Electronic “Sniffer” Systems to Identify the Acute Respiratory Distress Syndrome

Max T. Wayne¹, Thomas S. Valley^{2,3}, Colin R. Cooke^{2,3*}, and Michael W. Sjoding^{2,3,4}

¹Department of Internal Medicine, ²Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, ³Institute for Healthcare Policy & Innovation, and ⁴Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan

Abstract

Background: The acute respiratory distress syndrome (ARDS) results in substantial mortality but remains underdiagnosed in clinical practice. Automated ARDS “sniffer” systems, tools that can automatically analyze electronic medical record data, have been developed to improve recognition of ARDS in clinical practice.

Objectives: To perform a systematic review examining the evidence underlying automated sniffer systems for ARDS detection.

Data Sources: MEDLINE and Scopus databases through November 2018 to identify studies of tools using routinely available clinical data to detect patients with ARDS.

Data Extraction: Study design, tool description, and diagnostic performance were extracted by two reviewers. The Quality Assessment of Diagnostic Accuracy Studies-2 was used to evaluate each study for risk of bias in four domains: patient selection, index test, reference standard, and study flow and timing.

Synthesis: Among 480 studies identified, 9 met inclusion criteria, and they evaluated six unique ARDS sniffer tools. Eight studies had

derivation and/or temporal validation designs, with one also evaluating the effects of implementing a tool in clinical practice. A single study performed an external validation of previously published ARDS sniffer tools. Studies reported a wide range of sensitivities (43–98%) and positive predictive values (26–90%) for detection of ARDS. Most studies had potential for high risk of bias identified in their study design, including patient selection (five of nine), reference standard (four of nine), and flow and timing (three of nine). In the single external validation without any perceived risks of biases, the performance of ARDS sniffer tools was worse.

Conclusions: Sniffer systems developed to detect ARDS had moderate to high predictive value in their derivation cohorts, although most studies had the potential for high risks of bias in study design. Methodological issues may explain some of the variability in tool performance. There remains an ongoing need for robust evaluation of ARDS sniffer systems and their impact on clinical practice.

Systematic review registered with PROSPERO (CRD42015026584).

Keywords: acute lung injury; acute respiratory distress syndrome; diagnostic tool; identification; systematic review

(Received in original form October 24, 2018; accepted in final form December 6, 2018)

*C.R.C. is Deputy Editor of *AnnalsATS*. His participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

Supported by National Heart, Lung, and Blood Institute grants K23HL140165 (T.S.V.) and K01HL136687 (M.W.S.).

Author Contributions: C.R.C. and M.W.S.: conception and design of the study; M.T.W., T.S.V., and M.W.S.: substantial contributions to data collection; M.T.W. and M.W.S.: data analysis and interpretation; M.T.W.: drafted the manuscript; and, T.S.V., C.R.C., and M.W.S.: critical manuscript revisions. All authors provided final manuscript approval.

Correspondence and requests for reprints should be addressed to Michael W. Sjoding, M.D., M.Sc., University of Michigan, NCRC Building 16, G027W, 2800 Plymouth Road, Ann Arbor, MI 48109. E-mail: msjoding@umich.edu.

This article has an online supplement, which is accessible from this issue’s table of contents at www.atsjournals.org.

Ann Am Thorac Soc Vol 16, No 4, pp 488–495, Apr 2019

Copyright © 2019 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201810-715OC

Internet address: www.atsjournals.org

Almost one-fourth of patients requiring mechanical ventilation develop the acute respiratory distress syndrome (ARDS), conferring a 40% in-hospital mortality rate (1–5). Yet, 40% of patients with ARDS are not identified, and 35% do not receive low-tidal volume ventilation despite its proven mortality benefit (3). Underrecognition of patients with ARDS is believed to be a major reason for inadequate treatment (6, 7). For this reason, automated electronic “sniffer” systems that analyze electronic records have been developed to assist clinicians with the identification of ARDS in clinical practice (8–12).

