Objectives: Chest radiograph is considered the first-line diagnostic imaging modality for patients presenting with pulmonary symptoms in the ICU. In this meta-analysis, we aim to evaluate the diagnostic accuracy of chest radiograph, and when concomitantly studied lung ultrasound, in comparison with the gold-standard CT for adult critically ill patients with respiratory symptoms.

Data Sources: PubMed, EMBASE, and Gray literature.

Study Selection: Studies comparing chest radiograph, and if performed lung ultrasound, with CT for adult ICU patients with respiratory symptoms.

Data Extraction: Quality was scored with Quality Assessment of Diagnostic Accuracy Studies-2, and study setting, test characteristics, and study design were extracted.

Data Synthesis: In the meta-analysis, we included 10 full-text studies, including 543 patients, and found that chest radiograph has an overall sensitivity of 49% (95% CI, 40–58%) and specificity of 92% (86–95%). In seven studies, where also lung ultrasound was studied, lung ultrasound had an overall sensitivity of 95% (92–96%) and specificity of 94% (90–97%). Substantial heterogeneity was found. A planned subgroup analysis for individual pathologies was performed. The results of four abstract-only studies, included in the systematic review, were considered unlikely to significantly influence results of our meta-analysis. Study limitations were that most studies were of low power combined with methodological limitations.

Conclusions: This meta-analysis demonstrates that chest radiograph has a low sensitivity and reasonable specificity compared with CT for detecting lung pathology in critically ill patients. The studies also investigating lung ultrasound, showed lung ultrasound to be clearly superior to chest radiograph in terms of sensitivity with similar specificity, thereby opting to be the first-line diagnostic tool in these patients. (Crit Care Med 2018; XX:00–00)

Key Words: chest radiograph; diagnostics; intensive care; respiratory symptoms; ultrasound
Multiple studies on the use of lung ultrasound (LUS) in the ICU and emergency department have been published. 

Especially in the ICU, when dealing with critically ill patients, LUS has potential advantages over CT and chest radiograph (8) as it can be performed at the bedside and without the described limitations of CT.

Despite the aforementioned, chest radiograph is still the first line of diagnostic chest imaging for patients with pulmonary symptoms in the ICU.

In this systematic review and meta-analysis, we aim to investigate the diagnostic accuracy of chest radiograph, and when concomitantly investigating LUS, compared with the gold-standard CT in critically ill adult patients with respiratory symptoms.

OBJECTIVES

Primary Objective
To evaluate the diagnostic accuracy of “chest radiograph” compared with CT in critically ill adult patients (≥ 18 yr) with respiratory symptoms.

Secondary Objectives
To evaluate the diagnostic accuracy of LUS compared with CT in critically ill patients with pulmonary symptoms, when also investigated in the included studies.

MATERIAL AND METHODS

Study Design
This is a systematic review and meta-analysis. To improve the quality of our systematic review, Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed. The protocol was registered: PROSPERO#CRD42016041448.

Selection of Studies

Studies. Randomized controlled trials, cross-sectional case control, and observational studies.

Population. Patients greater than or equal to 18 years admitted to the ICU, with pulmonary symptoms (diagnosis included are among others: pneumothorax, pleura effusion, pulmonary edema, pneumonia, atelectasis, acute respiratory distress syndrome [ARDS]), who have an indication for CT.

Index Test. Chest radiograph for the diagnosis of pulmonary symptoms; data on accuracy of LUS were also collected from studies that also investigated LUS.

Outcome. All data concerning diagnostic accuracy (sensitivity, specificity, positive predictive value [PPV], and/or negative predictive value [NPV]).

Reference Standard. CT.

Exclusion Criteria
1) Measurements: chest radiograph for other reason than cardiopulmonary pathology.

2) Outcome: unable to determine accuracy, only determination of interobserver agreement, no CT for comparison.

3) No abstract available.

4) Language: abstract not in English.

