of precision.

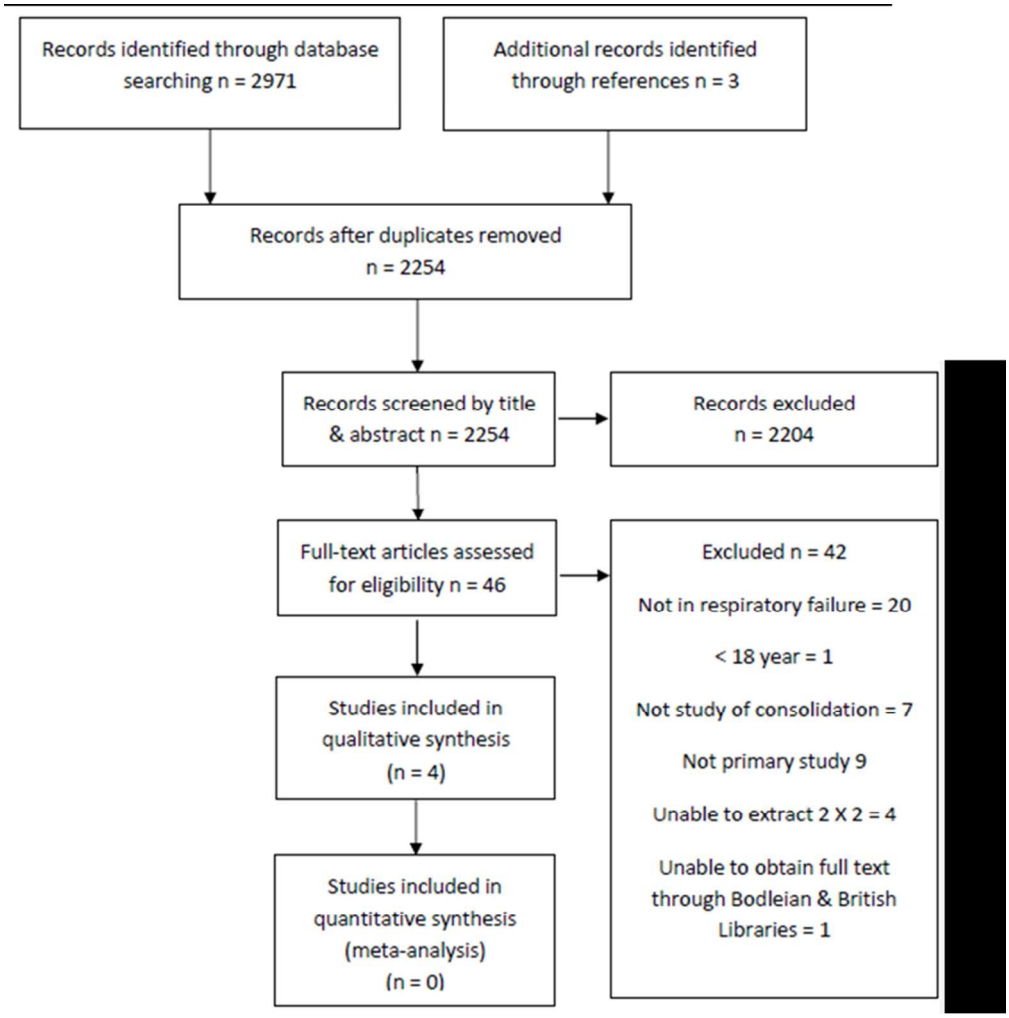


Figure 1. PRISMA flow diagram of study identification and selection.
148x149mm (150 x 150 DPI)

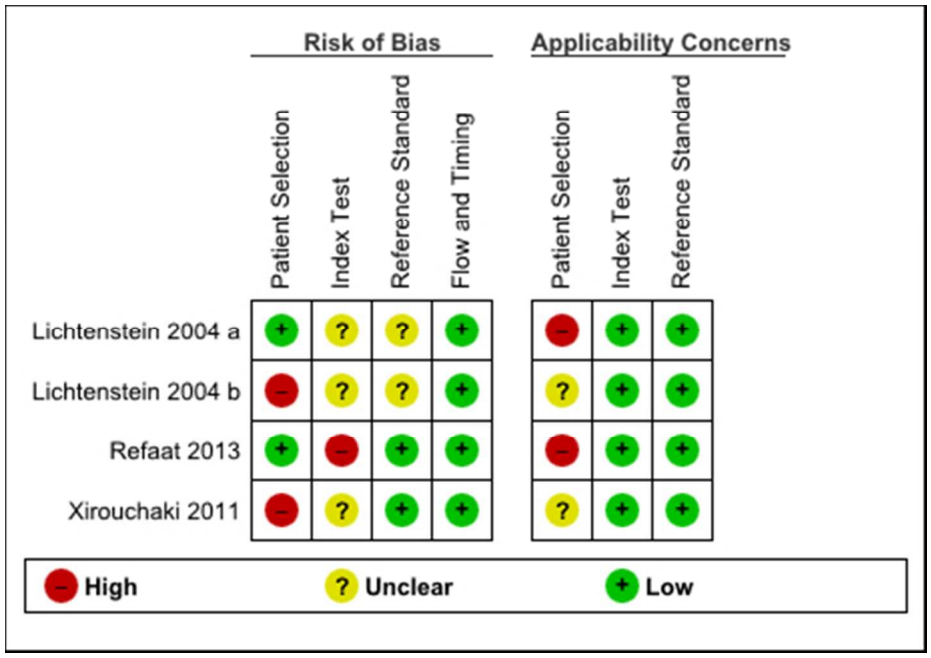


Figure 2. QUADAS2 risk of bias and applicability assessment of individual studies.
121x85mm (96 x 96 DPI)

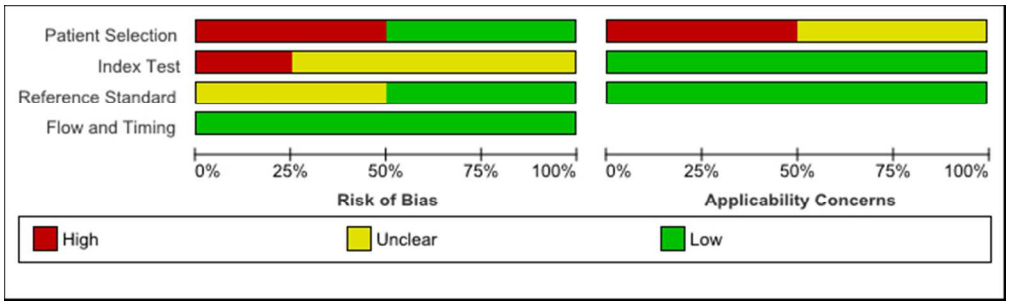


Figure 3. QUADAS2 risk of bias and applicability assessment across primary studies.
179x53mm (96 x 96 DPI)

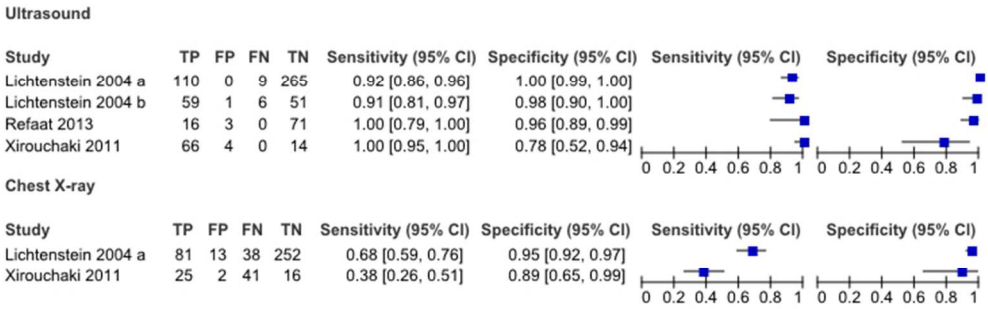


Figure 4. Sensitivity and specificity of ultrasound and chest X-ray for included studies. Note different units of analyses which preclude pooling of studies: lung regions (12 per patient) in Lichtenstein 2004a; lungs (2 per patient) in Lichtenstein 2004b and Xirouchaki 2011; and individual patients in Refaat 2013.
196x73mm (96 x 96 DPI)

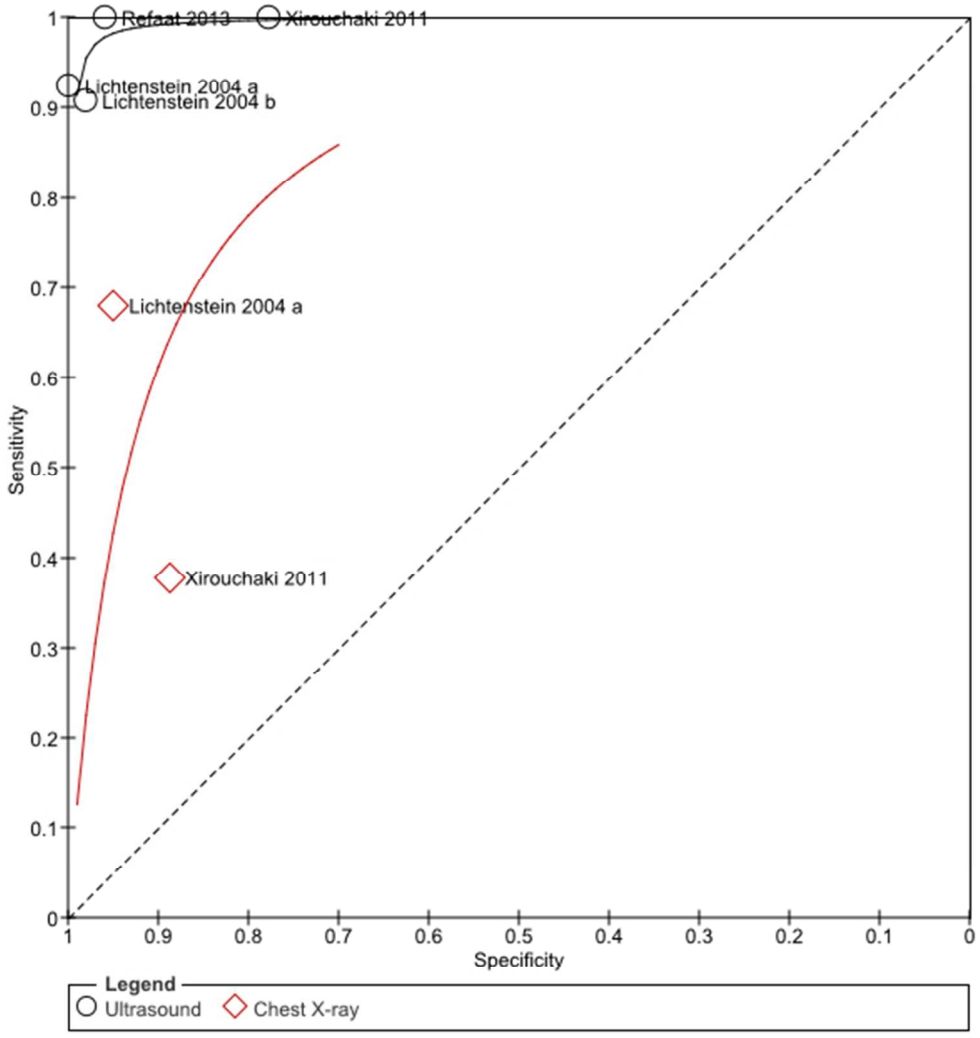


Figure 5. Sensitivity and specificity of ultrasound and chest X-ray for consolidation.
158x171mm (96 x 96 DPI)

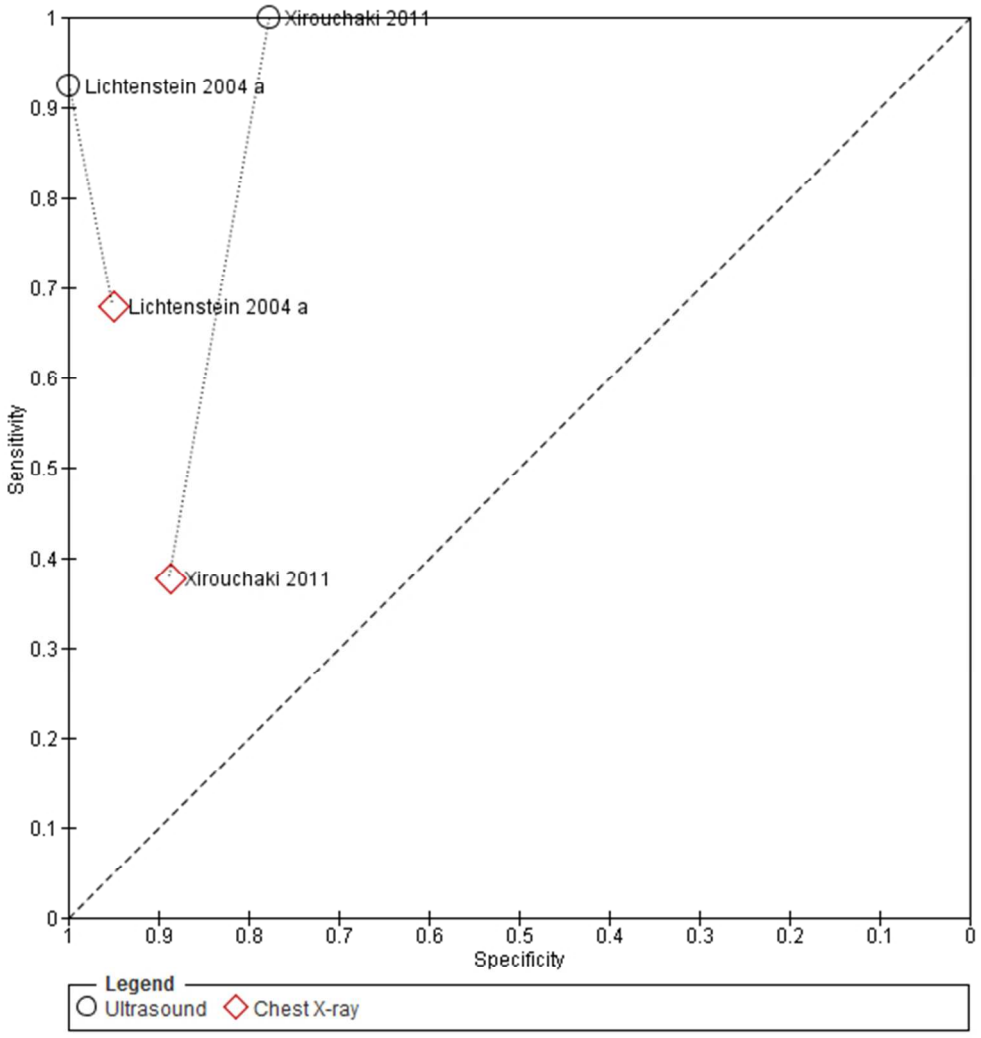


Figure 6. Direct comparisons for ultrasound and chest X-ray in two individual studies. Units of analysis are lungs (2 per patient) for Xirouchaki 2011, and lung regions for Lichtenstein 2004a (12 per patient).
158x171mm (96 x 96 DPI)

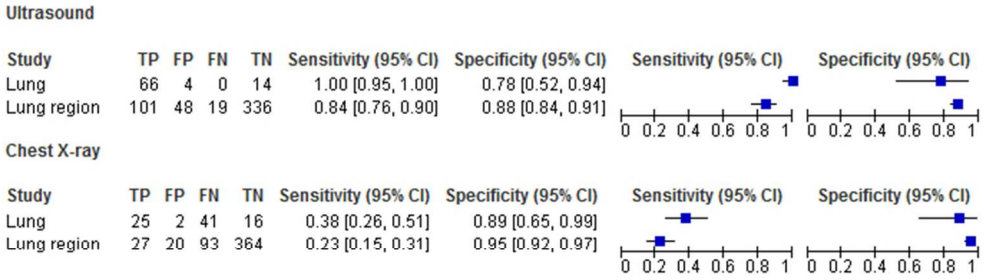


Figure 7. Impact of units of analyses on sensitivity and specificity of ultrasound and chest X-ray for consolidation. Data from Xirouchaki et al 2011 are stratified according to two units of analysis; 'lung' (2/patient) and 'lung region' (12/patient). In comparison to lung analysis, lung region analysis reduces sensitivity but inflates specificity. It also increases total study numbers, giving the appearance of tighter estimates of precision.
184x64mm (96 x 96 DPI)

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	6

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	13
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

ONLINE SUPPLEMENT- Appendices 1-3

The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure: A Systematic Review.

Appendix 1. Specific search strategies for three databases.

