PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Accuracy of Lung Ultrasonography in the Hands of Non-imaging Specialists to Diagnose and Assess the Severity of Community-Acquired Pneumonia in Adults: A Systematic Review
AUTHORS	Strøm, Julie; Haugen, Pia; Hansen, Malene; Graumann, Ole; Jensen, Martin Bach; Aakjær Andersen, Camilla

VERSION 1 – REVIEW

REVIEWER	James W Tsung MD, MPH Icahn School of Medicine at Mount Sinai New York, New York, USA
REVIEW RETURNED	24-Dec-2019

GENERAL COMMENTS	Lung ultrasound for community acquired pneumonia is a worthy topic to aggregate published literature periodically and perform a systematic review; the last systematic review that I can find was performed by Chavez M et al in Resp Research 2014. This topic is within the scope of a general medical journal such as BMJ Open. Introduction Page 2, line 43: based on prior meta-analyses (Chavez et al 2014) which show that lung US generally has excellent test characteristics this would suggest that LUS IS NOT A highly user dependent examination, unless specifically analyzing novices. Reference 14 by Brandli does not appear to be data that supports LUS as highly user dependent; the weight of the evidence seems to point to the contrary. The systematic review methodology appears appropriately followed and complete. In looking at the selected studies, I believe there is a significant omission in the paper by Lichtenstein et al. Ultrasound diagnosis of alveolar consolidation in the critically ill. Intensive Care Med 2014. Although these are ICU patients that had lung ultrasound by experienced sonologists they were not specifically nosocomial or ventilator associated pneumonia thus this data should be considered. Results: For your point estimates regarding test characteristic (sensitivity, specificity, kappa) kindly include 95% confidence intervals. It is understandable that you have not proceeded to pool data and perform meta-analyses as I generally agree that there is too much

heterogeneity with respect to study setting, patient population, and other technical factors such as probe used (curvilinear, phased, linear) and scanning technique (6 zone, vs 8 zone vs protocols including posterior thorax scanning). Perhaps this could be reiterated again in your limitations section.

Table 1. Diagnostic accuracy...

Please add 95% confidence intervals to your sensitivity and specificity values. Another column for useful statistic is "number of caPNA by reference study/N total study population" which will yield prevalence to help readers quickly assess study context (high prevalence likely admitted hospital patients vs low prevalence, likely outpatient clinic or ED patients).

Discussion:

Page 7, Line 56: re: "LUS to diagnose CAP is compromised by its inabililty to visualized pumonary lesions that are not in contact with the pleura." All this is true, it appears to happen very infrequently-- according to Lichtenstein et al Intensive Care Med 2014 (the paper that was not included or deselected from your such) this occurred in only 1.5% of cases of lung consolidation.

Page 8, Line 19: In general from pediatric data the learning curve appears steep, and substitution of chest X-ray with LUS is feasible in safe in a randomized controlled trial by Jones et al. Feasibility and Safety of Substituting Lung Ultrasonography for Chest Radiography When Diagnosing Pneumonia in Children: A Randomized Controlled Trial. Chest 2016

REVIEWER	Venkatakrishna Rajajee, MD
	University of Michigan, Ann Arbor, Michigan, USA
REVIEW RETURNED	06-Feb-2020

GENERAL COMMENTS	Title: Accuracy of Lung Ultrasonography in the Hands of Nonimaging Specialists to Diagnose and Assess the Severity of Community-Acquired Pneumonia in Adults: A Systematic Review
	Manuscript ID: bmjopen-2019-036067
	Summary: This is a systematic review of the use of lung ultrasound to diagnose community acquired pneumonia. Heterogeneity in the studies prevented formal meta-analysis. The authors performed a narrative systematic review and concluded that LUS in the hands of non-imaging specialists physicians working clinically has high accuracy in diagnosing pneumonia in adults.
	Assessment: This is a well written systematic review. The search criteria for publications is clearly stated, including key inclusion and exclusion criteria. Test positivity criteria in each study for the reference standard is mentioned. At least 2 authors reviewed each publication at both the abstract and full-text levels. Article authors were contacted for additional data. An established risk of bias tool (QUADAS-2) was used. Significant heterogeneity was recognized through a subjective assessment and, appropriately, a meta-analysis was not performed. A narrative review was performed.
	I do have some important concerns that need to be addressed-
	1. There is insufficient detail on what constituted test positivity for the index test (lung ultrasound) across studies. This is critical. Did the study look for consolidation, an A/B pattern, a B-profile without

sliding? Did subpleural consolidations count? This must be listed for the studies, in the same way the authors list test-positivity criteria for the reference standard.