5) Articles with only an abstract, but no full-text available or full text in English were excluded from qualitative assessment and meta-analysis but were included in the systematic review for separate quantitative analysis when enough data were available regarding sensitivity, specificity, and diagnostic accuracy.

These exclusion criteria were modified after screening of the abstracts. Several abstracts were identified without (English) full text that provided sufficient data on accuracy. Some studies were excluded based on more than one exclusion criteria. These studies are mentioned in one of the exclusion groups.

Literature Search Strategy

After consulting a medicine literature search specialist, we searched EMBASE and PubMed for relevant articles up until April 2016. The selected articles were checked for backward and forward citations by hand search. We used Mendeley Software (Elsevier Inc, New York, NY). Finally, we searched for Gray literature at OpenGrey (www.opengrey.eu), Bielefeld Academic Search Engine (BASE) (www.base-search.net), and at the National Library of Medicine’s clinical trial registry (www.clinicaltrials.gov). Full search strategy is available in Appendix 1 (Supplemental Digital Content 1, http://links.lww.com/CCM/D436).

Data Extraction
Selection of studies and data extraction were performed by two independent reviewers (M.H.W., P.R.T.). Disagreement was resolved by consensus meetings with a third reviewer (H.R.T.).

Characteristics of Data Collected. Setting, time between index test and reference test, test characteristics, study design.

Results. 2 × 2 table, sensitivity, specificity, PPV, NPV. Results were categorized according to the diagnosis investigated.

Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 tool was used (9). Quality assessment was done by two independent reviewers (M.H.W., P.R.T.), and disagreement was resolved with a third reviewer (H.R.T.). Quality assessment was only performed on the full-text studies that were included in the meta-analysis.

Statistical Analysis and Data Synthesis
Pooled estimates of sensitivity and specificity were obtained by fitting a bivariate model on the raw study data. In the bivariate model, pairs of sensitivity and specificity are jointly analyzed, and the correlation that exists between these two measures obtained in a single study is then taken into account through the inclusion of a random effect for study in the model.

Results are presented with a 95% CI. The models were estimated in STATA Version 14. The Midas module for STATA (StataCorp LLC, College Station, TX). was used to make forest plots and to estimate heterogeneity. Hierarchical summary receiver operating characteristic (HSROC) curves for chest radiograph and LUS, respectively, were made.
RESULTS
Details regarding the study selection are presented in Figure 1. The search yielded 9,230 articles. We included 10 studies, involving 543 patients, that underwent qualitative assessment. Additionally, we included two of the evaluated 617 non-English articles studies (10, 11) with only their abstract written in English and two studies (12, 13) of which we were unable to get a full text, resulting in 323 extra patients.

In e-Table 1 (Supplemental Digital Content 2, http://links.lww.com/CCM/D437), study characteristics are summarized. The last four articles (10–13) are abstract-only studies. In two studies, it was unclear who evaluated the chest radiograph (14, 15), and in three, it was unclear how many reviewers evaluated the chest radiograph (14, 16, 17). Two studies selected patients retrospectively (16, 18). There were seven full text and three abstract only, who also evaluated the diagnostic accuracy of LUS compared with CT (5, 11–15, 17, 19–21).

Figure 2 shows a forest plot for the individual study and pooled results for the primary outcome, sensitivity and specificity of chest radiograph. Figure 3 shows a forest plot for the secondary outcome, sensitivity and specificity of LUS. Pooled results of chest radiograph and LUS accuracy for the different pathologies investigated are presented in Table 1. All individual study results for chest radiograph and LUS accuracy are presented in e-Table 2 (Supplemental Digital Content 3, http://links.lww.com/CCM/D438) and e-Table 3 (Supplemental Digital Content 4, http://links.lww.com/CCM/D439). HSROC curves showing study heterogeneity for both chest radiograph and LUS are presented in e-Figure 1 (Supplemental Digital Content 5, http://links.lww.com/CCM/D440; legend, Supplemental Digital Content 9, http://links.lww.com/CCM/D444) and e-Figure 2 (Supplemental Digital Content 6, http://links.lww.com/CCM/D441; legend, Supplemental Digital Content 9, http://links.lww.com/CCM/D444), respectively.