MEDLINE

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

	Searches	Results	Search Type
1	exp Ultrasonography/	247016	Advanced
2	(ultrasonic adj3 (diagnos* or tomograph* or imaging*)).ti,ab.	3772	Advanced
3	(echotomograph* or echograph* or ultrasonograph* or sonograph* or ultrasound).ti,ab.	247848	Advanced
4	1 or 2 or 3	395876	Advanced
5	exp Pneumonia/	75417	Advanced
6	pneumon*.ti,ab.	139061	Advanced
7	bronchopneumon*.ti,ab.	3003	Advanced
8	Respiratory Tract Infections/	31632	Advanced
9	("lower respiratory tract infection*" or "lower respiratory infection*" or LRTI).ti,ab.	5539	Advanced
10	(lung adj3 (inflamm* or infect*)).ti,ab.	16851	Advanced
11	exp Critical Illness/	16852	Advanced
12	5 or 6 or 7 or 8 or 9 or 10 or 11	224498	Advanced
13	exp Lung/	230110	Advanced
14	exp Thorax/	40771	Advanced
15	(lung* or chest* or thora* or respirat* or alveol*).ti,ab.	1023348	Advanced
16	13 or 14 or 15	1117732	Advanced
17	4 and 12 and 16	897	Advanced
18	17	897	Advanced
19	limit 18 to (humans and ("all adult (19 plus years)" or "young adult (19 to 24 years)" or "adult (19 to 44 years)" or "young adult and adult (19-24 and 19-44)" or "middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)"))	431	Advanced
20	Animals/	5486116	Advanced
21	17 not 20	809	Advanced
22	adult.mp. or middle aged.sh. or age:.tw.	6843828	Advanced
23	4 and 12 and 16 and 22	522	Advanced
24	23 not 20	496	Advanced

EMBASE.

Embase 1974 to 2013 October 21

	Searches	Results	Search Type
1	exp echography/	503948	Advanced
2	exp echotomography/	950	Advanced
3	(ultrasonic adj3 (diagnos* or tomograph* or imaging*)).ti,ab.	4613	Advanced
4	(echotomograph* or echograph* or ultrasonograph* or sonograph* or ultrasound).ti,ab.	329397	Advanced
5	1 or 2 or 3 or 4	635012	Advanced
6	exp pneumonia/	188700	Advanced
7	pneumon*.ti,ab.	169236	Advanced
8	bronchopneumon*.ti,ab.	3736	Advanced
9	exp lower respiratory tract infection/di [Diagnosis]	35926	Advanced
10	(lung adj3 (inflamm* or infect*)).ti,ab.	19808	Advanced
11	(lower adj3 ("respiratory tract infection*" or "respiratory infection*")).ti,ab.	6896	Advanced
12	LRTI.ti,ab.	860	Advanced
13	exp critical illness/	21700	Advanced
14	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13	324922	Advanced
15	(lung* or chest* or thora* or respirat* or alveol*).ti,ab.	1246364	Advanced
16	exp Lung/	224539	Advanced
17	exp Thorax/	75787	Advanced
18	15 or 16 or 17	1332427	Advanced
19	5 and 14 and 18	3035	Advanced
20	19	3035	Advanced
21	limit 20 to (human and (adult <18 to 64 years> or aged <65+ years>))	1389	Advanced
22	adult.mp. or middle aged.sh. or age:.tw.	7323441	Advanced
23	animal experiment/	1722105	Advanced
24	19 and 22	1732	Advanced
25	24 not 23	1726	Advanced

Web of Knowledge

Set	Results	
# 10	1,413	#7 AND #4 AND #1 Timespan=All years Search language=English
# 9	764	#8 AND #7 AND #4 AND #1 Timespan=All years Search language=English
# 8	Approximately 8,982,030	Topic=(adult* OR "middle age*" OR aged) OR Title=(adult* OR "middle age*" OR aged) Timespan=All years Search language=English
# 7	Approximately 420,383	#6 OR #5 Timespan=All years Search language=English
# 6	Approximately 73,103	Topic=(("critical* ill*")) OR Title=(("critical* ill*")) Timespan=All years Search language=English
# 5	Approximately 352,396	Topic=(pneumon* OR bronchopneumon* OR bronchit*) OR Title=(pneumon* OR bronchopneumon* OR bronchit*) Timespan=All years Search language=English
# 4	Approximately 645,870	#3 OR #2 Timespan=All years Search language=English
# 3	Approximately 192,826	Topic=(((ultrasonic OR ultrasound) SAME (diagno* OR tomograph* OR imaging*))) OR Title=(((ultrasonic OR ultrasound) SAME (diagno* OR tomograph* OR imaging*))) Timespan=All years Search language=English
# 2	Approximately 634,720	Topic=(echotomograph* OR echograph* OR ultrasonograph* OR sonograph* OR ultrasound) OR Title=(echotomograph* OR echograph* OR ultrasonograph* OR sonograph* OR ultrasound) Timespan=All years Search language=English
# 1	Approximately 2,257,730	Topic=(lung* or chest* or thora* or respirat* or alveol*) OR Title=(lung* or chest* or thora* or respirat* or alveol*) Timespan=All years

Appendix 2. Data extraction form

STUDY IDENTIFIERS

Author	Year	Journal	Country

STUDY TYPE Cohort/Case control/RCT

INCLUSION CRITERIA

1, Timing	Within 24 hours (can still include if no)	
2. Index Test	Chest ultrasound	
3. Target condition	Consolidation	
4. Comparator	Chest CR	
5. Reference	Chest CT	

EXCLUSION

1. Not primary study		
2. Not in respiratory failure		
3. Not chest ultrasound		
4. Not to diagnose consolidation		
5. Unable to obtain translation		
6. Unable to extract 2 X2 data		
7. Unable to obtain paper		

PARTICIPANT DETAILS

1. Dates recruited		
2. Number		
3. Age		
4. Gender (M:F)		
5. Location		
6. Illness severity		
7. ? mechanical ventilator		

INDEX DETAILS

1. Sonographer		
2. equipment		
3. Extent of examination		

COMPARATOR- Erect or supine film?

REFERENCE – CT equipment, protocol, reader.

FLOW- interval between performance of ultrasound and CXR/CT

DIAGNOSTIC PERFORMANCE (and unit of analysis ….)

	CT +	CT -	Sens	Spec
US +				
US -				
CXR +				
CXR -				

Appendix 3. Quality Assessment Forms

Domain		Signalling Question/Checklist	*Rating Guidelines*
Patient Population	**Bias**	Consecutive/random sample?	*Rate 'yes' if stated consecutive, or able to ensure randomness. Rate 'unclear' if neither, or 'convenience,' or data implausible. Rate 'no' if stated non-consecutive/non-random, or inclusion based on clinical decision for CT.*
		Case/control avoided?	*Rate 'no' if two-gate case control design*
		Inappropriate exclusions avoided?	*Rate 'no' if exclusions would not have affected ability to undergo testing, or post-hoc Rate 'unclear' if ambiguous wording raises possibility of inappropriate exclusions*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'
	Applicability	Respiratory failure/respiratory support/intensive-care	*Rate 'low concern' if indicates respiratory failure or requirement for respiratory support including high flow O_2 or ventilation.*
			Rate 'unclear concern' if indication for intensive care or intubation not clearly respiratory failure
			Rate 'high concern' if some patients clearly did not have respiratory failure, or confined to a subgroup of respiratory failure aetiologies
Ultrasound (Index)	**Bias**	Sonographer blind to CT AND clinical data?	*Rate as stated, 'unclear' if not stated.*
	Applicability	Reasonable sonographer, scanner and protocol	*Rate 'high concern' if deviates from usual practice to extent of irreproducibility, very old equipment, or incomplete scanning protocol*
CT (Reference)	**Bias**	Is CT likely to correctly classify consolidation?	*Rate 'unclear' if very old scanner, 'no' if examination incomplete*
		CT reported blind to ultrasound AND	*Rate as stated, 'unclear' if not stated.*

		clinical data?	*Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if both responses 'yes'*
	Applicability	Reasonable scanner and protocol	*Rate 'high concern' if deviates from usual practice to extent of irreproducibility.*
Flow & Timing	**Bias**	Interval between ultrasound and CT appropriate?	*Rate 'unclear' if not stated.*
		All patients underwent CT?	*Rate 'unclear' if not stated.*
		All patients analysed?	*Rate 'unclear' if not stated.*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'

(Table 1 continued)

Protocol for a systematic review & meta-analysis: Chest ultrasound versus chest X-ray for diagnosing radiographic consolidation in patients with acute respiratory failure.

Abstract.

Introduction.

Diagnosing pneumonia as a cause of acute respiratory failure can be challenging. Standard chest X-rays have limited accuracy.

Objectives.

This systematic review and meta-analysis will estimate the diagnostic accuracy of chest ultrasound as an initial test for radiographic consolidation in acute respiratory failure, and compare it against that of chest X-ray.

Search Strategy.

Medline, EMBASE, and the Science Citation Index will be searched. Reference lists, citing alerts, and related articles will be examined. If required, authors will be contacted for additional information.

Study Selection.

Studies: randomised trials, cohort and nested (one-gate) case-control studies. *Timing:* initial testing. *Participants:* adults in acute respiratory failure. *Index:* chest ultrasound. *Comparator:* chest X-ray. *Target Condition:* radiographic consolidation. *Reference:* chest computer tomography.

Data Collection & Analysis.

Two reviewers will independently select studies for inclusion, assess study quality and extract data. The sensitivity and specificity of ultrasound for consolidation in acute respiratory failure will be determined and compared against that of chest X-ray. Summary point estimates and 95% confidence intervals for ultrasound and chest X-ray will be determined by bivariate and hierarchical models. Heterogeneity will be explored by subgroup analyses and meta-regression.

Interpretation.

Results will inform policy-makers and clinicians regarding benefits of introducing chest ultrasound for pneumonia diagnosis in acute respiratory failure, and identify patients who may benefit most.

A. Rationale.

Acute respiratory failure.

Acute respiratory failure (a low blood-oxygen level) is a life-threatening state which requires urgent admission to hospital, often to the intensive care unit. Immediate provision of supplemental oxygen is critical. The next priority is then to diagnose (and treat) the underlying cause of respiratory failure.

Pneumonia can cause respiratory failure.

One major cause of respiratory failure is pneumonia, an infection of the lung parenchyma usually caused by bacterial or viral pathogens, many of which are susceptible to antimicrobial therapy.

Diagnostic tests for pneumonia in respiratory failure.

The history and physical examination may point to the presence of pneumonia, but its *sine qua non* is the finding of lung shadowing on radiological imaging, otherwise known as 'consolidation'.

However, traditional imaging techniques have significant drawbacks in detecting 'consolidation' in this setting;

a) Bedside chest X-ray may fail to detect consolidation due to the suboptimal images obtained in acutely unwell patients (Ovenfors et al, 1978).
b) Chest computer tomography (CT) has greater accuracy (Mirvis et al, 1987) but involves the risk of transporting patients who require respiratory support away from the safety of their ward.

The advantages of chest ultrasound.

Chest ultrasound may overcome the drawbacks of traditional imaging for pneumonia (Figure 1);

a) Unlike X-ray, ultrasound does not depend on optimal positioning.
b) Unlike CT scanning, ultrasound can be brought to the patient's bedside.

Figure 1. Ultrasound machinery used for chest sonography.

Current uncertainty.

Narrative reviews (including our own; Hew & Heinze, 2012) and consensus guidelines (Volpicelli et al, 2012) have identified evidence for the use of ultrasound to diagnose pneumonia.

However, a systematic review of chest of ultrasound to diagnose pneumonia in acute respiratory failure is needed to summarise evidence on its diagnostic accuracy. A test should ideally not enter clinical practice until its diagnostic accuracy has been clearly defined. Without such data, the introduction of the test may cause errors in diagnostic reasoning and jeopardise patient safety.

It is also necessary to compare a new test against tests currently in clinical use for that purpose, in order to determine the best role (if any) for the new test.

B. REVIEW QUESTION & OBJECTIVES.

'In the initial testing of patients with acute respiratory failure,
what is the accuracy of chest ultrasound to diagnose CT-detected consolidation,
when compared to bedside chest X-ray?'

This systematic review will therefore:

a) Review evidence on the ***diagnostic test accuracy*** of chest ultrasound for radiological consolidation in patients with acute respiratory failure.
b) Perform a ***comparison*** of chest ultrasound diagnostic accuracy with that of chest X-ray.

C. METHODS- DATA COLLECTION.

Inclusion Criteria.

Inclusion criteria for studies in this systematic review have been developed in accordance with recommendations from the Cochrane Handbook of Diagnostic Test Accuracy (Bossuyt et al, 2008).

Studies.

Cohort studies (where all patients have acute respiratory failure) and nested one-gate case-control studies (comprising patients with and without radiographic consolidation) will be included. Randomised controlled trials allocating respiratory failure patients to either ultrasound or a standard comparator will be included.

Studies which evaluate only chest ultrasound will be included.

Studies which evaluate chest ultrasound against standard testing with chest X-ray will also be included. The index comparisons may be either *paired* or *randomised*.

Timing.

The best place for ultrasound along the diagnostic pathway must be specified, since altering its place in a sequence of tests may change its diagnostic accuracy (Leeflang, et al, 2008; Reitsma et al, 2012). Given its advantages over chest X-ray and CT, the best role for ultrasound is as an initial test. Studies

will thus be included if < 24 hours elapse between acute respiratory failure diagnosis and the chest ultrasound.

If insufficient studies with this design are located, studies which employ ultrasound r> 24 hours after diagnosis of acute respiratory failure will be included, and explored for their likely impact on statistical heterogeneity.

Participants.

Studies of adult patients (age > 18) admitted to hospital with acute respiratory failure will be included. Patients admitted to emergency wards, general wards, high-dependency and intensive-care units will be included.

Studies where patients are well enough to be discharged home directly from the emergency department within 24 hours of presentation will be excluded.

Acute respiratory failure will be defined as one of the following:

a) An arterial partial pressure of oxygen (PaO_2) < 60 millimetre of mercury (mm Hg), without supplemental oxygen.
b) An arterial oxygen saturation of < 90% measured by pulse oximetry, without supplemental oxygen.
c) Where supplemental oxygen is required in order to raise the PaO2 to > 60 mm Hg or arterial oxygen saturations to > 90%.
d) Studies which do not explicitly define respiratory failure but make reference to credible diagnostic conventions based on the principles above will also be considered for inclusion.

Index Test.

Studies which use chest ultrasound will be included. Chest ultrasound will be defined as the use of B (brightness)-mode ultrasound to systematically examine the lungs and pleura. The investigation may be performed by either clinicians or radiologists.

Comparator.

Chest X-ray is universally performed as the initial investigation for respiratory failure. Thus if studies evaluate chest X-ray against chest ultrasound, they will be included for direct comparison.

Studies evaluating only chest X-ray (without chest ultrasound) will be excluded from review. Given the potential for clinical heterogeneity between studies, indirect comparisons will not be performed.

Target condition.

As mentioned, ultrasound, chest X-ray and CT have characteristic imaging findings suggestive of pneumonia, *i.e.* 'consolidation'. The target condition will be framed as this radiological finding ('consolidation') rather than a clinical diagnosis ('pneumonia'). This allows comparison of like with like; images (on ultrasound) referenced to images (on radiology).

Studies will therefore be included if they measure the accuracy of chest ultrasound for the *radiological finding* of consolidation defined by the reference standard.

Studies which examine the accuracy of chest ultrasound to detect the *clinical diagnosis* of pneumonia (defined by a combination of history, examination and imaging) will be excluded, in order to avoid bias arising from integrating non-imaging data into the diagnostic algorithm.

Reference standard.

In patients with acute respiratory failure, X-rays obtained at the bedside result in poor image quality (Ovenfors et al, 1978) and poor diagnostic sensitivity (Henschke et al, 1996). Conversely, CT is highly sensitive for consolidation even in very ill patients (Rubinowitz et al, 2007). Studies will therefore only be included if they use chest CT scanning as the reference standard.

Studies may still be included if only some patients receive CT (differential verification), but only if data can be obtained for analysis in the subgroup of patients who underwent CT.

Chest CT will be defined as the use of helical CT to examine the thorax.

Search Strategy.

The search will be conducted according to guidance from the Cochrane Handbook of Diagnostic Test Accuracy (de Vet et al, 2008) and include the following principles:

a) The search strategy will be developed in consultation with a healthcare librarian experienced with supporting systematic reviews.
b) The search will be carried out independently by two reviewers. Disagreements will be resolved by discussion or consultation with a third reviewer.
c) Multiple electronic databases will be searched including, but not confined to, MEDLINE, EMBASE and the ISI Science Citation Index.
d) The search strategy will include the concepts: (i) index test *i.e.* ultrasound AND target condition *i.e.* pneumonic consolidation, or, (ii) index test *i.e.* ultrasound AND participants *i.e.* acute respiratory failure.
e) Each concept will be described by a large variety of terms (text words and subject headings).
f) For each database, a number of preliminary searches will be conducted using a range of textwords and subjects headings. Further searches will then be run using additional textwords and subjects headings derived from studies identified on initial searches.
g) In order to maximise search sensitivity, relevant search filters will NOT be used (such as the new EMBASE indexing term 'diagnostic accuracy study' or the MEDLINE subject heading 'sensitivity and specificity').
h) The reference lists of relevant studies will be examined and citation searches will be performed. Citing alerts in electronic journals and the 'related articles' feature in PubMed will be used to identify further relevant articles.
i) The search strategy will be fully described as an appendix in the final published review.

Study Identification.

Search results will be screened by two independent reviewers. Studies that appear relevant will be obtained and assessed for inclusion by each reviewer. Disagreements will be resolved by discussion or referral to a third reviewer. The process of study identification will be shown by a flow diagram.

Data Extraction.

Data extraction forms will be developed using Microsoft Access. The forms will be trialled on a small number of studies and modified appropriately.

Data will be extracted independently by two reviewers. Disagreements will be resolved by discussion or referral to a third reviewer.

Data items to be extracted will include:

a) *Study identification*- author, year, location.
b) *Study details*- cohort, case-control, randomised trial.
c) *Inclusion*- Timing, index test, target condition, comparator, reference.
d) *Exclusion*- The following hierarchy will be employed: (i) Not a primary study (ii) Patients not in respiratory failure (iii) Not a study of chest ultrasound (iv) Not a study to diagnose consolidation (v) Unable to obtain a translation (vi) Unable to extract 2 X 2 data (vii) Unable to obtain the paper.
e) *Participants*- number, age, gender, location, illness severity, whether on a ventilator.
f) *Index*- sonographer (clinician versus radiologist), equipment (high-end versus lightweight portable), thoroughness of sonographic examination (whether any views were excluded).
g) *Comparator and reference*- interval between performance of ultrasound and chest X-ray/CT scanning.
h) *Diagnostic performance*- 2 X 2 contingency tables of index and (where available) comparator tests denoting true positives, true negatives, false positives, false negatives.

Quality Assessment.

QUADAS2 (the Revised Tool for Quality Assessment of Diagnostic Accuracy Studies; Whiting et al, 2011) will be used to assess quality in this review. As per QUADAS2 the quality assessment process will be divided into 4 phases;

Articulating the review question.

Participants, index, reference, and flow and timing are all already defined in the review question earlier.

Tailoring the tool to the specific review.

Four steps are suggested (Whiting et al, 2011);

a) Tailoring the tool content. (Table 1).
b) Developing rating guidelines. (Table 1).
c) Piloting the tool and guidelines. This will be done on a randomly-selected included study.
d) Applying the tool to all included studies. This will be done on all included studies.

Drawing flow diagrams for each study.

A flow diagram for each included study will assist judgments regarding bias and applicability.

Applying the tool to each study to make judgments on bias and applicability.

Finally, the tool will be applied to each included study. The risk of bias and level of applicability for each study will be summarised in tables (Table 2) and graphs (Figure 2). [Summary scores for individual studies will not be undertaken as they are prone to error (Juni et al, 1999).]

Table 1. Rating guidelines (blue) developed for the review-specific quality assessment (black).

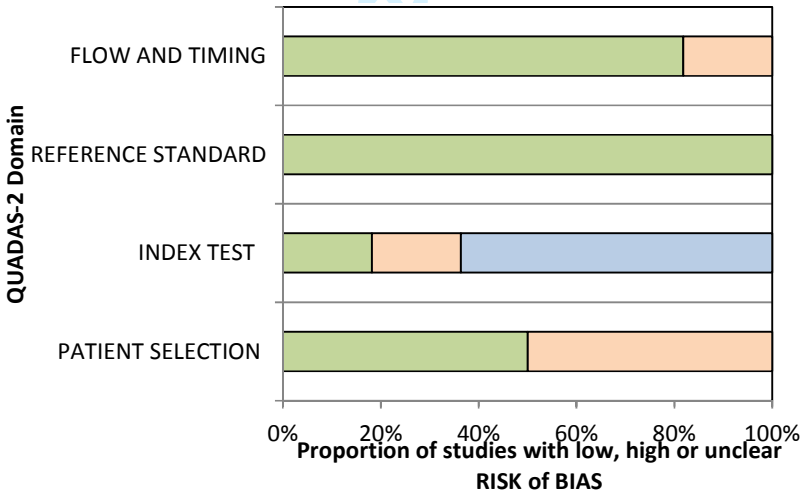
Domain		Signalling Question/Checklist	***Rating Guidelines***
Patient Population	**Bias**	Consecutive/random sample?	*Rate 'yes' if stated consecutive, or able to ensure randomness. Rate 'unclear' if neither, or 'convenience,' or data implausible. Rate 'no' if stated non-consecutive/non-random or inclusion based on clinical decision for CT.*
		Case/control avoided?	*Rate 'no' if two-gate case control design*
		Inappropriate exclusions avoided?	*Rate 'no' if exclusions would not have affected ability to undergo testing, or post-hoc Rate 'unclear' if ambiguous wording raises possibility of inappropriate exclusions*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'
	Applicability	Respiratory failure/respiratory support/intensive-care	*Rate 'low risk' if indicates respiratory failure or requirement for respiratory support including high flow O_2 or ventilation.*
			Rate 'unclear' if indication for intensive care or intubation not clearly respiratory failure
			Rate 'high risk' if some patients clearly did not have respiratory failure, or confined to a subgroup of respiratory failure aetiologies
Ultrasound (Index)	**Bias**	Sonographer blind to CT AND clinical data?	*Rate as stated, 'unclear' if not stated.*
	Applicability	Reasonable sonographer, scanner and protocol	*Rate 'high risk' if deviates from usual practice to extent of irreproducibility, very old equipment, or incomplete*

			scanning protocol
CT (Reference)	**Bias**	Is CT likely to correctly classify consolidation? CT reported blind to ultrasound?	*Rate 'unclear' if very old scanner, 'no' if examination incomplete* *Rate as stated, 'unclear' if not stated.* *Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if both responses 'yes'*
	Applicability	Reasonable scanner and protocol	*Rate 'no' if deviates from usual practice to extent of irreproducibility.*
Flow & Timing	**Bias**	Interval between diagnosis of respiratory failure and ultrasound < 24 hours? Interval between ultrasound and CT appropriate? All patients underwent CT? All patients analysed?	*Rate 'no' if > 24 hours* *Rate 'no' if > 24 hours.* *Rate 'unclear' if not stated.* *Rate 'unclear' if not stated.* *Rate 'unclear' if not stated.* *Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'*

Table 2. Suggested tabular presentation for quality assessment (adapted from Whiting et al, 2011).

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Study 1	?	☹	☺	☹	☺	☺	☺
Study 2							
Study 3							
etc.							

Figure 2. Graphical display of QUADAS2 quality assessment for risk of bias (from Whiting et al, 2011). A similar display will be generated for applicability concerns (not shown).



D. METHODS- DATA ANALYSIS.

Measures of test accuracy.

Primary outcomes.

a) *Diagnostic accuracy of Ultrasound.*
For each included study, a 2 X 2 contingency table will be extracted to allow calculation of sensitivity and specificity of ultrasound for consolidation with 95 % confidence intervals.

b) *Direct Comparison of Ultrasound versus chest X-ray.*
In studies where a direct comparison between ultrasound and chest X-ray is performed, sensitivity and specificity of chest X-ray for consolidation with 95% confidence intervals will also be derived.

Secondary outcomes.

Positive and negative likelihood ratios and diagnostic odds ratios will be calculated for each study. Positive and negative predictive values will also be presented, referenced specifically to the *prevalence* of pneumonia in the studies identified.

Missing data.

For studies where 2 X 2 contingency tables cannot be derived from the paper, corresponding authors will be contacted. If adequate data is still not obtained, these studies will be excluded from review and analysis.

Units of analysis.

Per patient.

It is anticipated that the unit of analysis in most studies will be individual patients.

Per lung.

It is possible that some studies, in an effort to increase sample size, may report ultrasound results separately for each lung.

Importantly, this latter method is prone to bias, since the two lungs in a patient with (or without) consolidation are not independent of each other. Furthermore, both sides of a patient would be scanned by a single sonographer, whose examination of the second lung is likely to be biased by findings from the first.

For this reason, the influence of the unit of analysis on statistical heterogeneity will be explored.

Descriptive statistics.

Paired Forest plots.

Paired Forest plots will be used to display 95% confidence intervals of sensitivity and specificity for each study.

Results will be stratified by study type, since this may influence the potential biases (*eg* spectrum bias in case-control designs, versus partial verification bias in cohort-type studies).

Receiver operator characteristics (ROC) plots.

Pairs of sensitivity and specificity from each study will be plotted in receiver operator characteristics (ROC) space. This will facilitate an assessment of whether sensitivity and specificity are negatively correlated (Reitsma et al, 2012).

Meta-analysis.

Feasibility.

The appropriateness of data pooling and meta-analysis will depend on the number of studies and participants, and the methodological and clinical homogeneity of included studies.

Methodology.

Meta-analysis if appropriate will be conducted with expert statistical assistance, given the complexity of this field and the rapid evolution in methodology (Reitsma et al, 2012).

The Moses-Littenberg model provided in Revman 5 (Moses t al, 1993) will not be used for meta-analysis, since it does not allow for random effects, nor does it provide estimates of heterogeneity between studies (Macaskill et al, 2010).

Instead, the Bivariate random effects (Reitsma et al, 2005) models will be employed to determine summary estimates of test accuracy and calculate reliable 95% confidence intervals around these parameters (Harbord et al, 2007). We are particularly interested in the summary estimates for ultrasound rather than analysis of the SROC itself.

Results from these models will be input into Revman 5 to depict;

a) A summary ROC plot for ultrasound,
b) A summary operating point for ultrasound, and,
c) A 95% confidence region around the summary point for ultrasound (Macaskill et al, 2010).

If bivariate random effects models do not converge, a univariate logistic regression random effects meta-analysis of sensitivity and specificity separately will be performed instead.

Heterogeneity.

Subgroup analysis- Describing heterogeneity.

The number of subgroup analyses will be kept to as low as possible to minimise chance findings.

However, the following pre-specified subgroups will be examined, based on;

a) The unit of analysis (person versus lung),
b) Need for mechanical ventilation (yes or no),
c) Location of patient management (general ward versus intensive care),
d) Ultrasound operator (clinician versus radiologist),
e) Sophistication of ultrasound equipment (high-end machines versus small portable devices).

Meta-regression- Explaining heterogeneity.

Depending on the total number of studies, the number of participants within studies, and the degree of heterogeneity, the influence of covariates on diagnostic accuracy may be further explored by meta-regression, available through both Bivariate and HSROC models (Reitsma et al, 2012).

Covariates potentially contributing to heterogeneity will be examined using descriptive ROC plots according to covariates. If there are sufficient studies, meta-regression will be considered. Such covariates include;

a) ***Clinical Heterogeneity*;** unit of analysis, need for mechanical ventilation, location of patient management, sonographer type, documented thoroughness of sonographic chest examination, and the sophistication of ultrasound equipment.
b) ***Methodological Heterogeneity***. Depending on the quality of the studies included, specific risks of bias (such as partial verification bias) may also be incorporated as covariates.

However, it is acknowledged that such study level covariates have limited power to detect differences in diagnostic accuracy between subgroups (Reitsma et al, 2012). Furthermore, it has been suggested that at least 10 studies per covariate are needed for robust meta-regression (Gagnier et al, 2012), and it is unlikely that a sufficient volume of studies will be included.

Direct comparison of ultrasound and chest X-ray.

Only studies undertaking both tests will be included for this comparative analysis; each patient may have either undergone both tests (for paired comparison) or be randomised to either test (for randomised comparison).

Even if insufficient data is available for meaningful direct comparison, indirect comparison will *not* be performed. It is likely that studies examining the diagnostic accuracy of chest X-ray may be different from those examining the accuracy of chest ultrasound, introducing significant bias to indirect comparisons.

Preliminary graphical analysis.

The sensitivity and specificity for both tests (ultrasound and chest X-ray) in each study will be plotted in ROC space using Revman 5, as single points joined by a line (Macaskill et al, 2010).

Test comparisons.

Depending on the data available from individual studies, tests comparisons may then be performed using the bivariate model, with outputs which may be entered into Revman to superimpose the summary estimates for each tests (ultrasound and chest X-ray) and their 95% confidence regions on the ROC scatterplot (Macaskill, 2010).

Sensitivity analyses.

The following sensitivity analyses will be performed in order to test the robustness of the primary outcomes.

a) Comparison of the summary operating point for ultrasound, with and without inclusion of studies with a high risk-of-bias in one or more domains on quality assessment.
b) Comparison of the summary operating point for ultrasound, with and without inclusion of studies with a single lung as the unit of analysis.

c) Comparison of the sensitivity and specificity of ultrasound versus chest X-ray, with and without inclusion of studies with a high risk of bias in one or more domains on quality assessment.

Software.

A number of software options will be considered;

a) Analyses will probably be performed with the latest version of Stata, using the *metandi* command for meta-analysis of test accuracy studies (Harbord, 2009). Stata or SAS codes for bivariate analysis including covariates will be used. Results would be input into Revman 5 for graphical display.
b) Alternatively, the freely available R-package *mada* command (Doebler 2012) may be used, which performs all the analyses described above including bivariate and HSROC models with covariates, and provides publication-ready figures as an integral part of the programme.
c) By the analysis stage of this systematic review, new statistical techniques and software options may have emerged that supersede the software options described above.

Thus the final decision regarding software selection will be made in consultation with a specialist biomedical statistician abreast of advances in the field, *just prior to performing the data analysis.*

E. Interpretation

Methodological conclusions.

Results of this review and meta-analysis will be of value in determining whether and how to apply ultrasound as an initial test. Its exact role will hinge on the diagnostic test performance established in the review and meta-analysis (Bossuyt et al, 2006);

a) If ultrasound is highly specific ***or*** sensitive for consolidation, it could serve to rule-in or rule-out the condition, terminating the diagnostic pathway as a *triage test*.
b) If ultrasound is more sensitive ***and*** specific than chest X-ray to diagnose consolidation, it could serve as a *replacement test*.
c) If ultrasound is more sensitive ***but*** less specific than chest X-ray to diagnose consolidation, it could serve as an *add-on test* if the initial chest X-ray is negative.

Clinical conclusions.

The actual values of sensitivity and specificity are of critical importance. Patients with consolidation detected on ultrasound are likely to be given empirical treatment for pneumonia. Patients without consolidation on ultrasound are likely to undergo testing for alternative diagnoses.

The consequences of a false negative result are therefore significant. If consolidation is missed, a potentially treatable condition (ie pneumonia) may go untreated. However, the consequences of a false positive result are equally significant. Reporting the presence of consolidation when it is

actually absent may mislead the clinician into premature diagnostic closure, with subsequent failure to consider alternative diagnoses, including the true diagnosis.

Subgroup analyses and meta-regression may identify important patient characteristics or test practices which influence the diagnostic accuracy of ultrasound in this scenario.

Finally, practical considerations (cost and availability) also have a bearing on the choice of testing.

F. Dissemination.

The results will be published in a critical care or respiratory medicine peer-reviewed journal preferably with an open-access option to allow wide dissemination.

The published document will be reported according to PRISMA guidelines (Preferred Reporting Items for systematic reviews and Meta-Analyses, Moher et al, 2009).

G. Logistics.

Registration.

To reduce the chance of duplication or redundancy, this Review Protocol will be registered at PROSPERO, an international registry of systematic reviews (Booth et al, 2012).

Review team.

a) Reviewer 1. Mark Hew, Respiratory Physician, Alfred Hospital, Melbourne, Australia.
b) Reviewer 2. John Corcoran, Clinical Research Fellow, Churchill Hospital, Headington, UK.
c) Reviewer 3. Najib Rahman, Director of Oxford Respiratory Trials Unit, Churchill Hospital, UK.
d) Biomedical Statistician. TBA.
e) Health Care Librarian with Searching expertise. TBA.

Timeline (Table 3).

Protocol development	month 1-2	June-July 2013
Literature search	month 3	August 2013
Relevance screening/inclusion assessment	month 4	September 2013
Data extraction & quality assessment	month 5	October 2013
Systematic review & meta-analysis	month 6	November 2013
Submission for publication	month 7	December 2013

H. REFERENCES.

Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012 Feb 9;1:2. doi: 10.1186/2046-4053-1-2.

Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089–92.

Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

Doebler P, Holling H. Meta-Analysis of Diagnostic Accuracy with mada. R-Project 2012. Accessed 19 June 2013 at http://cran.r-project.org/web/packages/mada/vignettes/mada.pdf

Gagnier JJ, Bombardier C, Boon H, Moher D, Beyene J: An empirical study using permutation-based resampling in meta-regression. Systematic Reviews 2012, 1:18.

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007; 8: 239-251.

Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata Journal 2009; 9: 211-229.

Henschke CI, Yankelevitz DF, Wand A, Davis SD, Shiau M. Accuracy and efficacy of chest radiography in the intensive care unit. Radiol Clin North Am. 1996;34(1):21-31.

