Lung ultrasonography versus chest radiography for the diagnosis of pediatric community acquired pneumonia in emergency department: a meta-analysis

Liang Wang¹, Wei Song², Yong Wang³, Jie Han³, Ke Lv¹

¹Department of Ultrasound, ²Department of Radiology, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, Beijing 100730, China; ³Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

Contributions: (I) Conception and design: K Lv, L Wang; (II) Administrative support: K Lv, L Wang; (III) Provision of study materials or patients: K Lv, L Wang; (IV) Collection and assembly of data: K Lv, L Wang; (V) Data analysis and interpretation: K Lv, L Wang, W Song, Y Wang, J Han; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Ke Lv. Department of Ultrasound, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, 1 Shuaifuyuan Wangfujing, Beijing 100730, China. Email: lvke@163.com.

Introduction

Pediatric community acquired pneumonia (pCAP) is a major public health and economic problem with a considerable impact on morbidity and mortality in children (1). According to the World Health Organization, pCAP is the most common infectious cause of death among children worldwide (2). It is estimated that pCAP accounts for 18% of the total number of deaths in children <5 years, more...
than tuberculosis, acquired immunodeficiency syndrome (AIDS), and malaria combined (3).

While pCAP is considered a clinical diagnosis, radiologic confirmation is often obtained. In clinical practice chest radiography (CR) may be requested by physicians in terms of diagnosis, differential diagnosis and complications (4). Therefore, despite the international guidelines, CR is frequently performed for children with suspected pCAP, thus increasing unnecessary pediatric radiation exposure (5).

Since 1986, many researchers have investigated the use of lung ultrasonography (LU) in evaluating pneumonia, first in the adult and recently in pediatric population (6-8). Characteristics such as being rapid, mobile, repeatable and easily performed at the bedside and without radiation exposure make LU particularly valuable. Recently many studies and meta-analyses have shown promising results on the accuracy of LU in diagnosing pneumonia and potentially replacing CR in pediatric population (3,8-10).

In the literature, previous studies establishing the accuracy of LU often used CR as reference standard and took into account different clinical settings all together. However, it is methodologically questionable to include studies with different clinical settings to establish the accuracy of LU in diagnosing pCAP. Emergency department (ED) is a specific clinical setting with limited medical resource relative to hospitalized patients in intensive care unit (ICU) or medical ward. Meanwhile, CR is not a sufficiently reliable reference standard in diagnosing pCAP due to its high inter-observer variability and relatively low sensitivity (71–87%) (11,12). Therefore, to make a more objective and specific analysis, our study is aimed to compare the accuracy of LU to CR in diagnosing pCAP in ED setting through a systematic review and a meta-analysis.

Methods

Search strategy

A systematic literature search was applied to PubMed (1946 to present) and Embase (1974 to present) databases. A combination of a controlled vocabulary of keywords around ‘pneumonia’, ‘ultrasonography’, ‘radiography’ and ‘emergency’ was used (more details in supplementary appendix online). The search of studies was not limited on the basis of publication dates. Unpublished researches or conference communications were not considered. Only articles written in English were considered. All relevant titles and abstracts were searched for full text by two authors (Wang and Lv) independent of each other. References from selected studies and review were evaluated manually to identify any further relevant study for analysis. The literature search and data analysis were carried out in December 2018.

Study selection

The inclusion criteria were as follows: (I) pediatric patients aged <18 years with clinically suspected pCAP in ED setting; (II) inclusion of both LU and CR in diagnostic work-up; (III) expert pediatrician clinical diagnosis as gold standard for diagnosis of pCAP; (IV) sufficient data to calculate the true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) values. Editorials, letters, comments, reviews, conference abstract, case reports and animal experimental studies were excluded. Two reviewers (Wang and Lv) independently evaluated the records. Their results were compared and any discrepancy was discussed until consensus was reached.

Data extraction

The data collected included: year of publication, the country in which the study was conducted, study design, reference standard, sample size, patient number, age distribution, gender distribution, CR plans, LU probe, LU criteria for pCAP, LU sonographer, and blinding. Furthermore, the values of TP, TN, FP and FN were extracted. If studies did not provide data to directly construct 2x2 contingency tables, they were calculated based on the diagnostic sensitivity and specificity provided in the studies.

