



# Derivation and Validation of an Automated Search Strategy to Retrospectively Identify Acute Respiratory Distress Patients Per Berlin Definition

# Xuan Song<sup>1,2,3</sup>, Timothy J. Weister<sup>4</sup>, Yue Dong<sup>5</sup>, Kianoush B. Kashani<sup>6</sup> and Rahul Kashyap<sup>5\*</sup>

<sup>1</sup> Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic, Rochester, MN, United States,
<sup>2</sup> Intensive Care Unit, Liaocheng Cardiac Hospital Affiliated to Shandong First Medical University, Shandong, China,
<sup>3</sup> Intensive Care Unit, DongE Hospital Affiliated to Shandong First Medical University, Shandong, China,
<sup>4</sup> Anesthesia Clinical Research Unit, Mayo Clinic, Rochester, MN, United States,
<sup>5</sup> Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, United States,
<sup>6</sup> Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN, United States

### **OPEN ACCESS**

#### Edited by:

Marcelo Arruda Nakazone, Faculty of Medicine of São José do Rio Preto, Brazil

#### Reviewed by:

Rudolf Oliveira, Federal University of São Paulo, Brazil Carmen Silvia Valente Barbas, University of São Paulo, Brazil

> \*Correspondence: Rahul Kashyap kashyap.rahul@mayo.edu

#### Specialty section:

This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: 06 October 2020 Accepted: 16 February 2021 Published: 11 March 2021

#### Citation:

Song X, Weister TJ, Dong Y, Kashani KB and Kashyap R (2021) Derivation and Validation of an Automated Search Strategy to Retrospectively Identify Acute Respiratory Distress Patients Per Berlin Definition. Front. Med. 8:614380. doi: 10.3389/fmed.2021.614380 **Purpose:** Acute respiratory distress syndrome (ARDS) is common in critically ill patients and linked with serious consequences. A manual chart review for ARDS diagnosis could be laborious and time-consuming. We developed an automated search strategy to retrospectively identify ARDS patients using the Berlin definition to allow for timely and accurate ARDS detection.

**Methods:** The automated search strategy was created through sequential steps, with keywords applied to an institutional electronic medical records (EMRs) database. We included all adult patients admitted to the intensive care unit (ICU) at the Mayo Clinic (Rochester, MN) from January 1, 2009 to December 31, 2017. We selected 100 patients at random to be divided into two derivation cohorts and identified 50 patients at random for the validation cohort. The sensitivity and specificity of the automated search strategy were compared with a manual medical record review (gold standard) for data extraction of ARDS patients per Berlin definition.

**Results:** On the first derivation cohort, the automated search strategy achieved a sensitivity of 91.3%, specificity of 100%, positive predictive value (PPV) of 100%, and negative predictive value (NPV) of 93.1%. On the second derivation cohort, it reached the sensitivity of 90.9%, specificity of 100%, PPV of 100%, and NPV of 93.3%. The strategy performance in the validation cohort had a sensitivity of 94.4%, specificity of 96.9%, PPV of 94.4%, and NPV of 96.9%.

**Conclusions:** This automated search strategy for ARDS with the Berlin definition is reliable and accurate, and can serve as an efficient alternative to time-consuming manual data review.

Keywords: automation, electronic health records, acute respriatory distress syndrome, adult, ICU

## INTRODUCTION

Acute respiratory distress syndrome (ARDS) is an acute inflammatory lung injury which occurs in the absence of cardiogenic pulmonary edema and leads to increased pulmonary vascular permeability, increased extravascular lung water, and loss of aerated lung tissue (1). Estimates of the hospitalbased incidence of moderate to severe ARDS vary from 1.6 to 7.7% of all intensive care unit (ICU) admissions and 8.0-19.7% of all ventilated patients (2-5). Additionally, the reported population-based incidence of ARDS varies from 10.1 to 86.2 cases per 100,000 person-years (6-9). Overall mortality associated with ARDS is  $\sim$ 40%, according to the most recent observational studies (10, 11). Currently, there is no disease-specific pharmacotherapy to increase survival, and ARDS management remains supportive; therefore, the identification of ARDS with the Berlin definition in the ICU is critical, not only to identify the cases early and start primary and secondary prevention strategies but also to identify ARDS cases for potential clinical prospective studies.

Traditional paper charts have been rapidly replaced by electronic medical records (EMRs). The use of EMRs as a tool to reduce cost and improve safety has been increasing over the years in both clinical practice and health care research (12). For research, EMRs has moved medicine into the era of "big data," where an unprecedented amount of information can allow for evaluation and identification of risk factors at the population level. ARDS is often not documented in addition to respiratory failure terms. ICD-9 terms are not specific to ARDS and often code to non-specific conditions such as "respiratory distress;" therefore, it is difficult to identify ARDS cases for clinical study. A manual chart review for ARDS diagnosis could be laborious and time-consuming, so the effective and accurate use of EMRs, structured search strategies, and data capturing to identify cases are critical. In retrospective studies related to ARDS, an automated search strategy would be useful to identify cases in a timely fashion with high precision. Other similar search strategies from our team to identify sepsis, post-operative complications, acute kidney injury, and extubation failure have been developed and validated (13-16). These investigators found that by using such electronic search strategies, they were able to achieve high sensitivity and specificity in detecting patients with the syndromes and complications mentioned above.

In this study, our primary aim was to develop and validate a reliable electronic search strategy to identify cases with ARDS with the Berlin definition. Our secondary aim was to compare the sensitivity and specificity of our automated search strategy with a reference standard generated by a comprehensive, manual review of the medical record.

## MATERIALS AND METHODS

This study was approved by the Mayo Clinic Institutional Review Board (IRB) for the use of existing medical records of patients who gave prior research authorization.

## **Study Population**

The study population consisted of all patients admitted to the ICUs at Mayo Clinic in Rochester, MN from January 1, 2009 to December 31, 2017. Among this population, two groups of 50 patients were selected by purposeful sampling for derivation. This random purposeful sampling was done to include a random number of ARDS patients for higher yield. Both the manual reviewer and gold standard were blinded to the results of this sampling.We used separate revision cohorts to be able to test each change in the search strategy in a different group of patients, and, therefore, be able to optimize the search strategy. An additional cohort of 50 random patients was selected for the validation cohort (**Figure 1**).

We used the Berlin definition of ARDS criteria (1). The Berlin definition partitions patients by PaO<sub>2</sub>/FiO<sub>2</sub> ratio into mild (PaO<sub>2</sub>/FiO<sub>2</sub> 200-300), moderate (PaO<sub>2</sub>/FiO<sub>2</sub> 100-199), and severe ARDS (PaO<sub>2</sub>/FiO<sub>2</sub> <100) and no longer includes the term "acute lung injury." This definition also clarifies several areas, including onset, which must be within 1 week of a known clinical insult or new or worsening respiratory symptoms; chest imaging, which must include bilateral opacities that are not fully explained by effusions, lobar collapse, or nodules; and origin of edema, which cannot be fully explained by cardiac failure or fluid overload and must be objectively evaluated (e.g., by echocardiography) if no apparent predisposing factor for ARDS is present. The Berlin definition also sets a minimum positive end-expiratory pressure (PEEP) level of 5 cm H<sub>2</sub>O during PaO<sub>2</sub>/FiO<sub>2</sub> determination because it has been recognized that changes in PEEP may reclassify patients from the current definition of ALI to ARDS (1).

## Manual Data Extraction Strategies

A manual review of patient EMRs was used for data extraction and ARDS adjudication. The manual reviewer was a practicing clinician, who reviewed the electronic medical charts with all available information. The reviewer assessed all included patients to identify patients who had ARDS per Berlin definition. The reviewer was not involved in the development or utilization of the automated electronic search strategy. Hence, the reviewer was not aware of the results of the automated search strategy.

We used the definition of Berlin ARDS criteria for manual chart review, and defined ARDS based on the presence of both of the following conditions simultaneously: (1) patients with  $PaO_2/FiO_2$  ratio <300,  $PEEP \ge 5 \text{ cm } H_2O$ , bilateral infiltrate or edema per chest X-ray, and (2) the presence of at least one risk factor for ARDS (i.e., sepsis/septic shock, pneumonia, pancreatitis, trauma, aspiration, multiple transfusion, drug overdose, and shock). We used the final adjudicated ARDS cases based on this process as the gold standard for the study.

## Automated Electronic Search Strategy

Data were used from Mayo Clinic ICU DataMart and Unified Data Platform, which are extensive data warehouses containing a near real-time normalized replica of Mayo Clinic's EMRs. These databases contain patient information, their laboratory test results, and clinical and pathological information from sources within the institution and have been previously validated (17,







18). Ventilator parameters (such as PEEP) were captured from the ventilators.

The automated electronic search strategy for identifying ARDS patients per Berlin definition was developed in the following sequential steps (**Figure 2**). First, patients were excluded who did not provide research authorization, along with those <18 years old. Second, ARDS patients were identified according to the following criteria: (1) PEEP  $\geq$ 5 during the ICU stay (this is "time zero") and identified the ventilator mode nearest (limited to ICU areas—procedures excluded); (2) Partial pressure of oxygen to fraction of inspired oxygen (PaO<sub>2</sub>/FiO<sub>2</sub>) ratio  $\leq$ 300, P/F ratios were first established based on matched PaO<sub>2</sub> and FiO<sub>2</sub> from labs. If FiO<sub>2</sub> labs were missing, FiO<sub>2</sub> from vital signs within ±15 min (nearest) the PaO<sub>2</sub> value were used; (3) Chest X-ray within 12 h and review radiology report for any of the

following combinations: bilateral infiltrates, bilateral opacities, or bilateral edema. If one of them was present, it was considered a positive radiology report. If all 3 criteria in second step were positive, then they were classified as potential ARDS patients. Third, patients with diagnosis of Idiopathic Pulmonary Fibrosis (IPF)/Interstitial Lung Disease (ILD)/pulmonary fibrosis were excluded by Charlson comorbidity index search. Fourth, patients with invasive mechanical ventilation <12 h were excluded, and the duration of mechanical ventilation was searched according to our previously published algorithm (19). Fifth, ARDS risk factors were searched for (i.e., sepsis/septic shock, pneumonia, aspiration, pancreatitis, trauma, drug overdose, shock, and multiple transfusions), and the search strategy for each risk factor was defined. Finally, patients with cardiogenic pulmonary edema, cardiogenic shock, and positive acute decompensated

TABLE 1   Automated search strategy sensitivity and specificity for ARDS per
Berlin definition.

	ARDS per Berlin definition			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Derivation cohort 1	91.3	100	100	93.1
Derivation cohort 2	90.9	100	100	93.3
Validation cohort	94.4	96.9	94.4	96.9

ARDS, acute respiratory distress syndrome; PPV: positive predictive value; NPV, negative predictive value.

heart failure during ICU admission were excluded, and patients with cardiogenic pulmonary edema risk factors were also excluded by clinical note searches (20), and these risk factors include history of Coronary Heart Disease (CAD), chronic heart failure (CHF), and New ST-changes/Left bundle branch block (Electrocardiography query within  $\pm 24$  h of 1st PEEP  $\geq 5$ ). The automated search algorithms were validated in comparison with the gold standard obtained by manual review.

### **Statistical Analysis**

For automation process, the only applicable analysis is sensitivity and specificity for a nominal variable (ARDS, yes/no). The sensitivity and specificity of the search algorithms were calculated by comparing the results to the gold standard obtained by manual review of the charts. We used JMP Pro 14 statistical software (SAS Institute Inc., Cary, NC, USA). *P*-values 0 < 0.05 were considered statistically significant.

## RESULTS

Between January 1, 2009, and December 31, 2017, 140,098 adult patients with research authorization were admitted to the participating ICUs, and 2,281 patients met the ARDS Berlin definition according to the automated search strategy. A total of 100 patients were chosen after purposeful sampling to be included in the two derivation cohorts, and an additional 50 patients were selected for the validation cohort.