Hew M, Heinze S. Chest Ultrasound in Practice: a review of utility in the clinical setting. Intern Med J. 2012 Aug;42(8):856-65.

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008 Dec 16;149(12):889-97.

Linnet K, Bossuyt PMM, Moons KGM, Reitsma JB. Quantifying the Accuracy of a Diagnostic Test or Marker. Clinical Chemistry 2012. 58:9: 1292–1301.

Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. J Clin Epidemiol 2004; 57: 925-932.

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Mirvis SE, Tobin KD, Kostrubiak I, Belzberg H. Thoracic CT in detecting occult disease in critically ill patients. AJR Am J Roentgenol. 1987;148(4):685-689.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097.

Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993; 12: 1293-1316.

Ovenfors C, Hedgecock MW. Intensive care unit radiology: problems of interpretation. Radiol Clin North Am 1978:16:407-439

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982-990.

Reitsma JB, Moons KGM, Bossuyt PM, Linnet K. Systematic Reviews of Studies Quantifying the Accuracy of Diagnostic Tests and Markers. Clin Chem. 2012 Nov;58(11):1534-45.

Rubinowitz AN, Siegel MD, Tocino I. Thoracic imaging in the ICU. Crit Care Clin. 2007 Jul;23(3):539-73.

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001; 20: 2865-2884.

Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein DA, Mathis G, Kirkpatrick AW, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med. 2012 Apr;38(4):577-91.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.

BMJ Open

The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure: A Systematic Review.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-007838.R1
Article Type:	Research
Date Submitted by the Author:	17-Apr-2015
Complete List of Authors:	Hew, Mark; Alfred Hospital, Allergy, Immunology & Respiratory Medicine Corcoran, John; Oxford University Hospitals NHS Trust, Oxford Centre for Respiratory Medicine Harriss, Elinor; University of Oxford, Bodleian Health Care Libraries Rahman, Najib; Oxford University, Oxford Centre for Respiratory Medicine Mallett, Susan; National Institute for Health Research Diagnostic Evidence Co-operative Oxford,
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice, Radiology and imaging, Intensive care, Respiratory medicine, Research methods
Keywords:	Adult thoracic medicine < THORACIC MEDICINE, Respiratory infections < THORACIC MEDICINE, Ultrasound < RADIOLOGY & IMAGING, Adult intensive & critical care < ANAESTHETICS, Thoracic medicine < INTERNAL MEDICINE, Chest imaging < RADIOLOGY & IMAGING

The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure: A Systematic Review.

Authors

1. Mark Hew PhD MSc FRACP
Allergy Immunology and Respiratory Medicine, Alfred Hospital, Melbourne, Victoria, Australia

2. John P Corcoran MA MRCP
Oxford Centre for Respiratory Medicine, Oxford University Hospitals NHS Trust, Oxford, UK, Oxford, United Kingdom

3. Elinor Harriss, MA MSc
Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom

4. Najib M Rahman DPhil MSc MRCP
Oxford Respiratory Trials Unit, University of Oxford, Oxford, United Kingdom

5. Susan Mallett BA DPhil
Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author

A/Prof Mark Hew, Head of Allergy, Asthma & Clinical Immunology Service, Alfred Hospital, 55 Commercial Road, Prahran, Victoria 3004 Australia.

m.hew@alfred.org.au Tel 613- 90762934 Fax 613- 90762245

Key words

Pneumonia Respiratory Infection ARDS Assisted ventilation

Abstract word count 299, Manuscript word count 2951, References 24, Tables 3, Figures 7

Structured Abstract.

Objectives: (i) Summarise chest ultrasound accuracy to diagnose radiological consolidation, referenced to chest computed tomography (CT) in patients with acute respiratory failure (ARF). (ii) Directly compared ultrasound to chest X-ray.

Setting: Hospitalised patients.

Participants: Studies were eligible if adult participants in respiratory failure underwent chest ultrasound to diagnose consolidation referenced to CT. Exclusion: (i) Not primary study (ii) Not respiratory failure (iii) Not chest ultrasound (iv) Not consolidation (v) Translation unobtainable (vi) Unable to extract data (vii) Unable to obtain paper. Four studies comprising 224 participants met inclusion.

Outcome measures: As planned, paired Forest plots display 95% confidence intervals of sensitivity and specificity for ultrasound and chest X-ray. Sensitivity and specificity from each study are plotted in receiver operator characteristics (ROC) space. Meta-analysis was planned if studies were sufficiently homogeneous and numerous (≥ 4). Although this numerical requirement was met, meta-analysis was prevented by heterogeneous units of analysis between studies.

Results: All studies were in intensive care, with either a high risk of selection bias or high applicability concerns. Studies had unclear or high risk of bias related to use of ultrasound. Only one study clearly performed ultrasound within 24 hours of respiratory failure diagnosis. Ultrasound sensitivity ranged from 0·91 (95% CI 0·81-0·97) to 1·00 (95% CI 0·95-1·00). Specificity ranged from 0·78 (95% CI 0·52-0·94) to 1·00 (0·99-1·00). In two studies, chest X-ray had lower sensitivity than ultrasound, but there were insufficient patients to compare specificity.

Conclusions: Four small studies suggest ultrasound is highly sensitive and specific for consolidation in acute respiratory failure, but high risk of bias and concerns about applicability in all studies may have inflated diagnostic accuracy. Further robustly-designed studies are needed to define the role of ultrasound in this setting.

Registration: http://www.crd.york.ac.uk/PROSPERO/ (CRD42013006472

Article Summary: Strengths & Limitations of this study

Strengths

- Comparison of sonographic consolidation to a reference of radiological consolidation
- Restricted to studies with reliable gold standard of chest CT
- Examination of the influence of units of analysis on diagnostic accuracy reporting

Limitations

- Small number of eligible studies
- Meta-analysis prevented by heterogeneous units of analysis

INTRODUCTION.

Acute respiratory failure (ARF) is common and deadly. Published incidence rates[1,2] suggest approximately 50,000 patients each year in the UK at the severest end of the ARF spectrum may require ventilatory support. A quarter of these have underlying pneumonia[2], and face mortality rates as high as 50%[3].

Mortality escalates further when the cause of ARF is misdiagnosed, which occurs in one in five patients[4] due in part to imaging limitations. Patients are difficult to position for chest X-ray[5] resulting in suboptimal films which may miss consolidation[6], the commonest pattern of pneumonic infiltrate. Conversely, chest computed tomography (CT) is highly sensitive but entails risks of transporting critically ill patients[7]. Both shortcomings of traditional imaging may be overcome by chest ultrasound. Unlike X-ray, ultrasound does not require optimal patient positioning. Unlike CT, ultrasound can be brought to the bedside.

Narrative reviews[8], consensus guidelines[9] and systematic reviews[10,11] all advocate the use of ultrasound to diagnose pneumonia but crucially, most studies of ultrasound accuracy have not examined patients in ARF settings. In those that do, the reference standard is often the final clinical diagnosis, risking incorporation bias if ultrasound itself forms part of that standard.

Further confusion arises when ultrasound accuracy studies use 'pneumonia' as the target condition. The commonest pneumonic infiltrate on imaging is the shadowing termed 'consolidation'. (Less frequently, pneumonia may cause other imaging findings apart from consolidation, and consolidation may occasionally be caused by conditions other than pneumonia). While ultrasound can diagnose consolidation (an *imaging* finding), only clinicians diagnose pneumonia (a *clinical* diagnosis) by expertly blending imaging findings with available clinical information[12]. However, such clinical incorporation bias distorts estimates of ultrasound accuracy. Instead, the most appropriate

target condition for the imaging finding of sonographic consolidation is another imaging finding, in this case radiographic consolidation on chest CT.

To address these key issues, we undertook a systematic review to summarise the accuracy of chest ultrasound to diagnose radiological consolidation, referenced to chest CT, in the specific setting of hospitalised patients with acute respiratory failure (ARF). We also directly compared ultrasound to chest X-ray, the commonest screening test for consolidation in acute respiratory failure. We excluded paediatric studies because children have a different range of aetiologies for ARF[13].

METHODS.

The protocol was registered at http://www.crd.york.ac.uk/PROSPERO/ (review registration number CRD42013006472) and attached as a supplement; key points are summarised here. The review is reported according to PRISMA guidelines (Preferred Reporting Items for systematic reviews and Meta-Analyses[14]).

Inclusion criteria.

Studies: Cohort, cohort with nested one-gate case-control studies (participants with and without consolidation), randomised controlled trials (ultrasound versus chest X-ray). *Timing:* 24 hours or less between acute respiratory failure (ARF) diagnosis and ultrasound scanning. If insufficient such studies were found, studies where ultrasound was performed more than 24 hours after diagnosis of ARF would be included. *Participants:* Adults (age 18 or greater) admitted to any hospital setting with ARF. Studies excluded if patients discharged home directly from the Emergency Department within 24 hours without ward admission. Acute respiratory failure (ARF) defined as: i) arterial partial pressure of oxygen (PaO_2) < 60 millimetres of mercury, without supplemental oxygen, or; ii) arterial oxygen saturation of < 90% on pulse oximetry, without supplemental oxygen, or; iii) supplemental

oxygen required to prevent i) or ii), or; iv) author diagnosis of acute respiratory failure. *Index:* B-mode ultrasound examining lungs and pleura. *Comparator:* Studies comparing chest X-ray to chest ultrasound. Studies evaluating only chest X-ray excluded. *Target condition*: Radiological consolidation. Studies referencing chest ultrasound to a clinical diagnosis of pneumonia excluded. *Reference standard.* Chest CT, defined as helical CT to examine the thorax. Studies could be included if only some patients received CT, but only when data could analysis in the CT subgroup.

Exclusion criteria.

The following hierarchy was employed: (i) Not a primary study (ii) Not respiratory failure (iii) Not chest ultrasound (iv) Not consolidation (v) Translation unobtainable (vi) Unable to extract data to populate 2X2 contingency tables (vii) Unable to obtain paper through both Bodleian (University of Oxford) and British Libraries.

Search.

A healthcare librarian assisted with strategy development. Several iterations were trialled using two reference studies. The full search was run on 22 October 2013, in Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present, Ovid Embase 1974 to 2013 October 21, and Web of Knowledge Science Citation Index Expanded (Appendix 1). An update search was run on 6 August 2014 prior to publication. Filters were not used and no language or date restrictions applied. Only published studies were included. Reference lists, citation searches, citing alerts in electronic journals and the 'related articles' feature in PubMed were also used.

Study Selection.

Titles and abstracts were screened according to inclusion and exclusion by two reviewers independently of each other, and results pooled. Full texts were assessed for inclusion

independently by two reviewers. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Data Collection.

Pre-specified data extraction forms (Appendix 2) were developed [15], trialled in one study and modified. Data items included participants, index, comparator, reference, flow and diagnostic performance. Data was independently extracted by two reviewers. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Quality assessment.

QUADAS2 (the Revised Tool for Quality Assessment of Diagnostic Accuracy Studies) [16] was tailored for this review (Appendix 3). Rating guidelines were developed, piloted in one included study, and applied to remaining studies by both reviewers independently. Differences were resolved by discussion and, if necessary, referral to a third reviewer.

Analysis Plan.

Most studies were expected to analyse patients, although we anticipated beforehand that studies might report results for each lung. We did not anticipate reporting by lung region. We considered analysis of any units other than patients as biased, since one lung (or lung region) of a patient is not independent of another.

Paired Forest plots were used to display 95% confidence intervals of sensitivity and specificity. Sensitivity and specificity from each study were plotted in receiver operator characteristics (ROC) space. Meta-analysis was planned if studies were sufficiently homogeneous and numerous (≥ 4). Although this numerical requirement was met, meta-analysis was prevented by heterogeneous units of analysis between studies. Details of the planned meta-analysis are available at the registered protocol. Where ultrasound was compared to chest X-ray, sensitivity and specificity for both tests

were plotted in ROC space using Revman 5[17]. Tests of unpaired proportions for large samples were used to compare tests within same study, as insufficient data was available for paired comparison.

RESULTS.

Study selection.

Figure 1 shows the PRISMA flowchart. Totals from both (original and update) searches are combined. Four studies met inclusion criteria[18-21]. 2X2 contingency tables could be extracted from all included studies.

We were concerned regarding possible duplication or overlap of cohorts because two included papers had the same first author and year of publication (Lichtenstein, 2004a[18], Lichtenstein, 2004b[19]). However, we found key differences between the studies which rendered this unlikely (Table 1).

Study characteristics.

Table 2 summarises settings and patient characteristics of studies. All four studies were intensive care cohorts, but each reported different severity measures making comparison of acute respiratory failure (ARF) severity difficult.

Table 3 summarises ultrasound methods and units of analyses. Only one study[18] met our criterion for preferred studies, with ultrasound undertaken within 24 hours of ICU admission (and thus probably of ARF diagnosis). In the other studies, timing of ultrasound in relation to ARF diagnosis was not stated. Scanning protocols were similar across all studies; each lung was divided into six regions; anterior, lateral, and posterior; with upper and lower divisions. Three studies employed micro-convex probes, the fourth used both linear and convex probes. Probe frequency ranged between 3.5 and 10 megahertz (MHz).

Risk of bias and applicability concerns.

Figures 2 and 3 summarise the quality assessment of individual primary studies. Applicability was considered in relation to our review question, which examined the diagnostic accuracy of chest ultrasound for CT-detected radiographic consolidation in adults with acute respiratory failure (ARF).

***Lichtenstein, 2004a*[18].**

Selection. There was a low risk of selection bias due to consecutive recruitment of patients with acute respiratory distress syndrome. However, since acute respiratory distress syndrome represents the highest acuity of ARF, concerns regarding applicability were high.

Index Test. The risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information, (although blinded to CT). Applicability concerns were low.

Reference Test. CT interpretation was blinded to ultrasound results, but clinical data may have been available, so risk of bias was unclear. Applicability concerns were low.

Flow & timing. CT was performed within six hours after ultrasound. Risk of bias was low.

***Lichtenstein, 2004b*[19].**

Selection. The risk of selection bias was high because subjects were recruited on the clinical need for CT. Most patients were likely in ARF based on their need for intubation and specific diagnoses; however, ARF was not specifically stated so applicability concerns were unclear.

Index Test. The risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information, (although blinded to CT). Applicability concerns were low.

Reference Test. CT interpretation was blinded to ultrasound results, but clinical data may have been available, so risk of bias was unclear. Applicability concerns were low.

Flow & timing. CT was performed within six hours after ultrasound. The risk of bias was low.

***Xirouchaki, 2011*[20].**

Selection. Risk of selection bias was high as subjects were recruited on their clinical need for CT. Again, most patients were likely in ARF based on their need for intubation and specific diagnoses; however, ARF was not specifically stated so applicability concerns were unclear.

Index Test. Risk of bias was unclear as it was not stated whether sonographers were blinded to clinical information (although blinded to CT). Applicability concerns were low.

Reference Test. CT was interpreted blinded to both clinical information and ultrasound results, giving a low risk of bias. Applicability concerns were low.

Flow & timing. CT was performed no longer than six hours after ultrasound. Risk of bias was low.

***Refaat, 2013*[21].**

Selection. Recruitment was consecutive and exclusions were reasonable, giving a low risk of selection bias. The aetiological diagnoses in this patient group were restricted to a subgroup of respiratory failure aetiologies, causing high applicability concerns.

Index Test. The risk of bias was high as the sonographer had access to clinical information. There were low applicability concerns.

Reference Test. CT was interpreted blind to ultrasound results, giving a low risk of bias. Applicability concerns were low.

Flow & timing. CT was performed within 24 hours of ultrasound; risk of bias was low.

Analysis.

Ultrasound.

Sensitivity for diagnosing CT-detected consolidation ranged from 0·91 (95% CI 0·81-0·97) to 1·00 (95% CI 0·95-1·.00, Figure 4). Specificity ranged from 0·78 (95% CI 0·52-0·94) to 1·.00 (0·99-1·00).

Ultrasound compared to chest X-ray.

Two studies, both of which only included ventilated patients[18,20] (Figures 4, 5 & 6) evaluated both ultrasound and chest X-ray in the same patient populations, the best study design to compare tests[22]. In both studies the sensitivity of ultrasound was significantly greater than that of chest X-ray; 0·24 higher (95% CI 0·15 to 0·34, p<0·.0001; Figures 4, 6) in the first study[18] and 0·62 (95% CI 0·50 to 0·74, p<0·0001) in the second[20] (Figures 4, 6). When compared using 12 lung regions per patient, specificity was higher for ultrasound (0·049, 95% CI 0·023 to 0·075, p=0·0003) in the first study[18], but lower in in the second (-0·049, 95% CI -0·23 to -0·075, p=0·0003)[20]. Specificity compared at 2 lung regions per patient lacked sufficient power to detect a difference (Figure 4).

Impact of unit of analysis.