Consolidations
From Voggenreiter et al (22), data for specificity of chest radiograph for diagnosis of consolidation were excluded because only one observation was available for the estimation of the specificity. Sensitivity from same study was not mentioned in Figure 2 because specificity was missing and the forest plot procedure requires both to be filled in. Razazi et al (21) did not report raw data for consolidations and was therefore also excluded. In addition, from three abstract-only studies (11–13), including 291 patients, the ranges of chest radiograph sensitivity and specificity were 22–40% and 75–100%, respectively. LUS sensitivity and specificity were 32–100% and 87–100%, respectively.
We excluded 48 hours later (T2) data from Rocco et al (20) because they involved the same patients at ICU arrival (T1). Three abstract-only studies, including 245 patients, found a chest radiograph sensitivity and specificity which ranged from 34% to 69% and 54% to 100%, respectively. Two out of these three studies, including 213 patients, reported LUS accuracy with a sensitivity from 47% to 100% and specificity 82–100%.

Pneumothorax

There were not enough data to determine study heterogeneity. One additional abstract-only (13) study, which investigated the presence of pneumothorax in 200 patients, reported a chest radiograph sensitivity of 40% and specificity of 96% and for LUS a sensitivity and specificity of both 100%.

Interstitial Syndrome

Two additional, abstract-only, studies were included resulting in an extra 213 patients. Their range for chest radiograph sensitivity were 42–100% and for specificity 82–100%. LUS sensitivity and specificity were 50–95% and 83–95%, respectively.

Other Pulmonary Pathology

Lung contusion was investigated in two full-text studies (15, 20) including 65 patients. There was one study investigating patients with ARDS (18), including 90 patients. It did not consider LUS. Results from this study were used in the overall sensitivity and specificity of chest radiograph. The chest radiograph sensitivity was 73% (CI, 61–83%) with a specificity of 70% (CI, 47–87%).

Quality Assessment

The risk of bias and applicability concerns are summarized in Table 2.

Most studies lacked in some way information about blinding to the reference standard while evaluating the index results or the other way around. Sometimes, this was mentioned, but mostly, it was simply not presented in the study.
TABLE 1. The Pooled Estimates of Sensitivity and Specificity of Chest Radiograph and Lung Ultrasound Compared to CT For Each Pathology Investigated in Critically Ill Patients

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>CXR Pooled Sensitivity (95% CI) (%)</th>
<th>CXR Pooled Specificity (95% CI) (%)</th>
<th>LUS Pooled Sensitivity (95% CI) (%)</th>
<th>LUS Pooled Specificity (95% CI) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall accuracy</td>
<td>49 (40–58)</td>
<td>92 (86–95)</td>
<td>95 (92–96)</td>
<td>94 (90–97)</td>
</tr>
<tr>
<td>Consolidations: CXR: ns = 7, np = 349; LUS: ns = 5, np = 249</td>
<td>69 (49–84)</td>
<td>90 (75–96)</td>
<td>97 (91–99)</td>
<td>91 (79–96)</td>
</tr>
<tr>
<td>Pleural effusion: CXR: ns = 6, np = 276; LUS: ns = 4, np = 137</td>
<td>55 (42–66)</td>
<td>82 (73–89)</td>
<td>98 (87–100)</td>
<td>94 (79–99)</td>
</tr>
<tr>
<td>Pneumothorax: CXR: ns = 2, np = 81; LUS: ns = 1, np = 42</td>
<td>21 (42–66)</td>
<td>100 (0–100)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Interstitial syndrome: CXR: ns = 2, np = 74; LUS: ns = 2, np = 74</td>
<td>53 (35–69)</td>
<td>91 (75–97)</td>
<td>95 (83–99)</td>
<td>91 (75–97)</td>
</tr>
<tr>
<td>Lung contusion: CXR: ns = 2, np = 65; LUS: ns = 2, np = 65</td>
<td>38 (25–51)</td>
<td>90 (72–97)</td>
<td>98 (87–100)</td>
<td>90 (72–97)</td>
</tr>
</tbody>
</table>

CXR = chest radiograph, LUS = lung ultrasound, np = number of patients, ns = number of studies.

