Author's response to reviews

Title: Improving Accuracy in Assessing Pulmonary Edema on Bedside Chest Radiographs Using a Standardized Scoring Approach

Authors:

Mattias Hammon (matthias.hammon@uk-erlangen.de)
Peter Dankerl (peter.dankerl@uk-erlangen.de)
Heinz L Voit-Höhne (heinz.voit-hoehne@klinikum-ansbach.de)
Ferdinand J Kammerer (ferdinand.kammerer@uk-erlangen.de)
Michael Uder (michael.uder@uk-erlangen.de)
Rolf Janka (rolf.janka@uk-erlangen.de)

Version: 4  Date: 7 December 2013

Author's response to reviews: see over
Dear Editor, dear Reviewers,

we are very grateful for your thoughtful and helpful comments to our manuscript. We revised the manuscript in accordance with your advice, and carefully proof-read the manuscript to minimize typographical, grammatical, and bibliographical errors.

Here below is one-by-one response to your comments.

Associate Editor Comments:

a) significantly improve the language

We and a native speaker (british english) revised the manuscript to improve the language.

b) be more specific in the description of the used methodology.

The Materials and Methods section was extended respectively.

c) clarify how the authors selected the criteria or weighting of the criteria for diagnosing pulmonary edema. Was any preliminary test performed from which they could then test their criteria and weighting in a larger population?

We used information of previously published papers [1, 12, 16, 17].


The authors of these papers addressed these issues.

d) provide inter-reader or intra-reader variability.

To show interreader variability, free-marginal multirater kappa values were calculated for the standard as well as the standardized score-based approach. A consulted statistician thinks that calculating intrareader variability would not reasonable in this context.

e) examine the data more closely and find out which of their imaging criteria were most useful in identifying the presence of edema.

In our study the evaluation process was not monitored. It was thus not possible to establish which imaging criteria best correlate with the EVLW categories. We added a paragraph to the limitations section covering this point.

f) remove the point in the discussion regarding reduction of the need for residents to elect to show certain cases to their staff for help before reporting, since this is controversial and dependent on country regulations.

The paragraph was removed.

Reviewer 1:

1. What was the routine used for chest radiography? What grids, if any, were employed? What distances were used? What timing was used with respect to the respiratory cycle. What detector systems were used?
The Materials and Methods section was extended respectively. We included all demanded information.

2. How did the authors select the particular criteria and the specific weighting of these criteria as they were used to score pulmonary edema? Was a preliminary study carried out to test the criteria from which a final test was to be carried out in a larger population?

We used information of previously published papers [1, 12, 16, 17].


The authors of these papers addressed these issues.

3. Was a test inter-reader or intra-reader variability carried out?

To show interreader variability, free-marginal multirater kappa values were calculated for the standard as well as the standardized software-based approach. A consulted statistician thinks that calculating intrarreader variability would not reasonable in this context.

4. Which of the imaging criteria and weighting best correlated with the EVLW categories?
In our study the evaluation process was not monitored. It was thus not possible to establish which imaging criteria best correlate with the EVLW categories. We added a paragraph to the limitations section covering this point.

5. In Figure 1 a few illustrations appear to be of non-mechanically ventilated patients who would have been excluded under your stated criteria. Why did you use these figures?

In Figure 1 we want to show good examples of all different imaging criteria. Therefore, we also used some bedside chest radiographs that were not included in the study. For this reason, some chest radiographs shown in Figure 1 are of non-mechanically ventilated patients. All patients that were included in the study were mechanically ventilated.

6. P. 6. “...week...”. Do you mean “...weak...”?

Corrected.

Reviewer 2:

ABSTRACT

EVLW is a measured variable normally expressed as mL/kg and not a score, if authors used a different way to express EVLW this should be clearly stated.

Thank you for this hint. We modified the Abstract.

I think that the scoring system improved diagnostic accuracy rather than assessment quality.

You are right. Corrected.
The sentence from "Simultaneously...." to "...standard" should be rewritten for clarity.

**We rephrased the sentence as follows:**
Extravascular lung water values (EVLW, PiCCO (Pulse Contour Cardiac Output) measurements) were determined instantly after bedside CXR imaging and served as a reference standard.