Three different units of analyses were reported across four studies (Table 3). Only one study used the patient as a unit of analysis[21]. A second study[19] used only the lung (two per patient), and a third[18] used only lung regions (12 per patient).

The fourth study[20] reported its results using the lung as the unit of analysis, but provided additional data using lung regions in an electronic supplement. This provided the opportunity to study the impact of different units of analysis on test characteristics within same dataset (Figure 7).

Importantly, for both ultrasound and chest X-ray, changing the unit of analysis from lung to lung region reduced sensitivity but enhanced specificity, and gave more precise estimates of accuracy (narrower confidence intervals). It also inflated the prevalence of consolidation.

Reported sources of ultrasound and chest X-ray error.

In one study[19], five of six false negative ultrasounds were in patients with posteriorly placed consolidation. This study evaluated patients only in the supine position, which may hinder the detection of posterior consolidation.

In another study[20], all four false positive ultrasounds detected only small areas of consolidation. This study only used a tissue-like pattern to diagnose consolidation which may have reduced specificity.

None of the studies proposed reasons for false positive or false negative chest X-ray results in their discussion.

Synthesis of results.

Meta-analysis was not performed due to heterogeneous units of analysis across studies.

Additional analyses.

Heterogeneity could not be explored due to the small number of studies, apart from comparisons between different units of analysis.

DISCUSSION.

Summary of evidence.

In four small studies, the reported sensitivity and specificity of ultrasound for CT-diagnosed consolidation was high among hospitalised patients with acute respiratory failure (ARF). Ultrasound sensitivity was greater than for chest X-ray, in two studies directly comparing both methods in the same patient populations. However, paired comparisons in individual patients which are the best evidence for comparing tests were not available.

This review identified four quality issues that impact the reported test accuracy of ultrasound in included studies. Firstly, patient selection in every eligible study was either at high risk of bias, or had concerns about applicability to our systematic review. These concerns included recruitment of participants in ICU at the severest acuity of ARF (spectrum bias), restriction to limited ARF aetiologies, and non-consecutive recruitment. The sensitivity of ultrasound for consolidation may

thus be markedly poorer in unselected populations with less severe ARF (and lower burdens of consolidation) or a wider range of ARF aetiologies.

Secondly, in no study were sonographers clearly blinded to clinical data. This is pertinent because sonographers (who in three studies were actually clinicians) could have integrated bedside clinical data with ultrasound evaluation, artificially inflating ultrasound sensitivity.

Thirdly, only one study specified that ultrasound was performed within 24 hours of ICU admission (and presumably, of ARF diagnosis). The more time elapses before ultrasound is performed, the more likely lung consolidation would progress to a detectable extent, but the less likely the test result would improve patient outcome. This would boost reported ultrasound sensitivity but overstate its utility as an initial test.

Fourthly, the two studies comparing ultrasound to chest X-ray were undertaken wholly among ventilated patients. This spectrum bias would augment ultrasound sensitivity since patients would be more likely to have extensive (and more easily detectable) consolidation. The necessarily supine chest X-rays would render films less sensitive for consolidation, again exaggerating the benefit of ultrasound.

The variable units of analyses employed across studies also introduce additional concerns. Different units of analyses had an evident effect on test accuracy. The use of lung regions for analysis (as opposed to lungs) diminished sensitivity, inflated specificity, and gave the misleading appearance of greater precision. Another drawback of different units of analyses across studies is that meta-analysis of results would be misleading because studies using lung regions would have undue numerical weight. For these reasons, we highly recommend future studies be conducted and reported always including patients as the unit of analysis. This is the most appropriate and relevant unit, particularly since individual patients are usually the unit of clinical management.

In addition, we strongly recommend future studies should compare different tests in the same patients and present results as 2 by 2 tables of paired results separately in disease-positive and disease negative patients. This is important to understand whether false-positive and false-negative results occur in the same or different patients, and to design subsequent studies.

Compared to previous systematic reviews[10,11], the distinguishing features of our review were: i) the emphasis on a single clinical presentation ie ARF; in this 'high stakes' patient group, the additional resources required to perform ultrasound are better justified than in less severe clinical presentations; ii) the requirement for a CT reference; providing greater confidence in estimates of diagnostic accuracy; iii) the focus on a single radiographic abnormality *i.e.* consolidation rather than the clinical diagnosis of pneumonia, removing the risk of bias of incorporating clinical information into the target condition, and; iv) the pre-registered systematic review protocol.

Limitations.

This review is limited by the small number of studies meeting inclusion as of August 2014. Four studies were performed by three investigator groups, limiting generalizability to other clinical environments. Where more than one ultrasound sign was used to diagnose consolidation, test characteristics of individual signs for consolidation were not assessed. The small number of studies and different units of analyses prevented meta-analysis, exploration of clinical and methodological heterogeneity, and pooled comparisons between ultrasound and chest X-ray.

Conclusion.

Based on a small body of evidence at high risk of selection bias and index test bias, ultrasound is both sensitive and specific for CT-detected consolidation in acute respiratory failure. Heterogeneous units of analyses between studies limited comparisons between studies.

While ultrasound may have a role as an add-on test in ARF when the chest X-ray is negative for consolidation, this possibility is tempered by the narrow evidence base available associated with

substantial risks of bias and applicability concerns. We conclude there is insufficient evidence to support the widespread introduction of ultrasound to detect consolidation in hospitalised patients diagnosed with acute respiratory failure.

Robustly designed studies are needed, controlling for the fundamental biases discussed above. They should aim to determine if an add-on, or replacement, test strategy is truly beneficial and identify clinical determinants of test accuracy. Ultrasound should be compared to current methods and also to emerging diagnostic alternatives using biomarkers[23] and other novel imaging[24]. The feasibility of implementing ultrasound should also be studied, coupled with clinical and cost-effectiveness modelling.

Author contributions.

MH- Original idea, first draft of protocol, first reviewer, first draft of manuscript. SM- Supervised protocol development and overall review. JPC- Second reviewer. NMR- Third reviewer. EKH- Search design and execution. All the authors analysed the data and wrote the manuscript.

Conflict of Interests

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

1. Lewandowski K, Metz J, Deutschmann C, et al. Incidence, severity, and mortality of acute respiratory failure in Berlin, Germany. Am J Respir Crit Care Med. 1995 Apr;151(4):1121-5.

2. Luhr OR, Antonsen K, Karlsson M, et al. Incidence and mortality after acute respiratory failure and acute respiratory distress syndrome in Sweden, Denmark, and Iceland. The ARF Study Group. Am J Respir Crit Care Med. 1999 Jun;159(6):1849-61.

3. Woodhead M, Welch CA, Harrison DA, Bellingan G, Ayres JG. Community-acquired pneumonia on the intensive care unit: secondary analysis of 17,869 cases in the ICNARC Case Mix Programme Database. Crit Care. 2006;10 Suppl 2:S1.

4. Ray P, Birolleau S, Lefort Y, et al. Acute respiratory failure in the elderly: etiology, emergency diagnosis and prognosis. Crit Care. 2006;10(3):R82.

5. Ovenfors C, Hedgecock MW. Intensive care unit radiology: problems of interpretation. Radiol Clin North Am 1978:16:407-439

6. Khan AN, Al-Jahdali H, Al-Ghanem S, Gouda A. Reading chest radiographs in the critically ill (Part I): Normal chest radiographic appearance, instrumentation and complications from instrumentation. Ann Thorac Med. 2009 Apr;4(2):75-87.

7. Bouhemad B, Zhang M, Lu Q, Rouby JJ. Clinical review: Bedside lung ultrasound in critical care practice. Crit Care. 2007;11(1):205. Review.

8. Hew M, Heinze S. Chest Ultrasound in Practice: a review of utility in the clinical setting. Intern Med J. 2012 Aug;42(8):856-65.

9. Volpicelli G, Elbarbary M, Blaivas M, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med. 2012 Apr;38(4):577-91.

10. Hu QJ, Shen YC, Jia LQ, et al. Diagnostic performance of lung ultrasound in the diagnosis of pneumonia: a bivariate meta-analysis. Int J Clin Exp Med. 2014 Jan 15;7(1):115-21

11. Chavez MA, Shams N, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. Respir Res. 2014 Apr 23;15:50.

12. British Thoracic Society (BTS) Community Acquired Pneumonia in Adults Guideline Group. Guidelines for the management of community acquired pneumonia in adults: update 2009.October 2009 Vol 64 Supplement III.

13. Schneider J, Sweberg T. Acute respiratory failure. Crit Care Clin. 2013 Apr;29(2):167-83.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097.

15. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. AJR Am J Roentgenol. 2003 Jul;181(1):51-5. Review.

16. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.

17. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

18. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology. 2004 Jan;100(1):9-15.

19. Lichtenstein DA, Lascols N, Mezière G, Gepner A. Ultrasound diagnosis of alveolar consolidation in the critically ill. Intensive Care Med. 2004 Feb;30(2):276-81.

20. Xirouchaki N, Magkanas E, Vaporidi K, et al. Lung ultrasound in critically ill patients: comparison with bedside chest radiography. Intensive Care Med. 2011 Sep;37(9):1488-93.

21. Refaat R, Abdurrahman L. The diagnostic performance of chest ultrasonography in the up-to-date work-up of the critical care setting. The Egyptian Journal of Radiology and Nuclear Medicine (2013) 44, 779–789.

22. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med. 2013 Apr 2;158(7):544-54

23 Bozinovski S, Hutchinson A, Thompson M, et al. Serum amyloid a is a biomarker of acute exacerbations of chronic obstructive pulmonary disease.Am J Respir Crit Care Med. 2008 Feb 1;177(3):269-78.

24. Masaryk T, Kolonick R, Painter T, Weinreb DB. The economic and clinical benefits of portable head/neck CT imaging in the intensive care unit. Radiol Manage. 2008 Mar-Apr;30(2):50-4.

TABLES

Table 1. Differences in the two studies by Lichtenstein.

	Lichtenstein 2004 a	Lichtenstein 2004 b
Institution	Pitié-Salpétrière Hospital (stated in text)	Hopital Ambroise-Pare (implied by; author affiliation; acknowledgement of the ICU department head; acknowledgement of the Radiology department head where scans took place)
Type of ICU	Surgical	Medical
CT scanner used	Tomoscan SR 7000 (Philips, Eindhoven, The Netherlands)	CT Twin Flash (Elscint Limited, Haifa, Israel)
Reason for CT	Research study protocol	Clinical decision
Recruitment Period	Unstated, but inferred as 1993-1997 (from another paper arising from the same CT ARDS study, Puybasset et al, 2000)	Unstated

Table 2. Included studies: patient characteristics.

Author Country	Study type/ Period	Demographics	Setting	Inclusion	Illness severity	Mechanical ventilation
Lichtenstein 2004a France	Cohort, likely 1993-1997	*n*=32, *Age* 58 +/-15 (SD), *M:F* Not stated	Surgical ICU	ARDS, (pneumonia 18, pulmonary contusion 4, aspiration pneumonia 4, fat embolism 1, septic shock 3, cardiopulmonary bypass 2)	Lung injury severity score 2.6 +/- 0.8(SD), *(i.e. severe),* ARDS severity score of 11 +/- 6 (SD), Mortality 42%	All
Lichtenstein 2004b, France	Cohort, Period not stated	*n*=60, *Age* 53 (range 20-84), *M:F* 37:23	Medical ICU	Patients with critical illness requiring chest CT	Not stated	30/60
Xirouchaki 2011, Greece	Cohort, Period not stated	*n*=42, *Age* 57.1 +/-21.5 (SD), *M:F* 34:8	Mixed ICU	Patients with critical illness requiring chest CT (sepsis/multiorgan failure 18, trauma 11, Airways disease 7, pulmonary oedema 2, post-operative respiratory failure 2)	APACHE2 16.5 +/-6.5 (SD)	All
Refaat 2013 Egypt	Cohort, 2012-13	*n*=90, *Age* 50 (45-65), *M:F* 55:35	Chest ICU	Respiratory failure (pneumonic consolidation 16, lung cancer 7, lung metastases 7, pleural effusion 36, pneumothorax 12, hydropneumothorax 6, mesothelioma 7)	Not stated	Not stated

n: number, M: male, F: female, ICU: intensive care unit, ARDS: acute respiratory distress syndrome, CT: computer tomography, APACHE2: Acute Physiology and Chronic Health Evaluation II.

Table 3. Included studies: Ultrasound technique, signs of consolidation and units of analysis

Study	Ultrasound timing	Sono-grapher	Probe/ scanner	Scan position	Scan protocol	US signs of consolidation	Unit of analysis	Consolidation prevalence
Lichtenstein 2004 a	Within 24 hours of ICU admission (approximated to ARF diagnosis)	1 intensivist (of 2), experience not quantified	Micro-convex 5 MHz, Portable (Hitachi 405)	Supine	12 lung regions	Tissue-like pattern, no change in dimensions with respiration. Air bronchograms not mandatory	Lung region (12/ patient)	31% of lung regions
Lichtenstein 2004 b	Unstated	2 intensivists (kappa coefficient 0.89) experience not quantified	Micro-convex 5 MHz, Portable (Hitachi 405)	Supine	12 lung regions	Tissue-like pattern, arising from the pleural line, irregular deep border (regular if lobar), no change in dimensions with respiration. Air bronchograms not used	Lung (2/ patient)	56% of lungs
Xirouchaki 2011	Unstated	1 intensivist, 4 years' experience	Micro-convex 5-9 MHz, Portable (Hitachi 8500)	Supine & lateral	12 Lung regions	Tissue-like pattern +/- power doppler. Irregular deep border not used	Lung (2/ patient) & Lung region (12/ patient)	24% of lungs, but 79% of lung regions

Refaat 2013	Unstated	1 radiologist, > 7 years' experience	Linear 7.5-10 MHz & convex 3.5 MHz, Portable Shenzhen mindray DP-1100 Plus)	Supine & lateral Erect when possible	12 lung regions	Hypoechoic pattern, non-homogenous echo-texture, irregular shape, serrated margin, air and fluid bronchograms	Patient	18% of patients

FIGURE LEGENDS

Figure 1. PRISMA flow diagram of study identification and selection.

Figure 2. QUADAS2 risk of bias and applicability assessment of individual studies.

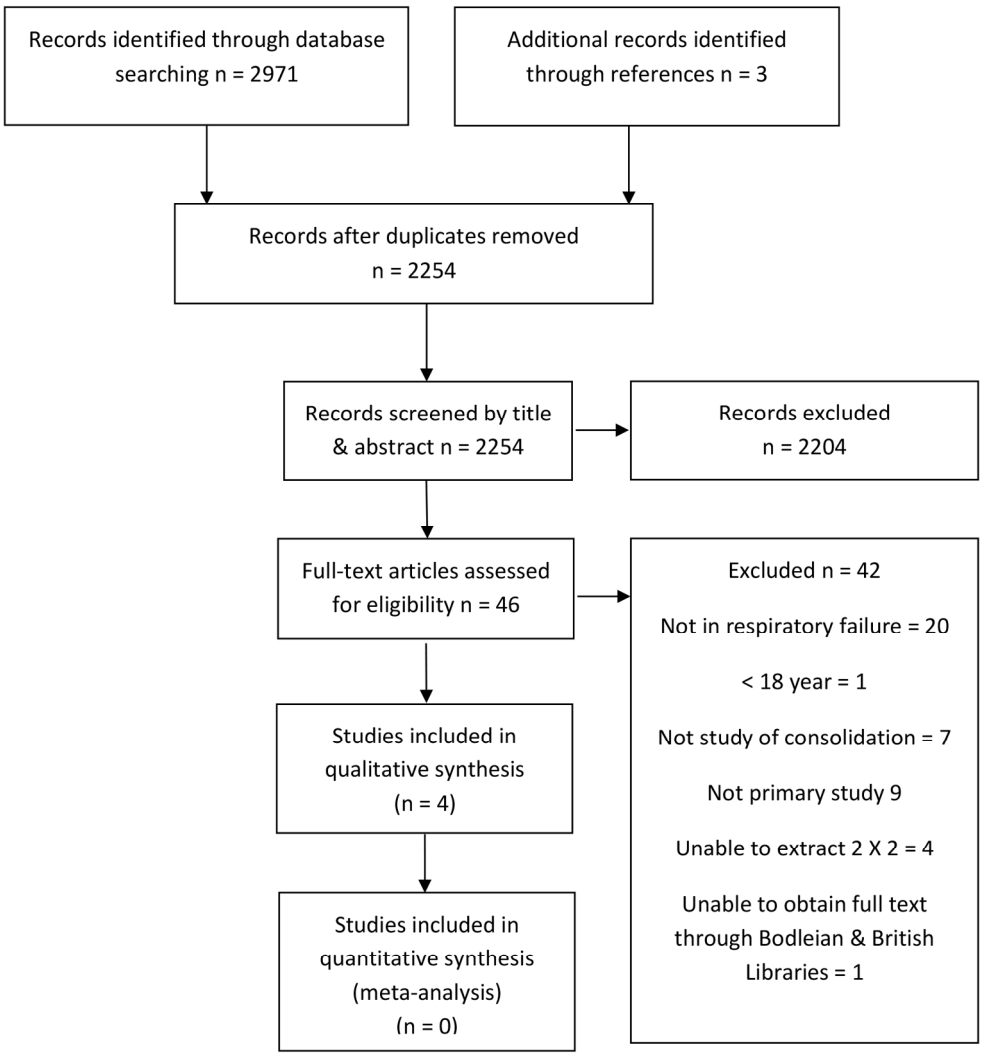
Figure 3. QUADAS2 risk of bias and applicability assessment across primary studies.