The automated search strategy identified ARDS patients with a sensitivity of 91.3%, specificity of 100%, positive predictive value (PPV) of 100%, and negative predictive value (NPV) of 93.1% in the first derivation cohort (Table 1). Disagreements between the automated search strategy and the manual review were observed in 2 patients in this data subset, both false negatives. In one of the cases, ARDS was missed by the digital algorithm as PaO<sub>2</sub>/FiO<sub>2</sub> ratio, and chest X-ray were not found, while in the other case sepsis developed >72 h after ICU admission. In the second derivation cohort (Table 1), the automated search strategy reached a sensitivity of 90.9%, specificity of 100%, PPV of 100%, and NPV of 93.3%. Disagreements between the automated search strategy and the gold standard occurred in 2 patients, both false negatives. The reasons for these false-negative cases were identical to those in the first derivation cohort. The manual vs. automated cohorts had same baseline characteristics as they were exact same cohorts (data not shown).

In the validation cohort, the automatic search strategy yielded a sensitivity of 94.4%, specificity of 96.9%, PPV of 94.4%, and NPV of 96.9% (**Table 1**). Disagreements between the automated search strategy and the reference standard occurred in 2 patients, both false negatives. One case was due to missing  $PaO_2/FiO_2$  ratio and chest X-ray, and the other case was because the patient used home Bilevel Positive Airways Pressure (BiPAP), and thus PEEP was not electronically recorded.

## DISCUSSION

In this study, we demonstrated an automated search strategy for ARDS that could effectively and accurately identify patients based on the accepted clinical definition (i.e., Berlin definition, among ICU patients). Several previously automated search strategies have been described in the literature (13–15, 21, 22); however, to date, the Berlin definition has not been used as a digital signature of ARDS patients.

As EMR utilization continues on an upward trajectory, the volume of available information to generate and validate digital signatures of different clinical syndromes in ICUs has grown. The accumulation of vast amounts of data provides opportunities to improve the processes of care and treatment. Manual chart review for ARDS diagnosis would likely be laborious and time-consuming; considering the significant shortage of human resources in clinical investigations, there seems to be a vital need to use EMRs for syndrome detection. Traditional ICD code searches for such conditions may not be completely sensitive or specific (18, 23), and changes in coding guidelines make them even less reliable. Thus, the development of automated search strategies can prove useful for clinical and research purposes.

Our study has several strengths. To our knowledge, this is the first study regarding the development and validation of an automated search strategy within EMRs for the identification of ARDS patients per Berlin definition. It is a valuable contribution in that it allows for a quick and reliable way to identify cases of ARDS retrospectively, which will ultimately enable pragmatic research on large cohorts of patients using existing EMRs. Using automated search strategies overcomes the barrier of timeconsuming manual review and mitigates human errors that occur during manual data extraction. This electronic signature provides strong support for educational and research activities and demonstrates a simple yet effective method that can be applied to other clinical conditions.

Several limitations of our study should be acknowledged. First, the accuracy of the EMR depends on the precision of written clinical notes. As with manual chart review, we assumed clinical documentation is accurate, while errors in the documentation are possible. In our institution, periodic quality checks on clinical notes are done with frequent audits. Therefore, we believe the impact of documentation errors in this digital signature is minimal. Secondly, this is an automatic search strategy to retrospectively identify ARDS patients per the Berlin definition. It cannot identify these patients in real time, but it lays a foundation for the development of ARDS software to identify ARDS patients in real time in the future. Finally, Mayo Clinic EMR structure may be different from other institutions, thus limiting its use. The generalization of the findings is limited at this point, given that no external validation was performed. Future studies should evaluate the method in different EMR systems and in different populations.