*The specificity by Voggenreiter et al (22) was excluded for consolidation. For pneumothorax, LUS sensitivity and specificity are left blank because there was only one study and there was not enough data to determine study heterogeneity.

Dashes indicate LUS sensitivity and specificity are left blank, because there was only one study and there was not enough data.
DISCUSSION

This meta-analysis on diagnostic accuracy of chest radiograph compared with CT in ICU patients with respiratory symptoms found that chest radiograph has an overall low sensitivity of 49% (CI, 40–58%) and a good specificity 92% (CI, 86–95%). The seven studies that also compared LUS with CT demonstrated that LUS was clearly superior to chest radiograph in terms of sensitivity 95% (CI, 92–96%), with similar specificity 94% (CI, 90–97%). The results of four abstract-only studies were considered unlikely to significantly influence the results of our meta-analysis because of similar characteristics in terms of study population and accuracy. All studies included had limitations in their design.

Our results contradict the widely held belief that chest radiograph should be the primary thoracic imaging modality in the ICU, considering its significantly low sensitivity for the investigated lung pathologies. Of course, not only test characteristics should influence a physician’s decision for diagnostic modality. Clinical suspicion and pretest probability will also guide the diagnostic process to rule in or out several conditions with high accuracy. Thereby, the potency of chest radiograph in ICU patients with respiratory symptoms to diagnose between different pulmonary conditions is further limited when the prevalence of a condition is low. Since prevalence of pulmonary pathology is lower in patients without symptoms, the additional value of chest radiograph is further impaired.

In stark contrast, LUS sensitivity was above 95% for all four lung pathologies, with a specificity for all pathologies of similar accuracy as chest radiograph. Therefore, LUS seems like a very good alternative in these patients also because of its bedside availability and fewer downsides over chest radiograph and CT.

Although the accuracy of chest radiograph for lung pathology is questioned in this meta-analysis, there are situations in which, in an ICU patient with pulmonary symptoms, it might be preferable to use chest radiograph instead of LUS. For example, chest radiograph is considered the gold standard for the detection of nasogastric tube and central venous catheter (miss) placement and complications (23, 24). One should note that LUS is operator dependent and can only be used to its full potential when the clinician masters the technique. By this, we emphasize the importance to adjust the choice of diagnostic modality on individual patient level. Furthermore, all modalities should be used in guidance with the clinic (e.g., history, laboratory results, and setting) and the ability of the physician to adequately use all the data to construct a diagnosis. CT has additional value compared with LUS in situations where higher resolution or quantification of abnormalities is required, for example, when to aid in difficult diagnosis or to advance our understanding of the pathogenesis and pathophysiology of a disease (25).

To our knowledge, this is the first systematic review on the diagnostic accuracy of chest radiograph and partly LUS, compared only with the gold-standard CT in patients admitted primarily to the ICU and with respiratory symptoms, in which a variety of common lung pathologies were evaluated. Ashton-Cleary (26) also explored the diagnostic performance of LUS and chest radiograph in a critical care population; however, they emphasized on LUS and compared it with different kinds of reference standards, including chest radiograph. We conducted an extensive search strategy and put emphasis on chest radiograph, with addition of LUS when concomitantly

### TABLE 2. Risk of Bias and Applicability Concerns Studies

<table>
<thead>
<tr>
<th>References</th>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient Selection</td>
<td>Index Test CXR</td>
</tr>
<tr>
<td>Lichtenstein et al (5)</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Xirouchaki et al (19)</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Kitazono et al (16)</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Figueroa-casas et al (18)</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Rocco et al (20)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Voggenreiter et al (22)</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Razazi et al (21)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+ = low risk, – = high risk, ? = uncertain risk, CXR = chest radiograph, NA = not available.
studied, since this is potentially the new first-line diagnostic tool in these patients.