Quality assessment

Two reviewers (Wang and Lv) independently assessed the methodological quality of the selected studies using the Quality Assessment of Studies of Diagnostic Accuracy (QUADAS-2) tool. The QUADAS-2 tool consists of four key domains: patient selection, index test, reference standard and flow and timing. Each domain is assessed in terms of risk of bias (low, high or unclear), and the first three domains relate to concerns regarding applicability. Disagreements in the evaluation of study quality between the two reviewers were resolved through discussion.

Statistical analysis

Metadisc1.4 software was used for statistical analysis. The
statistical significance was set at P<0.05. (I) Diagnostic performance. From each study, pooled sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratios with their 95% confidence interval (95% CI) were calculated. And the summary receiver operating characteristic (SROC) curves and area under the curve (AUC) were also analyzed. (II) Heterogeneity. First, we explored the heterogeneity caused by the threshold effect; if there is a strong positive correlation in Spearman’s correlation coefficient between the logit of sensitivity and logit of 1-specialty, threshold effects were considered to exist. Secondly, we explored and quantified the heterogeneity (apart from the threshold effect) by means of the inconsistency index (I²) test and chi-squared value test. If the selected studies had enough consistency (I²<50% and Q>0.05), the fixed effect model was used for analysis. If I²>50% or Q<0.05 suggested the existence of heterogeneity. Then, we explored the heterogeneity source by meta-regression techniques. If no obvious reason was found in heterogeneity, statistics (sensitivity, specificity, positive and negative likelihood ratios, and diagnostic odds ratio) pooling would adopt the random effect model. And the sensitivity analysis was conducted to verify the stability of the results. (III) Publication bias. The risk of publication bias was assessed across studies with a funnel plot of the studies’ logarithm of the diagnostic odds ratios against the standard error (SE) when more than nine studies were included.

Results

Search strategy and study selection

According to the search strategy, a total of 772 articles were retrieved, of which 56 articles were discarded due to duplication. According to the inclusion criteria, 689 items were excluded after screening the title and abstract. The remaining 27 articles were evaluated by reading the full text, of which 21 were further excluded. Eventually, 6 studies were reviewed (2,13-17). No additional study was included after screening the reference lists of included studies. Figure 1 showed the flowchart summarizing the process used to select the studies finally included.

Study characteristics and quality assessment

Among the 6 studies included, a total of 701 patients were included, of which pCAP was diagnosed in 575 (82.0%) cases. Details on study characteristics are summarized in Table 1. The QUADAS-2 tool indicated that the quality of included studies was generally fair (Table 2). Reference standard domain, however, was labeled as unclear in 4/6 studies. Because the criteria for pCAP diagnosis were not clearly reported. More importantly, in these studies CR was probably included as part of the diagnostic criteria for pCAP in addition to clinical presentation and clinical course. Therefore, the suboptimal quality of included studies could introduce a source of heterogeneity into the present meta-analysis.

Meta-analysis of diagnostic performance

By random-effects model meta-analysis of the 6 studies, the pooled sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratios of LU were 96.7% (95% CI, 94.9–98.0%), 87.3% (95% CI, 80.2–92.6%), 8.10 (95% CI, 2.30–28.47), 0.05 (95% CI, 0.02–0.10), 256.68 (95% CI, 35.39–1862.00), while they were 89.7% (95% CI, 87.0–92.1%), 93.7% (95% CI, 87.9–97.2%), 9.97 (95% CI, 5.11–19.45), 0.12 (95% CI, 0.08–
The AUC of SROC curves in diagnosing pCAP were 0.99 (95% CI, 0.98–1.00) and 0.97 (95% CI, 0.95–1.00) for LU and CR, respectively (Figure 2). Forest plots of sensitivity and specificity with 95% CI for LU and CR were provided in Figure 3.

**Heterogeneity tests**

Significant threshold effect was found for CR (P=0.000) but not for LU (P=0.397). Meanwhile, the I² values of sensitivity and specificity for both LU (40.7% and 80.7%) and CR (64.6% and 58.0%) were all close to or above 50% indicating substantial heterogeneity among studies.

Meta-regression analysis was then performed to explore the potential sources of heterogeneity. It included study design (prospective vs. retrospective), sample size of patients (>100 vs. <100), reference standard (ex-post diagnosis vs. non-ex-post diagnosis), LU sonographer (pediatrician vs. radiologist) and LU criteria for pCAP [consolidation vs. consolidation + B-line artifact (BLA)]. Meta-regression analysis failed to provide evidence of heterogeneity related to the potential effects of all above five confounding covariates.

Then, sensitivity analysis was used to estimate the potential sources of heterogeneity. The sensitivity analysis revealed that the pooled sensitivity, specificity and SROC showed little change compared with the previous results when excluding the 6 studies one by one.

**Publication bias**

There were only 6 studies included in our study. Thus, we did not show a funnel plot on account of the Cochrane handbook suggesting that funnel plots should only be used to assess the publication bias when more than nine are included.

**Discussion**

To our knowledge, this is the first meta-analysis that takes into account specifically the ED setting to compare the diagnostic accuracy of LU and CR for pCAP. In fact, there is a methodological concern on the inclusion of studies carried out in the medical ward or ICU to evaluate the diagnostic accuracy of an imaging modality in diagnosing pCAP. Because theoretically pediatric pneumonia that occurs in hospitalized patients may be partially attributed...
Table 2 Quality assessment of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk of bias</th>
<th>Applicability concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient selection</td>
<td>Index test</td>
</tr>
<tr>
<td>Copetti, 2008</td>
<td>Unclear</td>
<td>Low</td>
</tr>
<tr>
<td>Guerra, 2016</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Ianniello, 2016</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Boursiani, 2017</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Yilmaz, 2017</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Biagi, 2018</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

Figure 2 SROC curves of (A) LU and (B) CR for diagnosing pCAP. SROC, summary receiver operating characteristic; LU, lung ultrasonography; CR, chest radiography; pCAP, pediatric community acquired pneumonia; AUC, area under the curve; SE, standard error; Q*, Q index (the point in receiver operating characteristic curve space closest to the ideal top left-hand corner and where test sensitivity and specificity are equal).

Figure 3 Forest plots of sensitivity and specificity of included studies for (A) LU and (B) CR. LU, lung ultrasonography; CR, chest radiography; CI, confidence interval.
to hospital acquired pneumonia or induced ventilator-associated pneumonia rather than pCAP.

Compared with previous meta-analyses reported in the literature, our study is more focused on our clinical question. Despite this, we obtained a similar AUC value of LU (0.99) in diagnosing CAP compared with previous meta-analyses [0.97 in Orso et al. (10), 0.97 in Ye et al. (7), 0.98 in Orso et al. (9), and 0.98 in Pereda et al. (3)]. These data testify to the usefulness of the LU in diagnosing pCAP, despite heterogeneous populations.

Our meta-analysis showed that LU had much higher sensitivity in diagnosing pCAP compared with CR (96.7% vs. 89.7%). This finding was consistent with the results of a meta-analysis in adult CAP population, in which LU had a pooled sensitivity of 0.95 while CR had a pooled sensitivity of 0.77 (7). This difference of sensitivity between LU and CR in diagnosing pCAP is attributable to the superiority of LU in detecting lung consolidation. Lung consolidation is the typical feature of CAP on both LU and CR, which is defined as a hypoechogenic area with poorly defined borders and compact underlying comet tail artifacts on LU (13). Shah et al. found that LU is more sensitive than CR in detection of small lung consolidation (18). In their study, all 13 sub-centimeter lung consolidations detected by LU were found to be negative on CR. Similarly, in an adult study, the authors compared the accuracy of LU and CR in the detection of lung consolidation, taking computed tomography (CT) scan as the gold standard. They demonstrated that the sensitivity of LU (81.4%) was significantly higher than that of CR (64.3%) (19).

Although the detection of sub-centimeter lung consolidation significantly improved the sensitivity, it may have an opposite impact on specificity of LU in diagnosing pCAP. Because the pathological relevance of sub-centimeter lung consolidation is uncertain (20). The differential diagnosis list included bronchiolitis and bacterial/viral pneumonia (21). In the study of Biagi et al. LU was applied for the diagnosis of pneumonia in children with acute bronchiolitis (17). They found that 9/10 LU-FP cases were misdiagnosed as sub-centimetric pneumonia by LU. When only consolidation >1 cm was considered consistent with pneumonia, the specificity of LU increased from 83.9% to 98.4%, while the sensitivity decreased from 100% to 80%. Therefore, regarding children with sub-centimeter consolidations on LU, the authors suggested careful clinical and LU follow-up to discriminate patients who will need antibiotic treatment from those who will achieve a spontaneous resolution.