Individual ARDS sniffer systems can automatically analyze electronic health record data, including the text of radiology reports and laboratory data, to identify patients with ARDS in real time. Initial reports have described promising diagnostic performance of such tools when compared with adjudication of patients for ARDS by clinical reviewers (8, 10–15). However, in one subsequent evaluation, sniffer system performance was somewhat worse (16). Although there may be many potential reasons why these tools may have a degradation in performance, specific factors are understudied (17, 18). One concern is that subtle biases were present in tool development studies, which could artificially inflate initial results (19). Ultimately, it is not clear whether any current ARDS sniffer systems are ready for widespread implementation across health systems.

We performed a systematic review of currently published electronic sniffer systems for ARDS, examining the diagnostic performance of individual tools across studies while also evaluating for study biases using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) framework (20). We hypothesized that variation in study quality and potential risks for biases in study design could explain the variability in performance.

Methods

Information Sources and Search Strategy

We searched MEDLINE and Scopus databases through November, 29, 2018, to identify studies that used routinely available clinical data to identify patients with ARDS. With the assistance of a medical research librarian, we used Boolean logic to search for

studies with the following three concepts: “ARDS or acute lung injury” and “diagnosis” (or related terms) and “decision support systems” (or related terms). We used both controlled vocabulary terms and keywords to capture these concepts. The exact search strategy is available in the online supplement. The reference lists of all studies selected for full-text review were also examined for additional study titles that may have met study inclusion criteria. Abstracts of any additional studies identified in this manner were reviewed by two authors for potential inclusion as detailed below. We did not search conference abstracts, because our explicit goal was to examine published sniffer tools. All identified citations were imported into a reference manager (EndNote X8; Thomson Reuters).

This systematic review was prospectively registered with PROSPERO (CRD42015026584) before we proceeded with the search. The prespecified analysis plan did not include a meta-analysis, owing to concerns that there would not be sufficient studies of any one individual sniffer tool. We also planned to contact the original study authors for clarification if needed after review of the published articles.

Study Eligibility and Selection Criteria

Criteria for study eligibility were full-text studies testing tools that used data routinely available in electronic medical records or gathered by routine history taking to identify patients with ARDS or full-text studies that evaluated the impact of such tools on clinical outcomes. We focused on tools that could be applied to any patient in the emergency department, hospital ward, or intensive care unit (ICU). We excluded studies that 1) evaluated algorithms that predicted the development of ARDS at a future point, 2) used data not routinely available in clinical practice (e.g., studies testing experimental biomarkers), 3) identified transfusion-associated lung injury, 4) were performed in operating room settings, or 5) were published in a language other than English.

After the initial search, two investigators (M.T.W., T.S.V., or M.W.S.) independently screened all titles and abstracts for study inclusion. Any article flagged for full review by either investigator was reviewed. Full-text articles were independently reviewed by two authors

(M.T.W., T.S.V., and/or M.W.S.) for inclusion and exclusion criteria.

Data Abstraction

We developed an online data extraction tool for study review using REDCap (Vanderbilt University) (21). For each eligible study, two authors (M.T.W. and T.S.V.) independently extracted data on study characteristics, including patient selection, study setting, study design, number of subjects, and test characteristics. A copy of the extraction tool is provided in the online supplement. Studies were categorized as either 1) derivation only, 2) derivation and validation, 3) temporal validation, or 4) external validation. Derivation-only studies were defined as those in which a tool was derived using a single patient cohort. Derivation and validation studies were those in which the study tool was derived in one patient cohort and validated on a different cohort drawn from a similar patient population. Temporal validation studies were defined as studies in which a previously developed tool was tested on a patient population similar to the original validation (e.g., same institution) but during a different time period. External validation was defined as studies in which the tool was tested in a new cohort drawn from a different population compared with the original derivation study (e.g., separate medical system).

Discrepancies in data abstraction and study quality assessment were resolved by a third person (M.W.S.). Test characteristics abstracted included sensitivity, specificity, positive predictive value (PPV), negative predictive value, and area under receiver operating characteristic curve. We report sensitivity and PPV because of their importance for early warning system evaluation (22).