As with all systematic reviews, our review was sensitive for publication bias. We could not identify the effect of publication bias on results of a meta-analysis on diagnostic accuracy in the literature, but in meta-analysis on treatment effect, the effect could be overstated by an average 12% (27). To minimize publication bias, we searched in multiple gray literature databases.

Our meta-analysis was subject to substantial heterogeneity for several reasons. First of all, we investigated various pathologies, and these were all used in the forest plot (Figs. 2 and 3) as if they were equal to one another. Therefore, we also performed individual sensitivity and specificity analyses for each pathology to reduce heterogeneity (Table 1). Previous meta-analyses on LUS accuracy also found substantial heterogeneity, due to heterogeneity in study population, setting, and reference standard (28–30), among others. Therefore, in our methods, we used strict exclusion criteria for these factors to reduce heterogeneity.

Second, our systematic review included a limited number of studies of which only two studies (14, 16) had a study population of 100 participants or more. A study investigating the effect of small trials (31) in 13 meta-analyses found that small studies tend to have a more beneficial treatment effect. The authors advise to be careful with the interpretation of small trials especially when they lack high methodological quality (32). The same may hold true for studies on diagnostic accuracy with limited number of patients and cases, where a false-negative or positive result can affect accuracy more strongly than with larger studies. In addition, the included studies had methodological shortcomings. Last, the number of radiologists who evaluated chest radiograph or physicians who performed LUS were different for many studies. A few studies only used one reader for chest radiograph and one performer for LUS. Most studies did not mention the operator’s degree of experience. The unclear and varying level of experience between the operators is a potential form of bias. The aforementioned makes it difficult to draw firm conclusions.

In this systematic review, we primarily focused on the diagnostic accuracy of chest radiograph and LUS versus the gold-standard CT in varying lung pathology for patients admitted to the ICU to assess its use in clinical practice. However, Bossuyt et al (31) stressed the importance of the clinical utility, ease of implementation, and cost-effectiveness next to the accuracy of a diagnostic tool when evaluating its overall performance. Indeed, we found a better accuracy of LUS compared with chest radiograph, but we, and most studies included, did not look at the above-mentioned factors, which might lead to incorrect conclusions about the usefulness of chest radiograph and LUS in clinical practice. One study looked at how diagnostic information influenced therapeutic consequences based on CT results. In over half of their participants, interventions were performed based on CT results, and these results were not evident from the chest radiograph (22). In another study, LUS also showed to change patient management in almost half of patients (33). To our knowledge, there are no studies that contradict the utility of LUS.

There is a need for larger (multicenter) trials to compare chest radiograph and LUS accuracy with the gold-standard CT, but also to investigate the cost-effectiveness, the ease of implementation, and how it affects patient’s outcome and performed interventions. However, it is challenging to perform clinical research on patient-centered outcome for a diagnostic modality as isolated variable, as they are often combined with other modalities. Last, because different LUS protocols were used in the included studies, it is important to determine if LUS protocols are comparable or one protocol is superior to another.

CONCLUSIONS
This study demonstrates that chest radiograph has a low sensitivity and good specificity in critically ill patients with respiratory symptoms when CT is the gold standard. LUS was found to be superior to chest radiograph in these patients as estimates of sensitivity were much higher than for chest radiograph while LUS showed similar (or even higher) specificity. These results question the current use of chest radiograph as the first-line diagnostic modality for critically ill patients with respiratory symptoms. LUS seems to be a good alternative. Larger trials are needed that compare chest radiograph with LUS not only for accuracy but also for effects on outcome, clinical utility, ease of implementation, and cost-effectiveness.

ACKNOWLEDGMENTS
We thank Hans Ket for his advice and help with the literature search and Tom Hallowes for proofreading the article.

REFERENCES