Figure 4. Sensitivity and specificity of ultrasound and chest X-ray for included studies. Note different units of analyses which preclude pooling of studies: *lung regions* (12 per patient) in Lichtenstein 2004a; *lungs* (2 per patient) in Lichtenstein 2004b and Xirouchaki 2011; and *individual patients* in Refaat 2013.

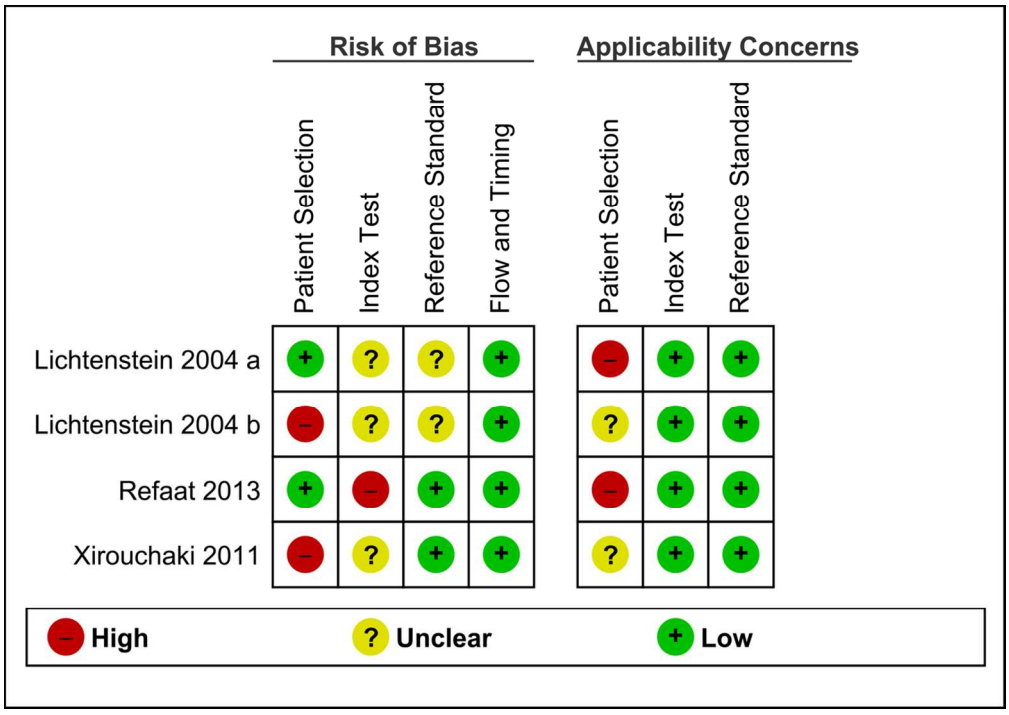
Figure 5. Sensitivity and specificity of ultrasound and chest X-ray for consolidation.

Figure 6. Direct comparisons for ultrasound and chest X-ray in two individual studies. Units of analysis are lungs (2 per patient) for Xirouchaki 2011, and lung regions for Lichtenstein 2004a (12 per patient).

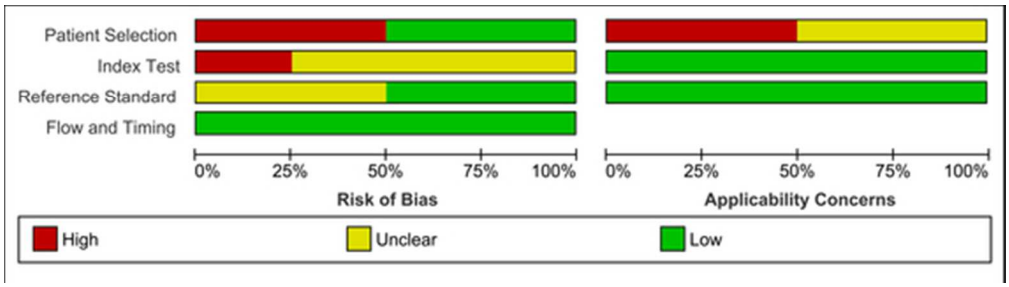
Figure 7. Impact of units of analyses on sensitivity and specificity of ultrasound and chest X-ray for consolidation. Data from Xirouchaki et al 2011 are stratified according to two units of analysis; 'lung' (2/patient) and 'lung region' (12/patient). In comparison to lung analysis, *lung region* analysis reduces sensitivity but inflates specificity. It also increases total study numbers, giving the appearance of tighter estimates of precision.



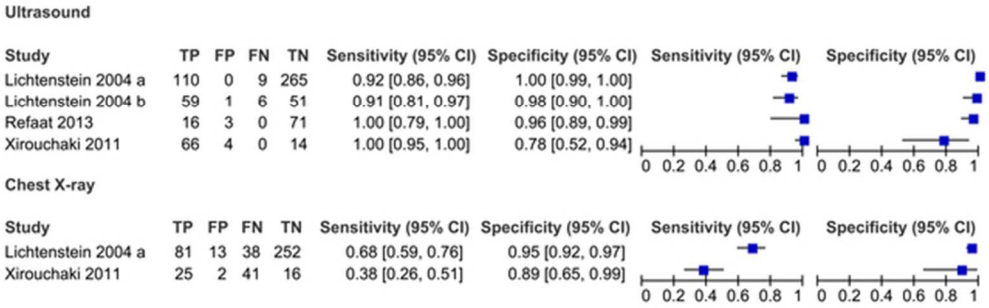
PRISMA flow diagram of study identification and selection.
161x172mm (300 x 300 DPI)



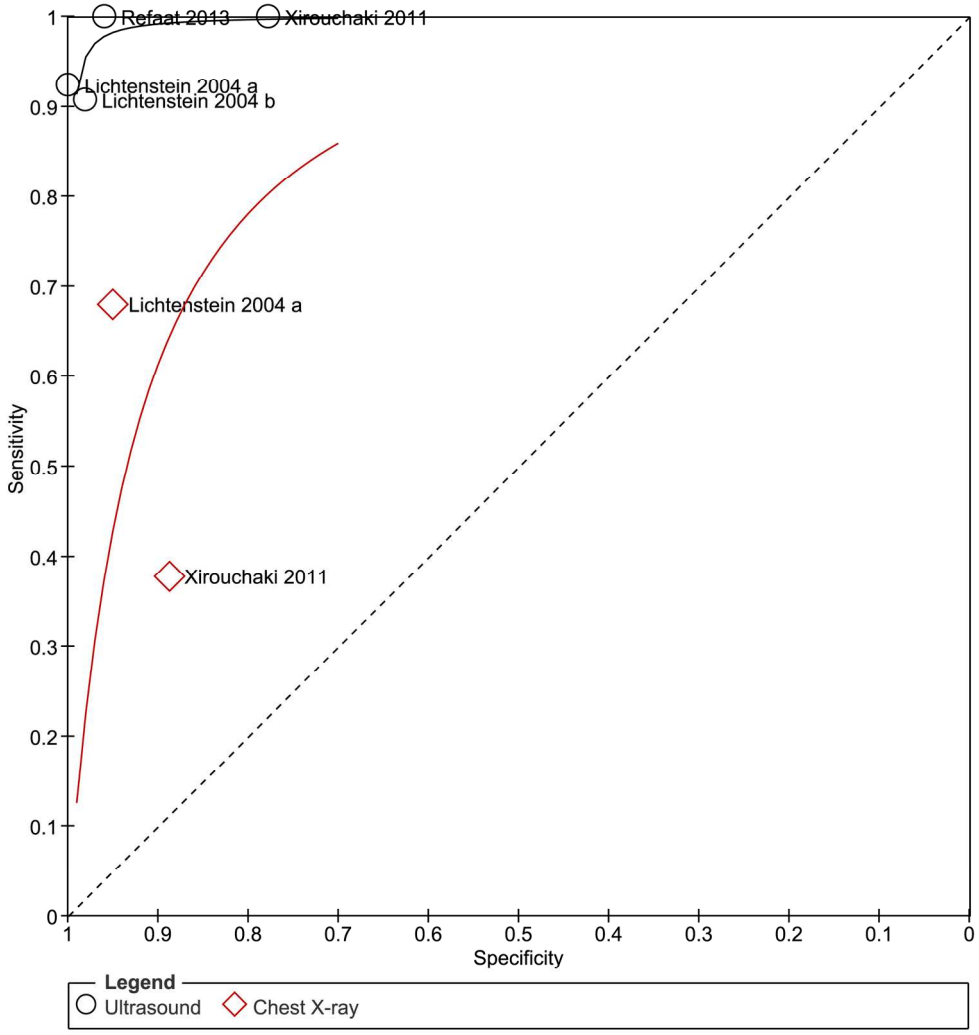
QUADAS2 risk of bias and applicability assessment of individual studies.
122x85mm (300 x 300 DPI)



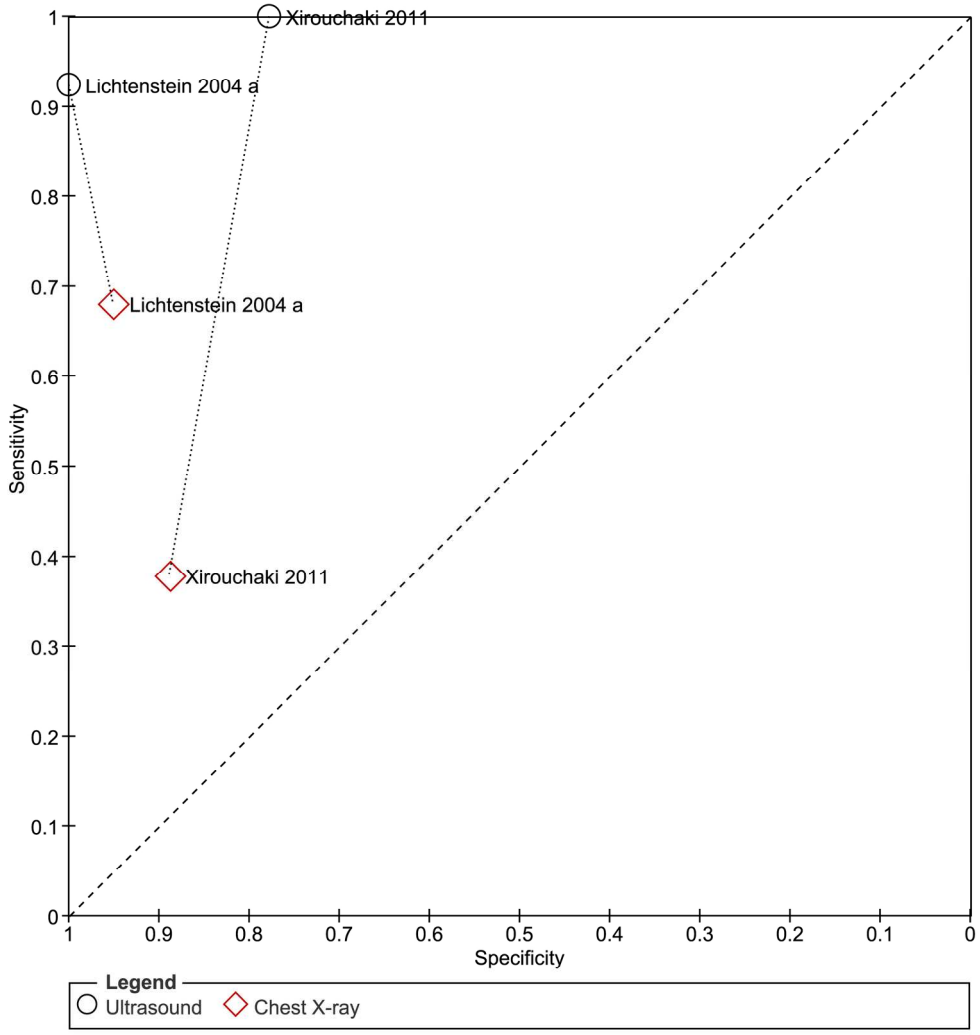
QUADAS2 risk of bias and applicability assessment across primary studies.
48x13mm (300 x 300 DPI)



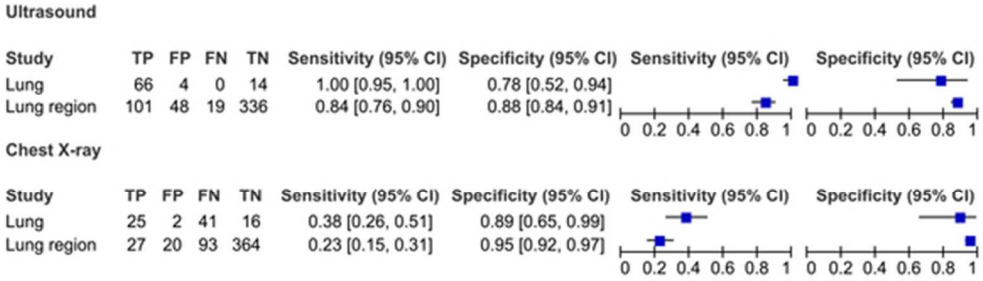
Sensitivity and specificity of ultrasound and chest X-ray for included studies. Note different units of analyses which preclude pooling of studies: lung regions (12 per patient) in Lichtenstein 2004a; lungs (2 per patient) in Lichtenstein 2004b and Xirouchaki 2011; and individual patients in Refaat 2013.
54x17mm (300 x 300 DPI)



Sensitivity and specificity of ultrasound and chest X-ray for consolidation.
188x203mm (300 x 300 DPI)



Direct comparisons for ultrasound and chest X-ray in two individual studies. Units of analysis are lungs (2 per patient) for Xirouchaki 2011, and lung regions for Lichtenstein 2004a (12 per patient).
188x203mm (300 x 300 DPI)



Impact of units of analyses on sensitivity and specificity of ultrasound and chest X-ray for consolidation. Data from Xirouchaki et al 2011 are stratified according to two units of analysis; 'lung' (2/patient) and 'lung region' (12/patient). In comparison to lung analysis, lung region analysis reduces sensitivity but inflates specificity. It also increases total study numbers, giving the appearance of tighter estimates of precision.
50x14mm (300 x 300 DPI)

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	6

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	13
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

ONLINE SUPPLEMENT

Appendix 1. Specific search strategies for three databases.

MEDLINE
Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

	Searches	Results	Search Type
1	exp Ultrasonography/	247016	Advanced
2	(ultrasonic adj3 (diagnos* or tomograph* or imaging*)).ti,ab.	3772	Advanced
3	(echotomograph* or echograph* or ultrasonograph* or sonograph* or ultrasound).ti,ab.	247848	Advanced
4	1 or 2 or 3	395876	Advanced
5	exp Pneumonia/	75417	Advanced
6	pneumon*.ti,ab.	139061	Advanced
7	bronchopneumon*.ti,ab.	3003	Advanced
8	Respiratory Tract Infections/	31632	Advanced
9	("lower respiratory tract infection*" or "lower respiratory infection*" or LRTI).ti,ab.	5539	Advanced
10	(lung adj3 (inflamm* or infect*)).ti,ab.	16851	Advanced
11	exp Critical Illness/	16852	Advanced
12	5 or 6 or 7 or 8 or 9 or 10 or 11	224498	Advanced
13	exp Lung/	230110	Advanced
14	exp Thorax/	40771	Advanced
15	(lung* or chest* or thora* or respirat* or alveol*).ti,ab.	1023348	Advanced
16	13 or 14 or 15	1117732	Advanced
17	4 and 12 and 16	897	Advanced
18	17	897	Advanced
19	limit 18 to (humans and ("all adult (19 plus years)" or "young adult (19 to 24 years)" or "adult (19 to 44 years)" or "young adult and adult (19-24 and 19-44)" or "middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)"))	431	Advanced
20	Animals/	5486116	Advanced
21	17 not 20	809	Advanced
22	adult.mp. or middle aged.sh. or age:.tw.	6843828	Advanced
23	4 and 12 and 16 and 22	522	Advanced
24	23 not 20	496	Advanced

EMBASE.