Study Quality

For each eligible study, two authors (M.T.W. and T.S.V.) independently assessed each study using criteria adapted from the QUADAS-2 (20). For each study, risk of bias was assessed in the following four domains:

1. *Patient selection*: Cohort study versus case-control study, inappropriate patient exclusions, population representative of patients in clinical practice
2. *Reference standard*: How ARDS was confirmed, whether the evaluation was blinded to new tool result

3. *Index test*: How the sniffer tool was performed and interpreted
4. *Flow and timing*: Reference standard applied similarly to all patients

A third author (M.W.S.) adjudicated any discrepancies. No study was excluded on the basis of study quality.

This review was performed in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) standards (23). The PRISMA-P checklist is included in the online supplement.

Results

We identified 480 citations using the initial search strategy (Figure 1). Of these, 52 citations were identified for full-text review, and 9 studies met inclusion criteria (Table 1). Reasons for exclusion included studies evaluating tools that predict ARDS development ($n = 17$) and studies evaluating tools that predict ARDS outcomes ($n = 6$). All studies were performed in the United States at single academic medical centers, and seven of nine were performed in an ICU setting. There was some heterogeneity

between studies regarding the patients included in the development or testing of tools. For example, some studies included only mechanically ventilated patients (9, 11–14), whereas others included patients with and without invasive mechanical ventilation (8, 15, 16). Among six unique sniffer tools evaluated, four were tested in a single study, whereas the tool originally described by Herasevich and colleagues in 2009 (8) and the ALI Selection System to Identify Subjects for Treatment/Trials (ASSIST) tool (12) were both evaluated in three studies.

Each of the six tools used electronic medical record data to identify patients with ARDS (Table 1). Four of six tools used natural language processing or free-text processing of chest radiograph reports to help identify cases of ARDS, whereas two tools (Chbat and colleagues [10] and Reamaroon and colleagues [15]) did not incorporate radiographic reports as a data source. In two of the models (Solti and colleagues [13] and Yetisgen-Yildiz and colleagues [14]), analysis of the radiograph reports alone were used. Sniffer tools described by Herasevich and colleagues in 2009 (8) and the ASSIST model (12) incorporated data from arterial blood gases (arterial oxygen tension pressure/fraction of

inspired oxygen) and used free-text processing of chest radiographs to identify patients with ARDS. The tool described by Reamaroon and colleagues (15) incorporated vital signs and laboratory data, whereas the tool described by Chbat and colleagues (10) used a patient's medical and surgical histories, ventilatory settings, medications, demographics, and laboratory data.

The purpose of the analysis varied across studies (Table 2). Four studies derived a tool and then validated the tool on a different patient cohort drawn from a similar population (10, 13–15). Four studies performed temporal validations of tools that were previously published or reported in a previous abstract, evaluating these tools on similar populations as the original validation but during a different time period (8, 9, 11, 12). One of these studies also deployed a sniffer tool at a single medical center, evaluating its real-world impact on detecting patients with ARDS and reducing exposure to potentially injurious ventilation (9). Only the study by McKown and colleagues externally validated two previously published tools (Herasevich 2009 and ASSIST), testing them in distinct patient populations within a separate medical system compared with their original validations (16).

Most studies reported good to excellent sensitivities (range, 79–98%), whereas only one reported lower sensitivity at 43% (14) (Table 2). A single derivation and validation study reported an excellent PPV of 90% (13), whereas all other studies had poor to moderate values (range, 26–74%). In the single external validation study of the Herasevich and ASSIST tools (16), the reported sensitivity and PPV were lower than previously reported (ASSIST, sensitivity decreased from 98% to 88% and PPV decreased from 74% to 46%; Herasevich 2009, sensitivity decreased from 96% to 79% and PPV decreased from 56% to 41%).

Quality Assessment

Using the QUADAS-2 framework, reviewers identified study design concerns in seven of nine studies evaluating ARDS sniffer tools that potentially placed them at high risk for bias (Table 3). Only the derivation and validation study performed by Reamaroon and colleagues (15) and the external validation study performed by McKown and colleagues (16) were considered by reviewers to have low risk of bias in each of the QUADAS-2 domains.