Apart from lung consolidation, BLA was another LU finding diagnostic for pCAP, which was used as diagnostic criterion in isolation from consolidation in 2 studies in our meta-analysis. However, the specificity of BLA in diagnosing pCAP is questionable. Our meta-analysis showed that LU had lower specificity in diagnosing pCAP compared with CR (87.3% vs. 93.7%). This difference was mainly attributable to the low specificity (45.5%) of LU reported by the study of Yilmaz et al. included in our meta-analysis. The authors found that increased BLA and/or pleural irregularity were determined in all six FP cases (4 children with upper airway infections and 2 children with pleural edema). In fact, when BLA is observed in the lung, it represents a sign of increased density due to the loss of aeration in the lung periphery (22). This condition may indicate different diseases including cardiogenic pulmonary edema, diffuse or focal interstitial lung disease, pneumonia and acute respiratory distress syndrome (23). Correct interpretation of BLA in LU is strongly influenced by associated sonographic signs and careful integration of all relevant clinical information. In fact, further studies are still needed regarding the specificity of particular lung ultrasound (LUS) findings for pCAP rather than large consolidation, such as isolated BLA and sub-centimeter consolidations.

Operator dependency is one of the limitations often cited with regard to the ultrasound imaging study. One study in our meta-analysis evaluated the inter-observer agreement between the experienced radiologist and the less experienced pediatrician who independently performed LU. The authors reported that there was an almost perfect agreement between pediatrician and radiologist in diagnosing pCAP by LU in 30 patients (kappa =0.93) (17). There is evidence that LUS might be accurate even when performed by less experienced operators.

Finally, there is some heterogeneity in the literature regarding the reference standard of pCAP, which is also reflected in the studies included in our meta-analysis. Although CT scan is the golden standard for the diagnosis of CAP, it is not ethically justifiable to indiscriminately use CT in pediatric patients. In our meta-analysis, CT scan was only used in a small proportion of patients in 2/6 studies (4/79 in Copetti et al. and 4/222 in Guerra et al.) for the diagnosis of recurrent and complicated cases. On the other hand, as the alternative choice, exclusively clinical evaluation by attending physicians does not seem to be sufficiently accurate method. In our opinion, a possible solution could be the “ex-post diagnosis” method, in which
the final diagnosis is established by external physicians not involved in the clinical case by using all available data. However, this “ex-post diagnosis” method was only employed in 2/6 studies in our meta-analysis.

**Limitations**

Our study had some limitations. First, articles in languages other than English were not included in our study. Second, considerable heterogeneity was present in this meta-analysis. Third, the small number of studies included in our meta-analysis is another limitation. Fourth, in addition to accuracy, the cost-effectiveness of LU and CR is another significant concern for diagnosing pCAP in ED. Because ED is a specific clinical setting with limited medical resource. However, it is beyond the aim of our study, and further research is needed in this field. Finally, a major limitation of this study is the question of the reference standard. Four/six studies probably included CR as part of the diagnostic criteria for pCAP in addition to clinical presentation and clinical course. This likely skews the analysis of CR results towards higher specificity, which may be the underlying cause for the significant threshold effect found for CR in our meta-analysis (P=0.000). This is somewhat controlled by following the clinical course; however, it likely remains a large limitation.

**Conclusions**

Our meta-analysis suggests that LU is an accurate tool in the diagnosis of pCAP in the ED setting with a superior sensitivity over CR. However, the specificity of LU is lower than that of CR, which may be attributable to the unspecific LU finding of sub-centimeter consolidation or the inclusion of CR as part of the diagnostic criteria for pCAP. Therefore, further studies are still needed to evaluate the diagnostic accuracy of LU for pCAP without inclusion of CR as a portion of the reference standard, and the specificity of particular LU findings for pCAP, such as sub-centimeter consolidations and isolated BLA.

**Acknowledgments**

None.

**Footnote**

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Institutional review board approval was not required because this study is a meta-analysis in which all analyses were based on previous published studies, and approval is waived for this study type at our institution (Peking Union Medical College Hospital).

**References**