Embase 1974 to 2013 October 21

	Searches	Results	Search Type
1	exp echography/	503948	Advanced
2	exp echotomography/	950	Advanced
3	(ultrasonic adj3 (diagnos* or tomograph* or imaging*)).ti,ab.	4613	Advanced
4	(echotomograph* or echograph* or ultrasonograph* or sonograph* or ultrasound).ti,ab.	329397	Advanced
5	1 or 2 or 3 or 4	635012	Advanced
6	exp pneumonia/	188700	Advanced
7	pneumon*.ti,ab.	169236	Advanced
8	bronchopneumon*.ti,ab.	3736	Advanced
9	exp lower respiratory tract infection/di [Diagnosis]	35926	Advanced
10	(lung adj3 (inflamm* or infect*)).ti,ab.	19808	Advanced
11	(lower adj3 ("respiratory tract infection*" or "respiratory infection*")).ti,ab.	6896	Advanced
12	LRTI.ti,ab.	860	Advanced
13	exp critical illness/	21700	Advanced
14	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13	324922	Advanced
15	(lung* or chest* or thora* or respirat* or alveol*).ti,ab.	1246364	Advanced
16	exp Lung/	224539	Advanced
17	exp Thorax/	75787	Advanced
18	15 or 16 or 17	1332427	Advanced
19	5 and 14 and 18	3035	Advanced
20	19	3035	Advanced
21	limit 20 to (human and (adult <18 to 64 years> or aged <65+ years>))	1389	Advanced
22	adult.mp. or middle aged.sh. or age:.tw.	7323441	Advanced
23	animal experiment/	1722105	Advanced
24	19 and 22	1732	Advanced
25	24 not 23	1726	Advanced

Web of Knowledge

Set	Results	
# 10	1,413	#7 AND #4 AND #1 Timespan=All years Search language=English
# 9	764	#8 AND #7 AND #4 AND #1 Timespan=All years Search language=English
# 8	Approximately 8,982,030	Topic=(adult* OR "middle age*" OR aged) OR Title=(adult* OR "middle age*" OR aged) Timespan=All years Search language=English
# 7	Approximately 420,383	#6 OR #5 Timespan=All years Search language=English
# 6	Approximately 73,103	Topic=(("critical* ill*")) OR Title=(("critical* ill*")) Timespan=All years Search language=English
# 5	Approximately 352,396	Topic=(pneumon* OR bronchopneumon* OR bronchit*) OR Title=(pneumon* OR bronchopneumon* OR bronchit*) Timespan=All years Search language=English
# 4	Approximately 645,870	#3 OR #2 Timespan=All years Search language=English
# 3	Approximately 192,826	Topic=(((ultrasonic OR ultrasound) SAME (diagno* OR tomograph* OR imaging*))) OR Title=(((ultrasonic OR ultrasound) SAME (diagno* OR tomograph* OR imaging*))) Timespan=All years Search language=English
# 2	Approximately 634,720	Topic=(echotomograph* OR echograph* OR ultrasonograph* OR sonograph* OR ultrasound) OR Title=(echotomograph* OR echograph* OR ultrasonograph* OR sonograph* OR ultrasound) Timespan=All years Search language=English
# 1	Approximately 2,257,730	Topic=(lung* or chest* or thora* or respirat* or alveol*) OR Title=(lung* or chest* or thora* or respirat* or alveol*) Timespan=All years

Appendix 2. Data extraction form

STUDY IDENTIFIERS

Author	Year	Journal	Country

STUDY TYPE Cohort/Case control/RCT

INCLUSION CRITERIA

1, Timing	Within 24 hours (can still include if no)	
2. Index Test	Chest ultrasound	
3. Target condition	Consolidation	
4. Comparator	Chest CR	
5. Reference	Chest CT	

EXCLUSION

1. Not primary study		
2. Not in respiratory failure		
3. Not chest ultrasound		
4. Not to diagnose consolidation		
5. Unable to obtain translation		
6. Unable to extract 2 X2 data		
7. Unable to obtain paper		

PARTICIPANT DETAILS

1. Dates recruited		
2. Number		
3. Age		
4. Gender (M:F)		
5. Location		
6. Illness severity		
7. ? mechanical ventilator		

INDEX DETAILS

1. Sonographer		
2. equipment		
3. Extent of examination		

COMPARATOR- Erect or supine film?

REFERENCE – CT equipment, protocol, reader.

FLOW- interval between performance of ultrasound and CXR/CT

DIAGNOSTIC PERFORMANCE (and unit of analysis ….)

	CT +	CT -	Sens	Spec
US +				
US -				
CXR +				
CXR -				

Appendix 3. Quality Assessment Forms

Domain		Signalling Question/Checklist	*Rating Guidelines*
Patient Population	**Bias**	Consecutive/random sample?	*Rate 'yes' if stated consecutive, or able to ensure randomness. Rate 'unclear' if neither, or 'convenience,' or data implausible. Rate 'no' if stated non-consecutive/non-random, or inclusion based on clinical decision for CT.*
		Case/control avoided?	*Rate 'no' if two-gate case control design*
		Inappropriate exclusions avoided?	*Rate 'no' if exclusions would not have affected ability to undergo testing, or post-hoc Rate 'unclear' if ambiguous wording raises possibility of inappropriate exclusions*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'
	Applicability	Respiratory failure/respiratory support/intensive-care	*Rate 'low concern' if indicates respiratory failure or requirement for respiratory support including high flow O_2 or ventilation.*
			Rate 'unclear concern' if indication for intensive care or intubation not clearly respiratory failure
			Rate 'high concern' if some patients clearly did not have respiratory failure, or confined to a subgroup of respiratory failure aetiologies
Ultrasound (Index)	**Bias**	Sonographer blind to CT AND clinical data?	*Rate as stated, 'unclear' if not stated.*
	Applicability	Reasonable sonographer, scanner and protocol	*Rate 'high concern' if deviates from usual practice to extent of irreproducibility, very old equipment, or incomplete scanning protocol*
CT (Reference)	**Bias**	Is CT likely to correctly classify consolidation?	*Rate 'unclear' if very old scanner, 'no' if examination incomplete*
		CT reported blind to ultrasound AND	*Rate as stated, 'unclear' if not stated.*

		clinical data?	*Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if both responses 'yes'*
	Applicability	Reasonable scanner and protocol	*Rate 'high concern' if deviates from usual practice to extent of irreproducibility.*
Flow & Timing	**Bias**	Interval between ultrasound and CT appropriate?	*Rate 'unclear' if not stated.*
		All patients underwent CT?	*Rate 'unclear' if not stated.*
		All patients analysed?	*Rate 'unclear' if not stated.*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'

(Table 1 continued)

Protocol for a systematic review & meta-analysis: The Diagnostic Accuracy of Chest Ultrasound for CT-detected Radiographic Consolidation in hospitalised Adults with Acute Respiratory Failure.

Version 2 (20 November 2013), supercedes Version 1 (9 August 2013)

Abstract.

Introduction.

Diagnosing pneumonia as a cause of acute respiratory failure can be challenging. Standard chest X-rays have limited accuracy.

Objectives.

This systematic review and meta-analysis will estimate the diagnostic accuracy of chest ultrasound as an initial test for radiographic consolidation in acute respiratory failure, and compare it against that of chest X-ray.

Search Strategy.

Medline, EMBASE, and the Science Citation Index will be searched. Reference lists, citing alerts, and related articles will be examined. If required, authors will be contacted for additional information.

Study Selection.

Studies: randomised trials, cohort and nested (one-gate) case-control studies. *Timing:* initial testing. *Participants:* adults in acute respiratory failure. *Index:* chest ultrasound. *Comparator:* chest X-ray. *Target Condition:* radiographic consolidation. *Reference:* chest computer tomography.

Data Collection & Analysis.

Two reviewers will independently select studies for inclusion, assess study quality and extract data. The sensitivity and specificity of ultrasound for consolidation in acute respiratory failure will be determined and compared against that of chest X-ray. Summary point estimates and 95% confidence intervals for ultrasound and chest X-ray will be determined by bivariate and hierarchical models. Heterogeneity will be explored by subgroup analyses and meta-regression.

Interpretation.

Results will inform policy-makers and clinicians regarding benefits of introducing chest ultrasound for pneumonia diagnosis in acute respiratory failure, and identify patients who may benefit most.

A. Rationale.

Acute respiratory failure.

Acute respiratory failure (a low blood-oxygen level) is a life-threatening state which requires urgent admission to hospital, often to the intensive care unit. Immediate provision of supplemental oxygen is critical. The next priority is then to diagnose (and treat) the underlying cause of respiratory failure.

Pneumonia can cause respiratory failure.

One major cause of respiratory failure is pneumonia, an infection of the lung parenchyma usually caused by bacterial or viral pathogens, many of which are susceptible to antimicrobial therapy.

Diagnostic tests for pneumonia in respiratory failure.

The history and physical examination may point to the presence of pneumonia, but its *sine qua non* is the finding of lung shadowing on radiological imaging, otherwise known as 'consolidation'.

However, traditional imaging techniques have significant drawbacks in detecting 'consolidation' in this setting;

a) Bedside chest X-ray may fail to detect consolidation due to the suboptimal images obtained in acutely unwell patients (Ovenfors et al, 1978).
b) Chest computer tomography (CT) has greater accuracy (Mirvis et al, 1987) but involves the risk of transporting patients who require respiratory support away from the safety of their ward.

The advantages of chest ultrasound.

Chest ultrasound may overcome the drawbacks of traditional imaging for pneumonia (Figure 1);

a) Unlike X-ray, ultrasound does not depend on optimal positioning.
b) Unlike CT scanning, ultrasound can be brought to the patient's bedside.

Figure 1. Ultrasound machinery used for chest sonography.

Current uncertainty.

Narrative reviews (including our own; Hew & Heinze, 2012) and consensus guidelines (Volpicelli et al, 2012) have identified evidence for the use of ultrasound to diagnose pneumonia.

However, a systematic review of chest of ultrasound to diagnose pneumonia in acute respiratory failure is needed to summarise evidence on its diagnostic accuracy. A test should ideally not enter clinical practice until its diagnostic accuracy has been clearly defined. Without such data, the introduction of the test may cause errors in diagnostic reasoning and jeopardise patient safety.

It is also necessary to compare a new test against tests currently in clinical use for that purpose, in order to determine the best role (if any) for the new test.

B. REVIEW QUESTION & OBJECTIVES.

'In the initial testing of patients with acute respiratory failure,
what is the accuracy of chest ultrasound to diagnose CT-detected consolidation,
when compared to bedside chest X-ray?'

This systematic review will therefore:

a) Review evidence on the ***diagnostic test accuracy*** of chest ultrasound for radiological consolidation in patients with acute respiratory failure.
b) Perform a ***comparison*** of chest ultrasound diagnostic accuracy with that of chest X-ray.

C. METHODS- DATA COLLECTION.

Inclusion Criteria.

Inclusion criteria for studies in this systematic review have been developed in accordance with recommendations from the Cochrane Handbook of Diagnostic Test Accuracy (Bossuyt et al, 2008).

Studies.

Cohort studies (where all patients have acute respiratory failure) and nested one-gate case-control studies (comprising patients with and without radiographic consolidation) will be included. Randomised controlled trials allocating respiratory failure patients to either ultrasound or a standard comparator will be included.

Studies which evaluate only chest ultrasound will be included.

Studies which evaluate chest ultrasound against standard testing with chest X-ray will also be included. The index comparisons may be either *paired* or *randomised*.

Timing.

The best place for ultrasound along the diagnostic pathway must be specified, since altering its place in a sequence of tests may change its diagnostic accuracy (Leeflang, et al, 2008; Reitsma et al, 2012). Given its advantages over chest X-ray and CT, the best role for ultrasound is as an initial test. Studies

will thus be included if < 24 hours elapse between acute respiratory failure diagnosis and the chest ultrasound.

If insufficient studies with this design are located, studies which employ ultrasound r> 24 hours after diagnosis of acute respiratory failure will be included, and explored for their likely impact on statistical heterogeneity.

Participants.

Studies of adult patients (age > 18) admitted to hospital with acute respiratory failure will be included. Patients admitted to emergency wards, general wards, high-dependency and intensive-care units will be included.

Studies where patients are well enough to be discharged home directly from the emergency department within 24 hours of presentation will be excluded.

Acute respiratory failure will be defined as one of the following:

a) An arterial partial pressure of oxygen (PaO_2) < 60 millimetre of mercury (mm Hg), without supplemental oxygen.
b) An arterial oxygen saturation of < 90% measured by pulse oximetry, without supplemental oxygen.
c) Where supplemental oxygen is required in order to raise the PaO2 to > 60 mm Hg or arterial oxygen saturations to > 90%.
d) Studies which do not explicitly define respiratory failure but make reference to credible diagnostic conventions based on the principles above will also be considered for inclusion.

Index Test.

Studies which use chest ultrasound will be included. Chest ultrasound will be defined as the use of B (brightness)-mode ultrasound to systematically examine the lungs and pleura. The investigation may be performed by either clinicians or radiologists.

Comparator.

Chest X-ray is universally performed as the initial investigation for respiratory failure. Thus if studies evaluate chest X-ray against chest ultrasound, they will be included for direct comparison.

Studies evaluating only chest X-ray (without chest ultrasound) will be excluded from review. Given the potential for clinical heterogeneity between studies, indirect comparisons will not be performed.

Target condition.

As mentioned, ultrasound, chest X-ray and CT have characteristic imaging findings suggestive of pneumonia, *i.e.* 'consolidation'. The target condition will be framed as this radiological finding ('consolidation') rather than a clinical diagnosis ('pneumonia'). This allows comparison of like with like; images (on ultrasound) referenced to images (on radiology).

Studies will therefore be included if they measure the accuracy of chest ultrasound for the *radiological finding* of consolidation defined by the reference standard.

Studies which examine the accuracy of chest ultrasound to detect the *clinical diagnosis* of pneumonia (defined by a combination of history, examination and imaging) will be excluded, in order to avoid bias arising from integrating non-imaging data into the diagnostic algorithm.

Reference standard.

In patients with acute respiratory failure, X-rays obtained at the bedside result in poor image quality (Ovenfors et al, 1978) and poor diagnostic sensitivity (Henschke et al, 1996). Conversely, CT is highly sensitive for consolidation even in very ill patients (Rubinowitz et al, 2007). Studies will therefore only be included if they use chest CT scanning as the reference standard.

Studies may still be included if only some patients receive CT (differential verification), but only if data can be obtained for analysis in the subgroup of patients who underwent CT.

Chest CT will be defined as the use of helical CT to examine the thorax.

Search Strategy.

The search will be conducted according to guidance from the Cochrane Handbook of Diagnostic Test Accuracy (de Vet et al, 2008) and include the following principles:

a) The search strategy will be developed in consultation with a healthcare librarian experienced with supporting systematic reviews.
b) The search will be carried out independently by two reviewers. Disagreements will be resolved by discussion or consultation with a third reviewer.
c) Multiple electronic databases will be searched including, but not confined to, MEDLINE, EMBASE and the ISI Science Citation Index.
d) The search strategy will include the concepts: (i) index test *i.e.* ultrasound AND target condition *i.e.* pneumonic consolidation, or, (ii) index test *i.e.* ultrasound AND participants *i.e.* acute respiratory failure.
e) Each concept will be described by a large variety of terms (text words and subject headings).
f) For each database, a number of preliminary searches will be conducted using a range of textwords and subjects headings. Further searches will then be run using additional textwords and subjects headings derived from studies identified on initial searches.
g) In order to maximise search sensitivity, relevant search filters will NOT be used (such as the new EMBASE indexing term 'diagnostic accuracy study' or the MEDLINE subject heading 'sensitivity and specificity').
h) The reference lists of relevant studies will be examined and citation searches will be performed. Citing alerts in electronic journals and the 'related articles' feature in PubMed will be used to identify further relevant articles.
i) The search strategy will be fully described as an appendix in the final published review.

Study Identification.

Search results will be screened by two independent reviewers. Studies that appear relevant will be obtained and assessed for inclusion by each reviewer. Disagreements will be resolved by discussion or referral to a third reviewer. The process of study identification will be shown by a flow diagram.

Data Extraction.

Data extraction forms will be developed using Microsoft Access. The forms will be trialled on a small number of studies and modified appropriately.

Data will be extracted independently by two reviewers. Disagreements will be resolved by discussion or referral to a third reviewer.

Data items to be extracted will include:

a) *Study identification*- author, year, location.
b) *Study details*- cohort, case-control, randomised trial.
c) *Inclusion*- Timing, index test, target condition, comparator, reference.
d) *Exclusion*- The following hierarchy will be employed: (i) Not a primary study (ii) Patients not in respiratory failure (iii) Not a study of chest ultrasound (iv) Not a study to diagnose consolidation (v) Unable to obtain a translation (vi) Unable to extract 2 X 2 data (vii) Unable to obtain the paper.
e) *Participants*- number, age, gender, location, illness severity, whether on a ventilator.
f) *Index*- sonographer (clinician versus radiologist), equipment (high-end versus lightweight portable), thoroughness of sonographic examination (whether any views were excluded).
g) *Comparator and reference*- interval between performance of ultrasound and chest X-ray/CT scanning.
h) *Diagnostic performance*- 2 X 2 contingency tables of index and (where available) comparator tests denoting true positives, true negatives, false positives, false negatives.