Figure 1. Flow diagram for literature search and results. *Other citations include those identified after performing a hand search of the references of articles identified in the original search. ARDS = acute respiratory distress syndrome.

Table 1. Characteristics of studies testing “sniffer” tools to identify acute respiratory distress syndrome

Study	Tool	Years	Setting	Centers	Center Type	n	Patients	Clinical Data Analyzed
Herasevich <i>et al.</i> , 2009 (8)	Herasevich 2009	2006–2006	ICU	Single	Academic	3,795	All ICU	ABG, FI_{O_2} , CXR reports
Azzam <i>et al.</i> , 2009 (12)	ASSIST	2005–2007	ICU	Single	Academic	199	Mechanically ventilated, trauma only	ABG, FI_{O_2} , CXR reports
Solti <i>et al.</i> , 2009 (13)	Solti 2009	—*	ICU	Single	Academic	857	Mechanically ventilated	CXR reports
Herasevich <i>et al.</i> , 2011 (9)	Herasevich 2009	2008–2009	ICU	Single	Academic	111	Mechanically ventilated	See Herasevich <i>et al.</i> , 2009
Koenig <i>et al.</i> , 2011 (11)	ASSIST	2004–2005	ICU	Single	Academic	1,270	Mechanically ventilated	See Azzam <i>et al.</i> , 2009
Chbat <i>et al.</i> , 2012 (10)	Chbat 2012	—*	ICU	Single	Academic	526	—*	Demographics, PMH, VS, laboratory examination results, ventilation settings CXR reports
Yetisgen-Yildiz <i>et al.</i> , 2013 (14)	Yetisgen-Yildiz 2013	—*	ICU	Single	Academic	55	Mechanically ventilated	
McKown <i>et al.</i> , 2017 (16)	ASSIST, Herasevich 2009	2006–2014	ICU	Single	Academic	2,841	All ICU	See Herasevich <i>et al.</i> , 2009, Azzam <i>et al.</i> , 2009
Reamaroon <i>et al.</i> , 2019 (15)	Reamaroon 2018	2016	—*	Single	Academic	401	Hypoxia or mechanically ventilated	Vital signs, laboratory examination results

Definition of abbreviations: ABG = arterial blood gas; ASSIST = ALI Selection System to Identify Subjects for Treatment/Trials; CXR = chest radiograph; FI_{O_2} = fraction of inspired oxygen; ICU = intensive care unit; PMH = past medical history; VS = vital signs.

*Not described in manuscript.

Five studies had high risk of bias in patient selection; four had high risk of bias in how the reference standard was applied to determine which patients developed ARDS; and three had high risk of bias with regard to flow and timing (e.g., the reference standard was not applied in the same way to all patients). In general, studies that reported a higher tool performance also had a higher total number of biases in study design domains (Figure 2).

Discussion

We systematically reviewed the published literature on electronic ARDS identification tools. We identified nine studies testing six different tools that used electronic medical record data to identify ARDS. The vast majority of the studies were either derivations and internal validations of new tools or temporal validations of previously described tools, whereas only one study was an external validation testing two of the previously developed models in a distinct cohort (16). Although most studies reported

reasonable sensitivity of the sniffer tools, they almost uniformly had lower PPV and also performed worse during an external validation. Only a single study examined the effect of implementing an automated electronic ARDS identification tool on clinical practice (9).

After the development of a new diagnostic or prognostic prediction tool, additional steps are required before widespread adoption can occur, including validation studies, impact assessments, and implementation evaluations (24). Validation study designs have a hierarchy of robustness, from temporal validations (new time period), to geographic validations (new institution), to domain validations (new patient group; e.g., adult vs. pediatric). The present systematic review identified one study meeting this second level of validation stringency. Impact assessment evaluates whether the tool changes physician behavior, clinical outcomes, or costs. Implementation studies evaluate practical barriers to clinical adoption and may consider usability, acceptability, and interoperability (25). A single study

performed an impact assessment of an ARDS sniffer tool and also elicited feedback on satisfaction with the system (9). Taken together, the present systematic review did not identify sufficient electronic ARDS “sniffer” system studies or the necessary range of evaluations to recommend widespread adoption.

