Reviewer’s report

Title: Automated identification of pneumonia in chest radiograph reports in critically ill patients

Version: 1 Date: 10 May 2013

Reviewer: Michael Klompas

Reviewer’s report:

Liu and colleagues present work on NLP tools to automate identification of pneumonia in critically ill patients using radiographic reports. As they note, a particular challenge in this work is management of the large number of indeterminate reports. Strengths of this study include the use of data from a large number of patients seen in a large number of ICUs, the use of robust software to facilitate NLP analysis, and their success developing algorithms that appear to cope successfully with the large number of uncertain reports. The authors are also to be commended for the descriptive detail they provide in the Online Supplement which helps make their work more transparent and hence more likely to be useful to others.

The major thing missing from this study is discussion of the potential application of this work. At first blush this might appear obvious – it could be used for surveillance or for studies on risk factors and treatments for pneumonia or for clinical decision support – but on further reflection I am not so sure. The major impediment to using the system is not a problem with the author’s NLP algorithm but a weakness of the underlying data. As the authors helpfully indicate, a substantial number of chest x-rays in critically ill patients are ambiguous. Indeed, the percentage of ambiguous radiographs (42%) far outweighs the percentage of unambiguously positive radiographs (7%). This is a major clinical and hence informatics problem because “possible pneumonia” is not the same as “no pneumonia.” It is highly likely, instead, that a substantial fraction of the patients with indeterminate radiographs do in fact have pneumonia. Surveillance, research, and clinical decisions support algorithms that failed to account for the hidden fraction of true pneumonias amongst the possible pneumonias will therefore miss a significant number of patients. The authors make brief mention in their Discussion that they hope to integrate their radiographic interpretation algorithms with additional clinical data. It would be nice to hear more about this but I worry that this might not solve the problem of ambiguity since there are no pathognomonic clinical signs for pneumonia in critically ill patients (e.g JAMA 2007;297:1583). So, while I think Liu and colleagues have accomplished a commendable informatics feat, it is not clear to me where their work will find its place clinically.

Major Compulsory Items to Address:

1. Please describe the performance of the algorithm (sens, spec, npv, ppv) for
“possible pneumonia” (add to results section and to Table 3).

2. Table 3 - please clarify how “validation/derivation” differs from “overall.” I presume it’s the addition of the development CXRs. If so, please include their performance characteristics in this Table as well in order to make the “overall” classification more transparent.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests