Validation Study of an Automated Electronic Acute Lung Injury Screening Tool

HELEN C. AZZAM, MD, MPH, SATIEET S. KHALSA, RICHARD URBANI, CHIRAG V. SHAH, MD, MS, JASON D. CHRISTIE, MD, MS, PAUL N. LANKEN, MD, BARRY D. FUCHS, MD

Abstract

Objective: The authors designed an automated electronic system that incorporates data from multiple hospital information systems to screen for acute lung injury (ALI) in mechanically ventilated patients. The authors evaluated the accuracy of this system in diagnosing ALI in a cohort of patients with major trauma, but excluding patients with congestive heart failure (CHF).

Design: Single-center validation study. Arterial blood gas (ABG) data and chest radiograph (CXR) reports for a cohort of intensive care unit (ICU) patients with major trauma but excluding patients with CHF were screened prospectively for ALI requiring intubation by an automated electronic system. The system was compared to a reference standard established through consensus of two blinded physician reviewers who independently screened the same population for ALI using all available ABG data and CXR images. The system’s performance was evaluated (1) by measuring the sensitivity and overall accuracy, and (2) by measuring concordance with respect to the date of ALI identification (vs. reference standard).

Measurements: One hundred ninety-nine trauma patients admitted to our level 1 trauma center with an initial injury severity score (ISS) ≥ 16 were evaluated for development of ALI in the first five days in an ICU after trauma.

Main Results: The system demonstrated 87% sensitivity (95% confidence interval [CI] 82.3–91.7) and 89% specificity (95% CI 84.7–93.4). It identified ALI before or within the 24-hour period during which ALI was identified by the two reviewers in 87% of cases.

Conclusions: An automated electronic system that screens intubated ICU trauma patients, excluding patients with CHF, for ALI based on CXR reports and results of ABGs is sufficiently accurate to identify many early cases of ALI.


Introduction

Acute lung injury (ALI) is a clinical syndrome defined by the American European Consensus Conference (AECC) criteria: (1) acute onset of bilateral pulmonary infiltrates on chest radiograph (CXR) that are consistent with pulmonary edema; (2) a ratio of the partial pressure of arterial oxygen to the fraction of inspired oxygen (PaO₂/FiO₂) ≤ 300; and (3) a pulmonary artery occlusion pressure ≤ 18 mm Hg, if measured, or no clinical evidence of congestive heart failure. The more severe form of this disorder, the acute respiratory distress syndrome (ARDS), requires a PaO₂/FiO₂ ≤ 200. Together, ALI and ARDS remain a major cause of morbidity, mortality, and cost in intensive care units (ICUs) worldwide. Most ICU patients with ALI require mechanical ventilatory support. Unfortunately, mechanical ventilation itself can exacerbate ALI and impact adversely on patient outcome. Based on evidence from two randomized clinical trials, mortality can be reduced significantly in patients with ALI by employing a lung protective ventilation (LPV) strategy, consisting of small tidal volumes and limiting alveolar pressures compared to ventilation using traditional larger tidal volumes. However, studies show that LPV has not been adopted widely into ICU practice, even at academic medical centers that participated in the clinical trials establishing its value in significantly reducing mortality. It is important to understand why LPV is not used in eligible patients to design effective interventions to increase use of this life-saving ventilation strategy. Although clinicians may not employ LPV in patients with ALI for a variety of reasons, several studies suggest that under-recognition of...
ALI by clinicians is an important factor. Acute lung injury may be under-recognized due to the subjective nature of the radiographic and hemodynamic criteria for ALI, the large number of diseases that can imitate ALI, and the fact that the PaO2/FiO2 ratio may not be routinely calculated by clinicians at the bedside.

Because early recognition of ALI is critical if the mortality benefits of LPV are to be achieved, we reasoned that an automated electronic system that could identify ALI accurately and alert clinicians would be a useful clinical tool. In addition, in this era of increasing focus on quality improvement and benchmark comparisons, such electronic data may help to increase diagnostic efficiency, objectivity, and surveillance of ALI for clinical and quality improvement purposes. Pilot studies have demonstrated the feasibility of surveillance and diagnostic tools for ventilator-associated pneumonia and transfusion-related acute lung injury.

Because early recognition of ALI is critical if the mortality benefits of LPV are to be achieved, we reasoned that an automated electronic system that could identify ALI accurately and alert clinicians would be a useful clinical tool. In addition, in this era of increasing focus on quality improvement and benchmark comparisons, such electronic data may help to increase diagnostic efficiency, objectivity, and surveillance of ALI for clinical and quality improvement purposes. Pilot studies have demonstrated the feasibility of surveillance and diagnostic tools for ventilator-associated pneumonia and transfusion-related acute lung injury.

We report herein the design and validation of an automated electronic system that uses electronic laboratory, radiographic, and demographic data from our hospital information systems to identify intubated patients with ALI. The design and methodology of this electronic ALI finder system was originally presented and published in abstract form. Validation of the system was performed in a cohort of patients with major trauma at risk for ALI by evaluating the system’s performance compared to two blinded physician reviewers, who independently screened the same population of patients for ALI.

**Methods**

**Design of ALI Screening Tool**

The overall design of the automated electronic system was based on the explicit methodology employed by a senior research coordinator for screening and diagnosing patients with ALI in our hospital using AECC criteria (Fig 1), as part of the ARDS clinical trials network (ARDSNet) sponsored by NIH NHLBI. The system screened all patients in ICUs for ALI requiring intubation 24 hours per day, 7 days per week. Although the AECC definition does not include the requirement for intubation, the mortality benefits of LPV were demonstrated in mechanically ventilated intubated patients.

As a first step, the system was programmed to identify intubated patients by either of two different mechanisms: (1) recognition of the electronic order placed for arterial blood gas (ABG) collection that designates intubation status, or (2) detection of the words “ETT”, “endotracheal tube”, or “tracheostomy tube” in any text field in the CXR report. If intubation was detected, a timeline was initiated into which all future data streams were incorporated.

Table 1

<table>
<thead>
<tr>
<th>Terms Indicating Bilaterality</th>
<th>Terms Synonymous with Infiltrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>infiltrates</td>
</tr>
<tr>
<td>Biapical</td>
<td>opacities</td>
</tr>
<tr>
<td>Bibasilar</td>
<td>Air and Space disease</td>
</tr>
<tr>
<td>Widespread</td>
<td>pneumonia</td>
</tr>
<tr>
<td>Diffuse</td>
<td>aspiration</td>
</tr>
<tr>
<td>Perihilar</td>
<td></td>
</tr>
<tr>
<td>Parahilar</td>
<td></td>
</tr>
<tr>
<td>Multifocal</td>
<td></td>
</tr>
<tr>
<td>Extensive</td>
<td></td>
</tr>
</tbody>
</table>

**Left- and right-sided infiltrates described in separate sentences**

ARDS

Pulmonary edema

Congestive heart failure

**ALI** = acute lung injury; **ARDS** = acute respiratory distress syndrome.
studies have shown that ALI cannot be distinguished reliably from CHF using the chest radiograph, even though patients with CHF were excluded from the current study.¹⁸ Only “final” radiology reports (i.e., electronically signed by attending radiologists) were screened, and subsequent modifications were ignored.

For cases identified by the system as meeting criteria for ALI, demographic information such as name, medical record number, and location were automatically imported into an on-line password-protected database on the hospital’s intranet. Electronic alerts were sent via e-mail to the study coordinator for these cases.

The automated system (© 2008 by the Trustees of the University of Pennsylvania) was designed to operate using relevant clinical information available electronically from the information systems in the University of Pennsylvania Health System (UPHS). It runs on a Linux system and is written in Perl. The system searched three main streams of data. The CXR reports arrive as free text from IDXrad (GE Healthcare), our hospital’s radiological information system. At our hospital, the median time between when a CXR is taken and the time it appears in IDXrad as a “preliminary” report is 1.02 hours (average 2.39 h) and as a “final” report is 2.01 hours (average 7.36 h). Laboratory results for ABGs from mechanically ventilated patients arrive from the Cerner laboratory system via MedView, a UPHS-developed clinician portal and clinical data warehouse used to access multiple clinical systems. The average time between when a laboratory sample is processed and time-stamped by Cerner and the time it is available in MedView is 48.5 seconds.

**Study Population**

The study was conducted at the Hospital of the University of Pennsylvania, one of the ARDSNet centers from 1994 to 2006 and an academic level 1 trauma center. One hundred ninety-nine severely injured trauma patients, without isolated head injury, and with an ISS ≥ 16 calculated by data from the first 24 hours of hospital admission, were assembled as part of a single-center prospective cohort study from October 2005 through Apr 2007.¹⁹ The ISS is a parameter that reflects extent and severity of injuries and has been used to identify patients at risk for ALI.¹⁸–²¹ Additional inclusion/exclusion criteria (including exclusion of patients with current or a known history of CHF) and Institutional Review Board approval were described in the original study.¹⁹