It remains unknown whether the performance characteristics of current ARDS sniffer tools would be acceptable to clinicians in practice. Because of their high sensitivity, these algorithms may have utility as ARDS screening tools. However, they would still require a human reviewer to determine whether ARDS is present in patients with positive screening results. Yet, performance variation in these tools across institutions again highlights the need for a rigorous validation to verify a tool’s higher sensitivity for identifying ARDS before use in clinical care. The lower PPV of the tools may lead to a false-positive rate that might not be acceptable to some providers. Given the growing concerns that alarm fatigue can contribute to adverse patient outcomes (26), future work is

Table 2. Reported tool performance and study conclusions

Author	Tool	Study Type*	Sensitivity [†]	PPV [†]	Primary Conclusion
Herasevich <i>et al.</i> , 2009 (8)	Herasevich 2009	Temporal validation	96 (94–98)	46 (42–50)	The automatic ALI sniffer accurately identifies patients who develop ALI in critically ill patients.
Azzam <i>et al.</i> , 2009 (12)	ASSIST	Temporal validation	87 (82–92)	74 (68–80)	An automated electronic system can prospectively identify trauma patients with ALI if patients with CHF are excluded.
Solti <i>et al.</i> , 2009 (13)	Solti 2009	Derivation and validation	91	90	An ML-based approach is comparable to physician annotation for identifying ALI on chest radiography reports.
Herasevich <i>et al.</i> , 2011 (9)	Herasevich 2009	Temporal validation	NR [‡]	59 [‡]	EMR surveillance can accurately detect potentially injurious tidal volumes and influence practice.
Koenig <i>et al.</i> , 2011 (11)	ASSIST	Temporal validation	98 (97–98)	74 (72–76)	An automated electronic system identifies patients with ALI with high accuracy at a large academic center.
Chbat <i>et al.</i> , 2012 (10)	Chbat 2012	Derivation and validation	85	69	A novel mathematical model using real-time ICU data can identify patients with ALI.
Yetisgen-Yildiz <i>et al.</i> , 2013 (14)	Yetisgen-Yildiz 2013	Derivation and validation	43	69	Text-processing methods using chest radiography reports in ICU patients meeting oxygenation criteria can identify ALI.
McKown <i>et al.</i> , 2017 (16)	ASSIST, Herasevich 2009	External validation	Herasevich 2009, 79 (76–82); ASSIST, 88 (86–90)	Herasevich 2009, 41 (39–44); ASSIST, 46 (43–48)	Published sniffer algorithms for ARDS may be useful as screening tools but are limited to screening rather than diagnosis, owing to poor specificity.
Reamaroon <i>et al.</i> , 2019 (15)	Reamaroon 2018	Derivation and validation	90	26	Accounting for ARDS diagnostic uncertainty in a model to detect ARDS improves model performance.

Definition of abbreviations: ALI = acute lung injury; ARDS = acute respiratory distress syndrome; ASSIST = ALI Selection System to Identify Subjects for Treatment/Trials; CHF = congestive heart failure; EMR = electronic medical record; ICU = intensive care unit; ML = machine learning; NR = not reported; PPV = positive predictive value.

*Derivation study = tool was derived in the study; derivation and validation = study tool was derived in one cohort and then validated in a different patient cohort drawn from a similar patient population; temporal validation = validation of previously derived tool in a similar but temporally distinct patient population; external validation = validation of study tool tested in new population compared with the original validation.

[†]95% confidence interval in parentheses (if reported).

[‡]Automated alert required patients to be identified by the algorithm described by Herasevich and colleagues 2009 (8) and to have received potentially injurious tidal volume. Performance was reported for alert-positive patients; therefore, only PPV is reported.

needed to better understand what is an acceptable false-positive rate or number needed to screen to identify a case of ARDS in this setting.