2. Another crucial detail to list across studies is the nature and extent of blinding of individuals performing and interpreting the index test, to the refence standard. Was this reported? If so was blinding performed?

3. Although this is not a meta-analysis, I would perform a sensitivity analysis, using a purely narrative/ subjective approach. For example, what is the accuracy when looking at only the studies of the highest methodological quality, which I would define as prospective, with adequate blinding, clearly defined and appropriate test-positivity criteria for the index test, with the most objective criterion for positivity of the reference test (i.e, only studies that used CT alone?).

4. Most importantly, please moderate the conclusion. The studies were very heterogenous, small in size, mostly single center, often retrospective, and I suspect blinding was incomplete or absent in some studies. Also, there is at least one study that reported a sensitivity of 68%. I would therefore state that there is significant heterogeneity across studies, that the majority of studies report high accuracy, but that because of the small size, problems with methodology and heterogeneity, there is a need for larger prospective studies with clearly established criteria for diagnosis and complete blinding.

VERSION 1 – AUTHOR RESPONSE

BMJ Open	
Suggestions of Reviewers 1 and 2	
Reviewer #1: Introduction Page 2, line 43: based on prior meta-analyses (Chavez et al 2014) which show that lung US generally has excellent test characteristics this would suggest that LUS IS NOT A highly user dependent examination, unless specifically analyzing novices. Reference 14 by Brandli does not appear to be data that supports LUS as highly user dependent; the weight of the evidence seems to point to the contrary.	Thank you for bringing this to our attention. We have rephrased the statement to refer to ultrasound in general as a user-dependent imaging modality and now, with reference to the "European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB)" Guideline "The Minimum Training Recommendations for the Practice of Medical Ultrasound": *"None of the existing literature, however, differentiates between LUS operators despite the fact that ultrasound generally is considered a highly user-dependent imaging modality"*

Reviewer #1: The systematic review methodology appears appropriately followed and complete. In looking at the selected studies, I believe there is a significant omission in the paper by Lichtenstein et al. Ultrasound diagnosis of alveolar consolidation in the critically ill. Intensive Care Med 2014. Although these are ICU patients that had lung ultrasound by experienced sonologists they were not specifically nosocomial or ventilator associated pneumonia thus this data should be considered.	We agree that the paper by Lichtenstein et al. "Ultrasound diagnosis of alveolar consolidation in the critically ill" Intensive Care Med, is a paper of great interest – unfortunately it does not meet our inclusion criteria: "to describe the use of LUS for diagnosing CAP". Lichtenstein et al. perform LUS to visualize alveolar consolidation, which is seen in numerous pulmonary conditions and not only in patients with pneumonia. Lichtenstein et al. describe that "the consolidations were seen in a setting of ARDS, infectious pneumonia, pulmonary embolism, trauma, cardiogenic pulmonary edema, surgery, unknown origin and miscellaneous." Hence, the indication for performing LUS in the paper by Lichtenstein et al. is to visualize the alveolar consolidation that was found on CT and not on the indication of suspected pneumonia and specifically to diagnose pneumonia as our inclusion criteria says. However, it is a paper of importance and we have now included it in the discussion regarding pulmonary lesions not in contact with the pleura (see later).
Reviewer #1: Results: For your point estimates regarding test characteristic (sensitivity, specificity, kappa) kindly include 95% confidence intervals.	Thank you for pointing this out. We have now included 95% confidence intervals where data for this was available.
Reviewer #1: It is understandable that you have not proceeded to pool data and perform meta-analyses as I generally agree that there is too much heterogeneity with respect to study setting, patient population, and other technical factors such as probe used (curvilinear, phased, linear) and scanning technique (6 zone, vs 8 zone vs protocols including posterior thorax scanning). Perhaps this could be reiterated again in your limitations section.	Thank you for this comment. We agree and have added the following sentence to the limitations section: *"Due to the significant heterogeneity across studies, it was not possible to pool data and perform a meta-analyses."*