**INTRODUCTION**
I think that the study hypothesis was that the scoring system would increase sensitivity and specificity of radiologists in diagnosing normal vs. elevated EVLW and this should be more clearly stated.

**The hypothesis was rephrased accordingly:**
[The purpose of this study was to investigate whether a standardized score-based assessment of bedside CXRs increases diagnostic accuracy for radiologists when distinguishing between normal and elevated EVLW-values].

**METHODS**
Why was informed consent waived?

**This investigation was approved by the institutional review board of the University of Erlangen. The review board waived the need for informed consent because all PiCCO measurements and chest radiographs were carried out because of a medical indication. No additional procedures needed to be done.**

How was the sample size calculated?

**Unfortunately, we were not able to calculate the sample size in advance because we could not find previously published papers on this topic.**

Why didn't you collect clinical characteristics of patients? I think this is a major limit of the study as edema clinical presentation is a spectrum rather than a yes
or no variable. Moreover, EVLW has relevant prognostic value which was not considered at all by authors.

When assessing bedside CXRs for pulmonary edema, the radiologist should make a clear (yes/no) decision. It is hardly possible correctly report/differentiate borderline cases. Therefore we investigated if score-based bedside chest radiograph assessment of patients with obviously normal/elevated EVLW value increases diagnostic accuracy of radiologists.

You are right. The EVLW value has relevant prognostic value. However, in this study we concentrated on one time point.

Again, I think that edema is a spectrum and I would not have discarded patients with EVLW values of 9-14.

We investigated if score-based bedside chest radiograph assessment of patients with obviously normal/elevated EVLW value increases diagnostic accuracy of radiologists. When assessing bedside CXRs for pulmonary edema, the radiologist should make a clear (yes/no) decision. However, the EVLW is a ratio scale: further studies need to be done to evaluate a possible correlation between the EVLW-value and the result determined by the scoring system without excluding borderline cases.

Were assessing radiologists blinded to study design? The study is designed as to have 10 patients with edema and 10 without lung edema, if radiologists assessing CXRs knew this design this would have introduced a bias as after selecting 10 CXRs with edema they would have known that the rest should have been normal. Again, this is an argument in favor of not having dropped 9-14 EVLW values.

The radiologists were not informed of the overall number of radiographs or the number of patients with normal or elevated EVLW-values. To permit a bias-free evaluation at least eight weeks elapsed between the first (standard approach) and the second (score-based approach) reading of the bedside CXRs, which were randomly presented.
For the reading where the software had been altered, the radiologists were not informed of the software score, nor did they know the scores of the various indications or the threshold distinguishing between pulmonary edema and no pulmonary edema.

RESULTS
I think that authors should find an appropriate test to see if the differences in sensitivity, specificity and K value were statistically significant. Threshold values and negative and positive predictive values should be indicated. It would be interesting to see if performance of the scoring system performed differently in the presence of cardiogenic vs. non-cardiogenic pulmonary edema.

Positive and negative predictive values were calculated and included in the Results section as well as in Table 2. Differentiating cardiogenic and non-cardiogenic pulmonary edema is a very interesting point that should be addressed in future work.

EVLW mostly is an indicator of edema trend over ICU stay and a powerful diagnostic predictor, but these aspects were not considered at all by authors.

You are right. The EVLW value has relevant prognostic value. However, in this study we concentrated on one time point.

DISCUSSION
How would you integrate this findings in clinical practice? Would you recommend scoring all CXRs or only those at risk for lung edema? What about patients with "intermediate" lung edema? If I don't have a PiCCO in place, how could I know if the patient EVLW is below 8 or over 15 and not between 9 and 14, so that I know that your scoring is validated? In facts, we do not know performance of your proposed scoring system in patients with EVLW 9-14, which actually may be a substantial portion of ventilated ICU patients.

This is a very important point we are focusing on in ongoing work.
In the present study we concentrated on patients with an obviously normal/elevated EVLW value. When assessing bedside CXRs for pulmonary edema, the radiologist should make a clear (yes/no) decision. However, the EVLW is a ratio scale: further studies need to be done to evaluate a possible correlation between the EVLW-value and the result determined by the scoring system without excluding borderline cases.

We added a paragraph to the limitations section which covers this point.

Table 1 was modified. Thank you again for your kind and thoughtful review. We hope that the revision addresses your concerns.

Sincerely,

the authors