Quality Assessment.

QUADAS2 (the Revised Tool for Quality Assessment of Diagnostic Accuracy Studies; Whiting et al, 2011) will be used to assess quality in this review. As per QUADAS2 the quality assessment process will be divided into 4 phases;

Articulating the review question.

Participants, index, reference, and flow and timing are all already defined in the review question earlier.

Tailoring the tool to the specific review.

Four steps are suggested (Whiting et al, 2011);

a) Tailoring the tool content. (Table 1).
b) Developing rating guidelines. (Table 1).
c) Piloting the tool and guidelines. This will be done on a randomly-selected included study.
d) Applying the tool to all included studies. This will be done on all included studies.

Drawing flow diagrams for each study.

A flow diagram for each included study will assist judgments regarding bias and applicability.

Applying the tool to each study to make judgments on bias and applicability.

Finally, the tool will be applied to each included study. The risk of bias and level of applicability for each study will be summarised in tables (Table 2) and graphs (Figure 2). [Summary scores for individual studies will not be undertaken as they are prone to error (Juni et al, 1999).]

Table 1. Rating guidelines (blue) developed for the review-specific quality assessment (black).

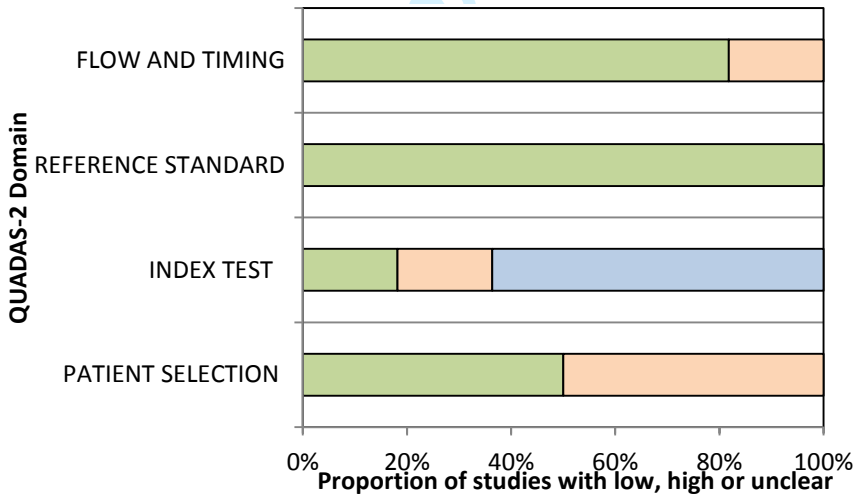
Domain		Signalling Question/Checklist	*Rating Guidelines*
Patient Population	**Bias**	Consecutive/random sample?	*Rate 'yes' if stated consecutive, or able to ensure randomness. Rate 'unclear' if neither, or 'convenience,' or data implausible. Rate 'no' if stated non-consecutive/non-random or inclusion based on clinical decision for CT.*
		Case/control avoided?	*Rate 'no' if two-gate case control design*
		Inappropriate exclusions avoided?	*Rate 'no' if exclusions would not have affected ability to undergo testing, or post-hoc Rate 'unclear' if ambiguous wording raises possibility of inappropriate exclusions*
			Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'
	Applicability	Respiratory failure/respiratory support/intensive-care	*Rate 'low risk' if indicates respiratory failure or requirement for respiratory support including high flow O_2 or ventilation.*
			Rate 'unclear' if indication for intensive care or intubation not clearly respiratory failure
			Rate 'high risk' if some patients clearly did not have respiratory failure, or confined to a subgroup of respiratory failure aetiologies
Ultrasound (Index)	**Bias**	Sonographer blind to CT AND clinical data?	*Rate as stated, 'unclear' if not stated.*
	Applicability	Reasonable sonographer, scanner and protocol	*Rate 'high risk' if deviates from usual practice to extent of irreproducibility, very old equipment, or incomplete*

			scanning protocol
CT (Reference)	**Bias**	Is CT likely to correctly classify consolidation? CT reported blind to ultrasound?	*Rate 'unclear' if very old scanner, 'no' if examination incomplete* *Rate as stated, 'unclear' if not stated.* *Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if both responses 'yes'*
	Applicability	Reasonable scanner and protocol	*Rate 'no' if deviates from usual practice to extent of irreproducibility.*
Flow & Timing	**Bias**	Interval between diagnosis of respiratory failure and ultrasound < 24 hours? Interval between ultrasound and CT appropriate? All patients underwent CT? All patients analysed?	*Rate 'no' if > 24 hours* *Rate 'no' if > 24 hours.* *Rate 'unclear' if not stated.* *Rate 'unclear' if not stated.* *Rate 'unclear' if not stated.* *Rate overall as high risk of bias if at least one 'no'; unclear risk of bias if at least one 'unclear'; low risk of bias if all responses 'yes'*

Table 2. Suggested tabular presentation for quality assessment (adapted from Whiting et al, 2011).

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Study 1	?	☹	☺	☹	☺	☺	☺
Study 2							
Study 3							
etc.							

Figure 2. Graphical display of QUADAS2 quality assessment for risk of bias (from Whiting et al, 2011). A similar display will be generated for applicability concerns (not shown).



D. METHODS- DATA ANALYSIS.

Measures of test accuracy.

Primary outcomes.

a) *Diagnostic accuracy of Ultrasound.*
For each included study, a 2 X 2 contingency table will be extracted to allow calculation of sensitivity and specificity of ultrasound for consolidation with 95 % confidence intervals.

b) *Direct Comparison of Ultrasound versus chest X-ray.*
In studies where a direct comparison between ultrasound and chest X-ray is performed, sensitivity and specificity of chest X-ray for consolidation with 95% confidence intervals will also be derived.

Secondary outcomes.

Positive and negative likelihood ratios and diagnostic odds ratios will be calculated for each study. Positive and negative predictive values will also be presented, referenced specifically to the *prevalence* of pneumonia in the studies identified.

Missing data.

For studies where 2 X 2 contingency tables cannot be derived from the paper, corresponding authors will be contacted. If adequate data is still not obtained, these studies will be excluded from review and analysis.

Units of analysis.

Per patient.

It is anticipated that the unit of analysis in most studies will be individual patients.

Per lung.

It is possible that some studies, in an effort to increase sample size, may report ultrasound results separately for each lung.

Importantly, this latter method is prone to bias, since the two lungs in a patient with (or without) consolidation are not independent of each other. Furthermore, both sides of a patient would be scanned by a single sonographer, whose examination of the second lung is likely to be biased by findings from the first.

For this reason, the influence of the unit of analysis on statistical heterogeneity will be explored.

Descriptive statistics.

Paired Forest plots.

Paired Forest plots will be used to display 95% confidence intervals of sensitivity and specificity for each study.

Results will be stratified by study type, since this may influence the potential biases (*eg* spectrum bias in case-control designs, versus partial verification bias in cohort-type studies).

Receiver operator characteristics (ROC) plots.

Pairs of sensitivity and specificity from each study will be plotted in receiver operator characteristics (ROC) space. This will facilitate an assessment of whether sensitivity and specificity are negatively correlated (Reitsma et al, 2012).

Meta-analysis.

Feasibility.

The appropriateness of data pooling and meta-analysis will depend on the number of studies and participants, and the methodological and clinical homogeneity of included studies.

Methodology.

Meta-analysis if appropriate will be conducted with expert statistical assistance, given the complexity of this field and the rapid evolution in methodology (Reitsma et al, 2012).

The Moses-Littenberg model provided in Revman 5 (Moses t al, 1993) will not be used for meta-analysis, since it does not allow for random effects, nor does it provide estimates of heterogeneity between studies (Macaskill et al, 2010).

Instead, the Bivariate random effects (Reitsma et al, 2005) models will be employed to determine summary estimates of test accuracy and calculate reliable 95% confidence intervals around these parameters (Harbord et al, 2007). We are particularly interested in the summary estimates for ultrasound rather than analysis of the SROC itself.

Results from these models will be input into Revman 5 to depict;

a) A summary ROC plot for ultrasound,
b) A summary operating point for ultrasound, and,
c) A 95% confidence region around the summary point for ultrasound (Macaskill et al, 2010).

If bivariate random effects models do not converge, a univariate logistic regression random effects meta-analysis of sensitivity and specificity separately will be performed instead.

Heterogeneity.

Subgroup analysis- Describing heterogeneity.

The number of subgroup analyses will be kept to as low as possible to minimise chance findings.

However, the following pre-specified subgroups will be examined, based on;

a) The unit of analysis (person versus lung),
b) Need for mechanical ventilation (yes or no),
c) Location of patient management (general ward versus intensive care),
d) Ultrasound operator (clinician versus radiologist),
e) Sophistication of ultrasound equipment (high-end machines versus small portable devices).

Meta-regression- Explaining heterogeneity.

Depending on the total number of studies, the number of participants within studies, and the degree of heterogeneity, the influence of covariates on diagnostic accuracy may be further explored by meta-regression, available through both Bivariate and HSROC models (Reitsma et al, 2012).

Covariates potentially contributing to heterogeneity will be examined using descriptive ROC plots according to covariates. If there are sufficient studies, meta-regression will be considered. Such covariates include;

a) ***Clinical Heterogeneity*;** unit of analysis, need for mechanical ventilation, location of patient management, sonographer type, documented thoroughness of sonographic chest examination, and the sophistication of ultrasound equipment.
b) ***Methodological Heterogeneity***. Depending on the quality of the studies included, specific risks of bias (such as partial verification bias) may also be incorporated as covariates.

However, it is acknowledged that such study level covariates have limited power to detect differences in diagnostic accuracy between subgroups (Reitsma et al, 2012). Furthermore, it has been suggested that at least 10 studies per covariate are needed for robust meta-regression (Gagnier et al, 2012), and it is unlikely that a sufficient volume of studies will be included.

Direct comparison of ultrasound and chest X-ray.

Only studies undertaking both tests will be included for this comparative analysis; each patient may have either undergone both tests (for paired comparison) or be randomised to either test (for randomised comparison).

Even if insufficient data is available for meaningful direct comparison, indirect comparison will *not* be performed. It is likely that studies examining the diagnostic accuracy of chest X-ray may be different from those examining the accuracy of chest ultrasound, introducing significant bias to indirect comparisons.

Preliminary graphical analysis.

The sensitivity and specificity for both tests (ultrasound and chest X-ray) in each study will be plotted in ROC space using Revman 5, as single points joined by a line (Macaskill et al, 2010).

Test comparisons.

Depending on the data available from individual studies, tests comparisons may then be performed using the bivariate model, with outputs which may be entered into Revman to superimpose the summary estimates for each tests (ultrasound and chest X-ray) and their 95% confidence regions on the ROC scatterplot (Macaskill, 2010).

Sensitivity analyses.

The following sensitivity analyses will be performed in order to test the robustness of the primary outcomes.

a) Comparison of the summary operating point for ultrasound, with and without inclusion of studies with a high risk-of-bias in one or more domains on quality assessment.
b) Comparison of the summary operating point for ultrasound, with and without inclusion of studies with a single lung as the unit of analysis.

c) Comparison of the sensitivity and specificity of ultrasound versus chest X-ray, with and without inclusion of studies with a high risk of bias in one or more domains on quality assessment.

Software.

A number of software options will be considered;

a) Analyses will probably be performed with the latest version of Stata, using the *metandi* command for meta-analysis of test accuracy studies (Harbord, 2009). Stata or SAS codes for bivariate analysis including covariates will be used. Results would be input into Revman 5 for graphical display.
b) Alternatively, the freely available R-package *mada* command (Doebler 2012) may be used, which performs all the analyses described above including bivariate and HSROC models with covariates, and provides publication-ready figures as an integral part of the programme.
c) By the analysis stage of this systematic review, new statistical techniques and software options may have emerged that supersede the software options described above.

Thus the final decision regarding software selection will be made in consultation with a specialist biomedical statistician abreast of advances in the field, *just prior to performing the data analysis.*

E. Interpretation

Methodological conclusions.

Results of this review and meta-analysis will be of value in determining whether and how to apply ultrasound as an initial test. Its exact role will hinge on the diagnostic test performance established in the review and meta-analysis (Bossuyt et al, 2006);

a) If ultrasound is highly specific ***or*** sensitive for consolidation, it could serve to rule-in or rule-out the condition, terminating the diagnostic pathway as a *triage test*.
b) If ultrasound is more sensitive ***and*** specific than chest X-ray to diagnose consolidation, it could serve as a *replacement test*.
c) If ultrasound is more sensitive ***but*** less specific than chest X-ray to diagnose consolidation, it could serve as an *add-on test* if the initial chest X-ray is negative.

Clinical conclusions.

The actual values of sensitivity and specificity are of critical importance. Patients with consolidation detected on ultrasound are likely to be given empirical treatment for pneumonia. Patients without consolidation on ultrasound are likely to undergo testing for alternative diagnoses.

The consequences of a false negative result are therefore significant. If consolidation is missed, a potentially treatable condition (ie pneumonia) may go untreated. However, the consequences of a false positive result are equally significant. Reporting the presence of consolidation when it is

actually absent may mislead the clinician into premature diagnostic closure, with subsequent failure to consider alternative diagnoses, including the true diagnosis.

Subgroup analyses and meta-regression may identify important patient characteristics or test practices which influence the diagnostic accuracy of ultrasound in this scenario.

Finally, practical considerations (cost and availability) also have a bearing on the choice of testing.

F. Dissemination.

The results will be published in a critical care or respiratory medicine peer-reviewed journal preferably with an open-access option to allow wide dissemination.

The published document will be reported according to PRISMA guidelines (Preferred Reporting Items for systematic reviews and Meta-Analyses, Moher et al, 2009).

G. Logistics.

Registration.

To reduce the chance of duplication or redundancy, this Review Protocol will be registered at PROSPERO, an international registry of systematic reviews (Booth et al, 2012).

Review team.

a) Reviewer 1. Mark Hew, Respiratory Physician, Alfred Hospital, Melbourne, Australia.
b) Reviewer 2. John Corcoran, Clinical Research Fellow, Churchill Hospital, Headington, UK.
c) Reviewer 3. Najib Rahman, Director of Oxford Respiratory Trials Unit, Churchill Hospital, UK.
d) Biomedical Statistician. TBA.
e) Health Care Librarian with Searching expertise. TBA.

Timeline (Table 3).

Protocol development	month 1-2	June-July 2013
Literature search	month 3	August 2013
Relevance screening/inclusion assessment	month 4	September 2013
Data extraction & quality assessment	month 5	October 2013
Systematic review & meta-analysis	month 6	November 2013
Submission for publication	month 7	December 2013

H. REFERENCES.

Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev. 2012 Feb 9;1:2. doi: 10.1186/2046-4053-1-2.

Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089–92.

Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: Searching for Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]. The Cochrane Collaboration, 2008.

Doebler P, Holling H. Meta-Analysis of Diagnostic Accuracy with mada. R-Project 2012. Accessed 19 June 2013 at http://cran.r-project.org/web/packages/mada/vignettes/mada.pdf

Gagnier JJ, Bombardier C, Boon H, Moher D, Beyene J: An empirical study using permutation-based resampling in meta-regression. Systematic Reviews 2012, 1:18.

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007; 8: 239-251.

Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata Journal 2009; 9: 211-229.

Henschke CI, Yankelevitz DF, Wand A, Davis SD, Shiau M. Accuracy and efficacy of chest radiography in the intensive care unit. Radiol Clin North Am. 1996;34(1):21-31.

Hew M, Heinze S. Chest Ultrasound in Practice: a review of utility in the clinical setting. Intern Med J. 2012 Aug;42(8):856-65.

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008 Dec 16;149(12):889-97.

Linnet K, Bossuyt PMM, Moons KGM, Reitsma JB. Quantifying the Accuracy of a Diagnostic Test or Marker. Clinical Chemistry 2012. 58:9: 1292–1301.

Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. J Clin Epidemiol 2004; 57: 925-932.

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Mirvis SE, Tobin KD, Kostrubiak I, Belzberg H. Thoracic CT in detecting occult disease in critically ill patients. AJR Am J Roentgenol. 1987;148(4):685-689.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097.

Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993; 12: 1293-1316.

Ovenfors C, Hedgecock MW. Intensive care unit radiology: problems of interpretation. Radiol Clin North Am 1978:16:407-439

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982-990.

Reitsma JB, Moons KGM, Bossuyt PM, Linnet K. Systematic Reviews of Studies Quantifying the Accuracy of Diagnostic Tests and Markers. Clin Chem. 2012 Nov;58(11):1534-45.

Rubinowitz AN, Siegel MD, Tocino I. Thoracic imaging in the ICU. Crit Care Clin. 2007 Jul;23(3):539-73.

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001; 20: 2865-2884.

Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein DA, Mathis G, Kirkpatrick AW, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med. 2012 Apr;38(4):577-91.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.