For interventions considered low risk for harm, such as lung-protective mechanical ventilation, performance characteristics of current ARDS sniffer tools may be sufficient. As performed in the Herasevich and colleagues 2011 study, these tools can be programmed to alert

providers that a patient with ARDS may not be receiving lung-protective ventilation (9). Implementation of the system in that study resulted in a reduction in patient exposure to potentially injurious tidal volumes while also having a false-positive alert rate less than 50%. One unintended consequence of such a system is that there may have been a reduction in potentially injurious tidal volumes among patients without ARDS (owing to patient

misclassification). However, this unintended consequence is not likely to be associated with an increased risk for harm (27, 28). In contrast, other ARDS treatments, such as prone positioning or paralytics, may cause harms or require significant critical care resources, necessitating an algorithm with a higher PPV (29).

We originally hypothesized that bias in study design and execution could explain the variability in the ARDS tool

Table 3. Risks of study bias evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 framework

Study	Patient Selection	Reference Standard	Index Test	Flow and Timing	Summary of Risk of Potential Bias(es) Identified
Herasevich <i>et al.</i> , 2009 (8)	—	+	—	+	Reviewers were not blinded to sniffer tool result; patients flagged by tool were reviewed differently from those not flagged
Azzam <i>et al.</i> , 2009 (12)	+	—	—	—	Patients with CHF history excluded
Solti <i>et al.</i> , 2009 (13)	+	+	—	—	Secondary use of patients enrolled in a clinical study, which may not be representative of the general ICU population; nonstandard definition of ARDS as reference standard
Herasevich <i>et al.</i> , 2011 (9)	—	—	—	+	Tool evaluated only among patients flagged by tool
Koenig <i>et al.</i> , 2011 (11)	+	+	—	+	Patients excluded if admitted to cardiac care unit, if after nontrauma surgery in first 48 h, or $FiO_2 < 50\%$; reference standard not applied similarly to all patients; reviewer not blinded to tool result for most patients
Chbat <i>et al.</i> , 2012 (10)	+	—	—	—	Methods implied case-control design
Yetisgen-Yildiz <i>et al.</i> , 2013 (14)	+	+	—	—	Secondary use of patients enrolled in a clinical study, which may not be representative of the general ICU population; nonstandard definition of ARDS as reference standard
McKown <i>et al.</i> , 2017 (16)	—	—	—	—	N/A
Reamaroon <i>et al.</i> , 2019 (15)	—	—	—	—	N/A

Definition of abbreviations: + = high risk of bias; — = low risk of bias; ARDS = acute respiratory distress syndrome; CHF = congestive heart failure; FiO_2 = fraction of inhaled oxygen; ICU = intensive care unit; N/A = not applicable.

QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) risk of bias includes four domains:

Patient selection: Cohort study versus case-control study, inappropriate patient exclusions

Reference standard: How ARDS was confirmed, whether evaluation was blinded to new tool result

Index test: How the sniffer tool was performed and interpreted

Flow and timing: Reference standard applied similarly to all patients

performance across studies. We evaluated each study for risk of bias in four major domains according to criteria adapted from QUADAS-2 (20). Patient selection and reference standard application were the two domains identified for which high risks of

bias were most common. Some studies validated tools using clinical study cohorts that may not be representative of patients encountered in clinical practice (13, 14). Others excluded patients with a history of congestive heart failure (11, 12). Because

an increasing number of patients admitted cardiac ICUs have noncardiac primary diagnoses (30), and because as many as 8% may have ARDS (8), an ideal sniffer system should also perform well in this setting.

Figure 2. (A and B) Acute respiratory distress syndrome “sniffer” tool performance in each study and the corresponding number of study design domains with high risk of bias identified. ASSIST = ALI Selection System to Identify Subjects for Treatment/Trials tool.

How the reference standard was applied and how ARDS was confirmed comprised another design domain with high risk of bias in some studies. For example, researchers were not blinded to the result of the sniffer tool, which may bias their assessment of whether ARDS developed (8, 11). More generally, difficulty with reliably diagnosing patients with ARDS is well described (31–34) and poses a unique challenge to studies validating an ARDS sniffer tool. Simply because the reliability of ARDS diagnosed by a single reviewer is moderate, the performance of an ARDS algorithm might vary if two different clinicians generate the reference standard (35). To combat this problem, future studies should rely on the combination of multiple independent reviewers to establish the gold standard and minimize measurement error (33).

It remains unknown how well automated electronic health record “sniffer” algorithms perform when implemented outside the institution in which they were developed. Particularly when algorithms analyze text documents, differences in the language used or how reports are generated may pose a major problem. One explanation for why the ARDS tools performed worse

when externally validated is that radiologists at different institutions report chest imaging findings in critically ill patients differently. A single-center study examining the free-text reports of chest computed tomographic scans found significant variation in the length of the findings and impression sections, as well as some variation in terminology (36). The authors’ conclusion was that developing algorithms to interpret the natural language of such reports would be a difficult task because of this variation. Because radiology reports are such an important source of data analyzed by most ARDS algorithms, tailoring these algorithms to institution-specific reporting practices may be necessary if such variation is present.

There are several limitations to our study. Like any systematic review, this study could be susceptible to bias related to which studies were included. We attempted to mitigate this by having a broad search strategy and reviewing citations of every included article. The review also included many studies that might be considered of low to moderate quality. We attempted to account for variation in study quality by using a standardized tool and the QUADAS-2 framework to identify these

biases. It is also unclear whether endpoints assessed (sensitivity and PPV) represent clinically relevant measures that have the potential to impact clinical outcomes. Good performance of these tools is important for limiting false alarms while also identifying all patients with ARDS, but only one study evaluated the effect of the tool on relevant clinical outcomes. Thus, we were not able to adequately assess the potential effect of these tools on important health outcomes.

Conclusions

Studies of electronic sniffer tools identifying patients with ARDS reported good to excellent performance in their derivation and internal validation cohorts but performed worse in a single external validation study. Potential biases were common across studies, particularly with regard to patient selection and reference standard application. There remains a critical need for robust research evaluating tools that assist clinicians in the identification of ARDS in clinical practice. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- 1 Rubenfeld GD, Caldwell E, Peabody E, Weaver J, Martin DP, Neff M, *et al*. Incidence and outcomes of acute lung injury. *N Engl J Med* 2005; 353:1685–1693.
- 2 Fröhlich S, Murphy N, Doolan A, Ryan O, Boylan J. Acute respiratory distress syndrome: underrecognition by clinicians. *J Crit Care* 2013; 28:663–668.
- 3 Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, *et al*; LUNG SAFE Investigators; ESICM Trials Group. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* 2016;315: 788–800.
- 4 Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A; Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000;342:1301–1308.
- 5 Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, *et al*; American Thoracic Society, European Society of Intensive Care Medicine, and Society of Critical Care Medicine. An official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine clinical practice guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2017; 195:1253–1263.
- 6 Sjoding MW, Hyzy RC. Recognition and appropriate treatment of the acute respiratory distress syndrome remains unacceptably low. *Crit Care Med* 2016;44:1611–1612.
- 7 Clark BJ, Moss M. The acute respiratory distress syndrome: dialing in the evidence? *JAMA* 2016;315:759–761.
- 8 Herasevich V, Yilmaz M, Khan H, Hubmayr RD, Gajic O. Validation of an electronic surveillance system for acute lung injury. *Intensive Care Med* 2009;35:1018–1023.
- 9 Herasevich V, Tsapenko M, Kojicic M, Ahmed A, Kashyap R, Venkata C, *et al*. Limiting ventilator-induced lung injury through individual electronic medical record surveillance. *Crit Care Med* 2011;39:34–39.
- 10 Chbat NW, Chu W, Ghosh M, Li G, Li M, Chiofolo CM, *et al*. Clinical knowledge-based inference model for early detection of acute lung injury. *Ann Biomed Eng* 2012;40:1131–1141.
- 11 Koenig HC, Finkel BB, Khalsa SS, Lanken PN, Prasad M, Urbani R, *et al*. Performance of an automated electronic acute lung injury screening system in intensive care unit patients. *Crit Care Med* 2011;39:98–104.
- 12 Azzam HC, Khalsa SS, Urbani R, Shah CV, Christie JD, Lanken PN, *et al*. Validation study of an automated electronic acute lung injury screening tool. *J Am Med Inform Assoc* 2009;16:503–508.
- 13 Solti I, Cooke CR, Xia F, Wurfel MM. Automated classification of radiology reports for acute lung injury: comparison of keyword and machine learning based natural language processing approaches. *Proceedings IEEE Int Conf Bioinformatics Biomed* 2009;2009:314–319.
- 14 Yetisgen-Yildiz M, Bejan CA, Wurfel MM. Identification of patients with acute lung injury from free-text chest x-ray reports. In: *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing (BioNLP 2013)*. August 4–9, 2013, Sofia, Bulgaria, pp. 10–17.
- 15 Reamaron N, Sjoding MW, Lin K, Iwashyna TJ, Najarian K. Accounting for label uncertainty in machine learning for detection of acute respiratory distress syndrome. *IEEE J Biomed Health Inform* 2019;23: 407–415; DOI: 10.1109/JBHI.2018.2810820.
- 16 McKown AC, Brown RM, Ware LB, Wanderer JP. External validity of electronic sniffers for automated recognition of acute respiratory distress syndrome. *J Intensive Care Med* [online ahead of print] 24 Jul 2017; DOI: 10.1177/0885066617720159.