**Validation Study Design**

The cohort of 199 patients was screened independently by the automated electronic system and two physician reviewers for ALI. The electronic system screened CXR reports and laboratory data for these patients automatically starting at the time of their admission to the Surgical ICU, as part of ongoing surveillance of all hospitalized patients. The physician reviewers, who were blinded to the cases identified by the electronic system, screened the cohort for ALI by viewing the CXR images and by manually calculating the PaO₂/FiO₂ ratios using all available CXR and laboratory data for each patient.¹⁹ Acute lung injury (ALI) was diagnosed by the physician reviewers when a mechanically ventilated patient met laboratory and radiographic criteria for ALI within 24 hours of each other, during the first 5 days in the ICU after presenting with major trauma. The date and time when a patient met criteria for ALI were also documented. A third reviewer was available for adjudication of discordant cases. Only cases that both physician reviewers interpreted independently as ALI, or discordant cases found to be consistent with ALI after adjudication, were considered true cases. All reviewers (JDC, PNL, CVS) had experience and training in interpreting chest radiographs as investigators in ARDSNet clinical trials of patients with ALI. The reviewers considered a CXR as “positive” for ALI only if the image unequivocally showed bilateral infiltrates consistent with pulmonary edema.

**Statistical Analysis**

Sensitivity, specificity, positive and negative predictive values, and diagnostic accuracy of the system were calculated using the consensus diagnoses of the physician reviewers as the reference standard. Results were expressed as proportions with 95% confidence intervals. We then compared the day on which ALI was first identified by the automated system to the day on which it was first identified by the two reviewers. The calendar day on which patients were admitted to the Emergency Department was considered Day 0, regardless of the time of day. The calendar days following Day 0 were labeled as Day 1, Day 2, etc.

**Sample Size**

The study was powered to evaluate an estimated sensitivity of 95%, such that the lower bound of the confidence interval did not include 90%. Using the modified Wald method, we calculated that a sample size of 200 would allow us to obtain 95% confidence intervals that range from 0.91 to 0.97 around an estimated 95% specificity.²²

**Results**

Of the 199 patients with major trauma, fifty-three were identified by the two physician reviewers as meeting criteria for ALI. Thus, the incidence of ALI in the study population was 26.6% (53/199) (95% CI 20.6 — 33.3). The automated system identified forty-six of these fifty-three cases as well as sixteen additional ones (Table 2). The performance parameters for ALI are shown in Table 2. The performance parameters of the automated system are shown in Table 3. The performance parameters were calculated using the consensus diagnoses of the physician reviewers as the reference standard. Results were expressed as proportions with 95% confidence intervals. We then compared the day on which ALI was first identified by the automated system to the day on which it was first identified by the two reviewers. The calendar day on which patients were admitted to the Emergency Department was considered Day 0, regardless of the time of day. The calendar days following Day 0 were labeled as Day 1, Day 2, etc.

**Table 2** | 2×2 Table that Compares Cases Identified as ALI by the Automated System and by the Two Physician Reviewers

<table>
<thead>
<tr>
<th>Identified by Automated System</th>
<th>Not Identified by Automated System</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified by Two Physician Reviewers</td>
<td>46</td>
<td>16</td>
</tr>
<tr>
<td>Not identified by Two Physician Reviewers</td>
<td>7</td>
<td>130</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>146</td>
</tr>
</tbody>
</table>

ALI = acute lung injury.

**Table 3** | Performance Parameters (with 95% confidence intervals) of Automated System

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy*</th>
</tr>
</thead>
<tbody>
<tr>
<td>86.8% (82.3–91.7)</td>
<td>89.0% (84.7–93.4)</td>
<td>74.2% (67.9–80.1)</td>
<td>94.9% (92.0–98.0)</td>
<td>88% (83.5–92.5)</td>
</tr>
</tbody>
</table>

PPV = positive predictive value; NPV = negative predictive value. *Accuracy = (true positives + true negatives)/total study population.
We next reviewed the radiographic reports of the sixteen patients identified by the system but not by the physician reviewers (i.e., false positives). In all of these reports, we found that the attending radiologists had indicated the presence of bilateral infiltrates, which triggered the system to identify them as qualifying radiographs. The specific reason for the disagreement between the two reviewers and the staff radiologists is unknown. One possibility is that the two physician reviewers may not have agreed with the radiologists that the infiltrates were bilateral in some or all of these 16 patients. Alternatively, the reviewers may have judged that the infiltrates, although bilateral, were not unequivocally consistent with pulmonary edema.

Discussion

Our study’s major finding is that an automated electronic system identified cases of ALI with both sensitivity and specificity greater than 87% compared with two physician reviewers. In 87% of the cases, the system identified ALI on or before the day on which it was identified by the reviewers.

This electronic system is an exportable tool for objective and automated surveillance for ALI and LPV eligibility. Given the estimated 200,000 cases of ALI each year (62.4 cases per person-year population),23 and the estimated mortality reduction between 8.8 and 33% in this population using LPV,3,4 use of such systems may have a significant impact on public health if shown to increase use of LPV.

In view of the many reports indicating that physicians under-use LPV in patients with ALI,4–9 systems-based solutions will likely be required to change physician behavior. Several studies have described the use of automated prompting systems to promote adherence to clinical guidelines.24,25 One study evaluated an automated tool that calculates the PaO2/FiO2 ratio to screen for transfusion-related ALI.25 However, to our knowledge the automated electronic system described in this report represents the first to diagnose ALI using laboratory data, CXR reports, and relevant demographic patient information. In addition, it appears to be unique in its ability to detect intubation status; using both CXR and laboratory data appeared to be effective in capturing intubated patients as the automated electronic system did not miss any patients in this study (relative to the physician reviewers) due to failure to recognize intubation status.

Table 4 includes an itemized list of the radiographic descriptors found in the radiology reports for the seven cases of ALI (as judged by the two physician reviewers) that were miscategorized by the automated system (i.e., false negatives). Four of the seven cases were not identified by the system due to unanticipated variations in language. For example, the system was programmed to identify “congestive heart failure” or “CHF”, but not “fluid overload”, “congestive failure”, or “congestion.” Additional cases were missed because the system was not programmed to recognize the word “contusion” as an infiltrate. Finally, two cases were missed by the system because the radiographic report used the phrase “bilateral atelectasis” or “bilateral plural effusions.”

Table 4 ■ Itemized List of Radiographic Reports for Cases Not Diagnosed as ALI by the Automated System (false negatives)

<table>
<thead>
<tr>
<th>Case</th>
<th>Radiographic Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>“patchy bibasilar atelectasis”</td>
</tr>
<tr>
<td>Case 2</td>
<td>“fluid overload”</td>
</tr>
<tr>
<td>Case 3</td>
<td>1st CXR: “pleural effusions” 2nd CXR: “congestion”</td>
</tr>
<tr>
<td>Case 4</td>
<td>1st CXR: “contusion” on the left, “atelectasis” on the right</td>
</tr>
<tr>
<td></td>
<td>2nd CXR: “aspiration” on the right, “opacity” in the left upper lobe</td>
</tr>
<tr>
<td>Case 5</td>
<td>“contusion left midlung”</td>
</tr>
<tr>
<td>Case 6</td>
<td>“bilateral atelectasis” and “right upper lobe hemorrhage/contusion”</td>
</tr>
<tr>
<td>Case 7</td>
<td>“congestive failure or fluid overload”</td>
</tr>
</tbody>
</table>

ALI = acute lung injury; CXR = chest radiograph.
We hypothesize that, by linking the computerized ALI recognition system with an automated prompt to clinicians, this system may increase the use of LPV and thereby improve the outcome of patients with ALI. Furthermore, our system can prompt clinicians to consider using LPV in all cases of ALI that it identifies, but especially in early cases of ALI, which may not have otherwise been recognized. In this way, it may also serve to educate clinicians about the AECC definition of ALI and the limitations of that definition.6,26

One of the strengths of this system is the timeliness with which patients were identified compared to the two physician reviewers, as the system diagnosed ALI on or before the time physician reviewers did so 87% of instances. The ABGs are automatically collected by the system every 6 hours, and are therefore processed within an average of 3 hours. The median delay between the time a CXR was taken and the time it appeared in our system as a “final” report was 4.0 hours (average 15.4 h). Although delays between when a CXR was taken and when it was read and stamped as “Final” could theoretically have led to delayed identification by the system, the concordance analysis suggests this is not the case. Therefore, the automated system is capable of making real-time diagnoses of ALI and initiating a rapid response by staff.

Although our system performed well when compared to a physician-determined standard, the concordance was not perfect. Of the seven ALI cases not identified by the system (false negatives), four cases would have been identified successfully had we anticipated some minor variations in language used by our radiologists. Our system can be modified easily to include these variations to potentially improve its performance. In contrast, three cases were missed because the radiologists used the terms atelectasis or pleural effusions to describe radiographic densities. These cases, in which there was disagreement among experts on the nature of the radiographic densities, illustrate the difficulties in diagnosing ALI using the chest radiograph. We do not plan to revise our radiographic query to pick up similar cases, since inclusion of all cases with atelectasis or pleural effusions would likely lead to an increase in the false-positive rate.