Reviewer #1: Table 1. Diagnostic accuracy Please add 95% confidence intervals to your sensitivity and specificity values. Another column for useful statistic is "number of caPNA by reference study/N total study population" which will yield prevalence to help readers quickly assess study context (high prevalence likely admitted hospital patients vs low prevalence, likely outpatient clinic or ED patients).	Thank you, we have now included 95% confidence intervals where data for this was available. We agree, that another column will enhance transparency. We have added a column to Table 1 with the following heading: *"Pneumonia positive (n) / Total number of patients examined for pneumonia (N)"*
Reviewer #1: Discussion: Page 7, Line 56: re: "LUS to diagnose CAP is compromised by its inabililty to visualize pulmonary lesions that are not in contact with the pleura." All this is true, it appears to happen very infrequently--according to Lichtenstein et al Intensive Care Med 2014 (the paper that was not included or deselected from your such) this occurred in only 1.5% of cases of lung consolidation.	Excellent point. We have added: *"However, according to Lichtenstein et al. who looked for lung consolidation in intensive care patients, this occurred in only 1.5% cases of lung consolidation."*
Reviewer #1: Page 8, Line 19: In general, from pediatric data the learning curve appears steep, and substitution of chest X-ray with LUS is feasible in safe in a randomized controlled trial by Jones et al. Feasibility and Safety of Substituting Lung Ultrasonography for Chest Radiography When Diagnosing Pneumonia in Children: A Randomized Controlled Trial. Chest 2016	Thank you for pointing this out. In line with your comment we have added the following to the manuscript: *"Though, the learning curve appears steep from pediatric data and in a randomized controlled trial by Jones et al. they found that substitution of CXR with LUS when evaluating children suspected of having pneumonia was feasible and safe, also in the hands of novice sonographers (≤ 25 examinations)."*
Reviewer #2: Assessment: This is a well written systematic review. The search criteria for publications is clearly stated, including key inclusion and exclusion criteria. Test positivity criteria in each	Thank you for the compliment.

study for the reference standard is mentioned. At least 2 authors reviewed each publication at both the abstract and full-text levels. Article authors were contacted for additional data. An established risk of bias tool (QUADAS-2) was used. Significant heterogeneity was recognized through a subjective assessment and, appropriately, a meta-analysis was not performed. A narrative review was performed.	
Reviewer #2: I do have some important concerns that need to be addressed- 1) There is insufficient detail on what constituted test positivity for the index test (lung ultrasound) across studies. This is critical. Did the study look for consolidation, an A/B pattern, a B-profile without sliding? Did subpleural consolidations count? This must be listed for the studies, in the same way the authors list test-positivity criteria for the reference standard.	Thank you for pointing this out. Due to space limit in the tables included in the manuscript, we have added a table called *"e-Table 2. Procedure and characteristics of LUS"* as supplemental material and refer to this in study characteristics: *"Definition of pneumonia based on LUS were divergent. Still, presence of subpleural or alveolar consolidation or a tissue-like lesion was part of the definition in all studies except one, in which no definition was described. The physicians performing and interpreting LUS were generally blinded to the reference standard; however, in four studies this matter was unclear. The definitions of pneumonia, blinding and scanning procedure and characteristics of LUS are listed in e-Table 2."*
Reviewer #2: 2) Another crucial detail to list across studies is the nature and extent of blinding of individuals performing and interpreting the index test, to the reference standard. Was this reported? If so was blinding performed?	Thank you for commenting on this. Whether the individuals performing and interpreting the index test were blinded to the reference standard was part of the methodological assessment and manifest itself as a high or unclear risk of bias regarding the index test in "e-Table 3. QUADAS-2 quality assessment". However, due to the nature of QUADAS-2, we agree, that this is not transparent to the readers. We have now listed this in *"e-Table 2. Procedure and characteristics of LUS"* and refer to this in study characteristics:

	"Definition of pneumonia based on LUS were divergent. Still, presence of subpleural or alveolar consolidation or a tissue-like lesion was part of the definition in all studies except one, in which no definition was described. The physicians performing and interpreting LUS were generally blinded to the reference standard; however, in four studies this matter was unclear. The definitions of pneumonia, blinding and scanning procedure and characteristics of LUS are listed in e-Table 2."
Reviewer #2: 3) Although this is not a meta-analysis, I would perform a sensitivity analysis, using a purely narrative/ subjective approach. For example, what is the accuracy when looking at only the studies of the highest methodological quality, which I would define as prospective, with adequate blinding, clearly defined and appropriate test-positivity criteria for the index test, with the most objective criterion for positivity of the reference test (i.e, only studies that used CT alone?).	Thank you for this proposal. In the results on *"Diagnostic accuracy of LUS"* we have added: *"The studies by Liu et al. and Amatya et al. were the two studies of highest methodological quality (e-Table 3). Both studies compared LUS to CT (Table 1) and LUS was performed by emergency physicians whose prior experience and training was described (Table 2). However, they differed with regards to the procedure and characteristics of LUS in terms of areas examined and definition of pneumonia on LUS. They found sensitivities of respectively 0.95 (95% CI, 0.89-0.98) and 0.91 (95% CI, 0.78-0.98) and specificities of 0.99 (95% CI, 0.92-1.00) and 0.61 (95% CI, 0.36-0.83)."* And in the Discussion: *"We highlighted the results of Liu et al. and Amatya et al. due to the quality of the studies, still, the studies were not completely comparable in other parameters. Both studies found high and comparable sensitivities of 0.95 (95% CI, 0.89-0.98) and 0.91 (95% CI, 0.78-0.98) respectively. However, in Amatya et al., LUS specificity was 0.61 (95% CI, 0.36-0.83) and significantly lower than the specificity in Liu et al. of 0.99 (95% CI, 0.92-1.00). According to Amatya et al., this was due to a higher prevalence of pulmonal co-*

	morbidities which resulted in false positive LUS results. Low specificity may lead to over-diagnosis of pneumonia and inappropriate use of antibiotics."
Reviewer #2: 4) Most importantly, please moderate the conclusion. The studies were very heterogenous, small in size, mostly single center, often retrospective, and I suspect blinding was incomplete or absent in some studies. Also, there is at least one study that reported a sensitivity of 68%. I would therefore state that there is significant heterogeneity across studies, that the majority of studies report high accuracy, but that because of the small size, problems with methodology and heterogeneity, there is a need for larger prospective studies with clearly established criteria for diagnosis and complete blinding.	Thank you for pointing this out. The conclusion now states: *"We found significant heterogeneity across studies. In the majority of studies, LUS in the hands of the non-imaging specialists demonstrated high sensitivities and specificities in diagnosing pneumonia. However, due to problems with methodology and heterogeneity there is a need for larger studies with uniform and clearly established criteria for diagnosis and blinding."*

VERSION 2 – REVIEW

REVIEWER	James W. Tsung MD, MPH Icahn School of Medicine at Mount Sinai
REVIEW RETURNED	25-Mar-2020

GENERAL COMMENTS	This manuscript has been revised and is acceptable for publication. In the last sentence of the limitations, I would recommend changing the "not possible" to "not appropriate." There have been prior meta-analyses that inappropriately pooled the data, just as you could, but due to the heterogeneity you identified such a result would be meaningless thus "not appropriate."

REVIEWER	Venkatakrishna Rajajee University of Michigan, Ann Arbor, Michigan, USA
REVIEW RETURNED	17-Mar-2020

GENERAL COMMENTS	The authors have addressed all of my concerns.

VERSION 2 – AUTHOR RESPONSE

Thank you for the comments and minor revision.

In the last sentence of Limitations, "not possible" has been changed to "not appropriate", as we fully agree with reviewer 1 on the comment.