- 17 Halamka JD, Tripathi M. The HITECH era in retrospect. *N Engl J Med* 2017;377:907–909.
- 18 Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, *et al.* Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18.
- 19 Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469–476.
- 20 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.*; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529–536.
- 21 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–381.
- 22 Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care* 2015;19:285.
- 23 Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, *et al.*; PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
- 24 Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61: 1085–1094.
- 25 Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol* 2015;3:32.
- 26 Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care* 2010;19:28–34; quiz 35.
- 27 Serpa Neto A, Oliveira Cardoso S, Manetta JA, Moura Pereira VG, Crepaldi Espósito D, de Oliveira Prado Pasqualucci M, *et al.* Association between use of lung-protective ventilation with lower tidal volumes and clinical outcomes among patients without acute respiratory distress syndrome: a meta-analysis. *JAMA* 2012;308: 1651–1659.
- 28 Sjoding MW, Gong MN, Haas CF, Iwashyna TJ. Evaluating delivery of low tidal volume ventilation in six ICUs using electronic health record data. *Crit Care Med* 2019;47:56–61.
- 29 Sjoding MW. Translating evidence into practice in acute respiratory distress syndrome: teamwork, clinical decision support, and behavioral economic interventions. *Curr Opin Crit Care* 2017;23: 406–411.
- 30 Sinha SS, Sjoding MW, Sukul D, Prescott HC, Iwashyna TJ, Gurm HS, *et al.* Changes in primary noncardiac diagnoses over time among elderly cardiac intensive care unit patients in the United States. *Circ Cardiovasc Qual Outcomes* 2017;10:e003616.
- 31 Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. *Chest* 1999;116:1347–1353.
- 32 Meade MO, Cook RJ, Guyatt GH, Groll R, Kachura JR, Bedard M, *et al.* Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2000;161:85–90.
- 33 Sjoding MW, Hofer TP, Co I, Courey A, Cooke CR, Iwashyna TJ. Interobserver reliability of the Berlin ARDS definition and strategies to improve the reliability of ARDS diagnosis. *Chest* 2018;153:361–367.
- 34 Sjoding MW, Hofer TP, Co I, McSparron JI, Iwashyna TJ. Differences between patients in whom physicians agree and disagree about the diagnosis of ARDS. *Ann Am Thorac Soc* 2019;16:258–264; DOI: 10.1513/AnnalsATS.201806-434OC.
- 35 Sjoding MW, Cooke CR, Iwashyna TJ, Hofer TP; Potential Effect on Clinical Study Results. Acute respiratory distress syndrome measurement error. *Ann Am Thorac Soc* 2016;13:1123–1128.
- 36 Huesch MD, Cherian R, Labib S, Mahraj R. Evaluating report text variation and informativeness: natural language processing of CT chest imaging for pulmonary embolism. *J Am Coll Radiol* 2018;15: 554–562.