For the sixteen patients identified as ALI by the system but not by the physician reviewers (false positives), there was a disparity between the radiologists’ interpretations of the chest radiographs (upon which the automated system relies) and the reviewers’ interpretations. This is not unexpected, given the subjective nature and poor interrater reliability of the CXR interpretation of ALI.9 Given the overall safety of LPV and its potential efficacy, it seems preferable to have an automated system that has a high sensitivity at the risk of identifying occasional false-positive cases of ALI.

Our study has several limitations. The first is using two physician reviewers as our diagnostic reference standard. Unfortunately, there is no true “gold standard” for the diagnosis of ALI.6 In addition, there is considerable interobserver variability in the interpretation of CXRs among critical care physicians, particularly in relation to findings such as pleural effusions, atelectasis, isolated lower lobe involvement, small lung volumes, and increased interstitial markings.9 Although this interobserver variability was decreased in our study by including two experienced physician reviewers as well as an adjudication process, it is still possible that our reference standard was “tarnished.” Nonetheless, we believe our system to be a useful adjunct for individual clinicians as well as an objective and effective surveillance tool.

Furthermore, concerns about the definition of ALI as delineated by the AECC have been raised. In one study that compared clinical criteria to the histopathological finding of diffuse alveolar damage in ALI cases diagnosed by clinical criteria, the sensitivity of the AECC criteria was 75% and the specificity was 84%.6 However, this definition is still widely used in everyday clinical practice and it represents the standard criteria by which patients are recruited into clinical trials. More importantly, the morbidity and mortality benefits of LPV were found in intubated patients with ALI defined by the AECC criteria. Thus, despite the inherent problems in defining ALI clinically, use of an automated electronic system such as we report has the potential to improve outcomes in this population. If future modification(s) are made to the AECC definition of ALI, these can likely be integrated into this automated system.

There are several issues that may impact the generalizability of our results. First, the performance of our system when screening a general population of at-risk patients is unknown, since this study used a select population of patients who had major trauma as a uniform risk factor for ALI and from whom patients with a history of congestive heart failure or suspected current congestive heart failure were excluded. We would anticipate that the accuracy of this type of system would be lower if used to screen populations that included patients with congestive heart failure. To avoid such a problem, one could exclude certain ICUs from being screened by the system, such as a postcardiac surgery ICU. In addition, our system may not be generalizable to nonintubated patients. However, this was a validation study designed to test the tool in a population at high risk for ALI, in which there existed a reliable reference standard, to evaluate its diagnostic performance as a clinical and research tool. Furthermore, the system was designed primarily to improve implementation of life-saving ventilator care (LPV), which is not relevant in a nonintubated population. Future studies testing our system in other at-risk groups are needed.

In addition, the electronic system is dependent on the use by attending radiologists of terms recognized by the automated system. Furthermore, since the software dictionary used in our system was developed from the terms and phrases employed by our hospital radiologists, this system may not provide consistent diagnostic accuracy among different staff or in other hospitals. Although all possible terms cannot be anticipated, pilot studies showed the terms used by our radiologists to describe ALI to be relatively constant (data not shown). It is likely this terminology is also relatively conserved between institutions. The true impact of this limitation can only be assessed by testing the system’s performance in other institutions. Fortunately, patients with ALI typically have multiple films performed, all of which would be reviewed by the system; this would tend to compensate for the potential adverse impact on performance
due to inconsistencies in language used by different staff radiologists.

Natural language processing has been shown to be superior to text-based methods that rely on detection of relevant phrases in the diagnosis of pneumonia from radiology reports, and could improve the accuracy of our automated system. However, the keyword search described in our system is more complex than the keyword searches evaluated in these papers. Furthermore, cost and availability precluded the use of natural language processing technology at our institution at this time.

Finally, the electronic system was designed and tested within a single hospital using software that integrated a unique set of hospital information systems. However, software programming required to replicate our system in other information systems environments could be easily applied. Furthermore, with the advent and increasing use of inpatient electronic medical records, programming or “plugging in” the software design into these commercial products may be facilitated.

In conclusion, this study demonstrates that an automated electronic system is capable of prospectively identifying many trauma patients with ALI early and accurately, when CHF patients are excluded. As such, the system has many potential applications in clinical, research, and quality improvement efforts. This ALI recognition system can be linked with an automated prompt to clinicians to consider potential applications in clinical, research, and quality improvement efforts. This ALI recognition system can be linked with an automated prompt to clinicians to consider the diagnosis of ALI and the early use of LPV. As such, this novel system offers the potential to improve the outcome of patients with ALI. Further studies are recommended to assess the impact of this system on clinical outcomes.

References