## CONCLUSIONS

Here we reported the derivation and validation of an automated electronic search query algorithm for identifying ARDS patients according to the Berlin definition. Sensitivity and specificity approached 100% in this study and may continue to improve as processes develop around electronic notes searches, following the iterative development model previously described. The development of this type of automated search strategy is widely applicable to clinical research; it may improve the efficiency and accuracy of patient identification, thus furthering knowledge on the subjects and potentially improving outcomes. Ultimately, it may enable pragmatic research on large cohorts of patients using existing EMRs, for early and rapid identification of ARDS patients.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## REFERENCES

- Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*. (2012) 307:2526–33. doi: 10.1001/jama.2012.5669
- Estenssoro E, Dubin A, Laffaire E, Canales H, Saenz G, Moseinco M, et al. Incidence, clinical course, and outcome in 217 patients with acute respiratory distress syndrome. *Crit. Care Med.* (2002) 30:2450–6. doi: 10.1097/00003246-200211000-00008
- Roupie E, Lepage E, Wysocki M, Fagon JY, Chastre J, Dreyfuss D, et al. Prevalence, etiologies and outcome of the acute respiratory distress syndrome among hypoxemic ventilated patients. SRLF Collaborative Group on Mechanical Ventilation. Societe de Reanimation de Langue Francaise. *Intensive Care Med.* (1999) 25:920–9. doi: 10.1007/s001340050983
- Caser EB, Zandonade E, Pereira E, Gama AM, Barbas CS. Impact of distinct definitions of acute lung injury on its incidence and outcomes in Brazilian ICUs: prospective evaluation of 7,133 patients. *Crit. Care Med.* (2014) 42:574– 82. doi: 10.1097/01.ccm.0000435676.68435.56
- Brun-Buisson C, Minelli C, Bertolini G, Brazzi L, Pimentel J, Lewandowski K, et al. Epidemiology and outcome of acute lung injury in European intensive care units. Results from the ALIVE study. *Intensive Care Med.* (2004) 30:51– 61. doi: 10.1007/s00134-003-2022-6
- Bersten AD, Edibam C, Hunt T, Moran J. Incidence and mortality of acute lung injury and the acute respiratory distress syndrome in three Australian States. Am. J. Respir. Crit. Care Med. (2002) 165:443– 8. doi: 10.1164/ajrccm.165.4.2101124
- Hughes M, MacKirdy FN, Ross J, Norrie J, Grant IS. Acute respiratory distress syndrome: an audit of incidence and outcome in Scottish intensive care units. *Anaesthesia*. (2003) 58:838–45. doi: 10.1046/j.1365-2044.2003.03287.x
- Li G, Malinchoc M, Cartin-Ceba R, Venkata CV, Kor DJ, Peters SG, et al. Eight-year trend of acute respiratory distress syndrome: a population-based study in Olmsted County, Minnesota. Am. J. Resp. Crit. Care Med. (2011) 183:59–66. doi: 10.1164/rccm.201003-0436OC
- 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) 353:1685–93. doi: 10.1056/NEJMoa050333

## ETHICS STATEMENT

This study was approved by the Mayo Clinic Institutional Review Board (IRB) for the use of existing medical records of patients who gave prior research authorization. Informed consent was obtained from all patients included in the study.

# **AUTHOR CONTRIBUTIONS**

RK conceptualized the study. XS and TW designed the study. XS performed data analysis and drafted the manuscript. YD and KK critically revised the article for important intellectual content. All authors read and approved the final manuscript.

## FUNDING

The Department of Anesthesiology and Division of Pulmonary and Critical Care Medicine at Mayo Clinic (Rochester, MN) supported this work with no direct financial support.

# ACKNOWLEDGMENTS

We thank the Anesthesia Clinical Research Unit for their help with electronic data abstraction.

- Villar J, Ambros A, Soler JA, Martinez D, Ferrando C, Solano R, et al. Age, PaO2/FIO2, and plateau pressure score: a proposal for a simple outcome score in patients with the acute respiratory distress syndrome. *Crit. Care Med.* (2016) 44:1361–9. doi: 10.1097/CCM.00000000001653
- Villar J, Blanco J, Anon JM, Santos-Bouza A, Blanch L, Ambros A, et al. The ALIEN study: incidence and outcome of acute respiratory distress syndrome in the era of lung protective ventilation. *Intensive Care Med.* (2011) 37:1932– 41. doi: 10.1007/s00134-011-2380-4
- Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. J. Am. Med. Inform. Assoc. (2011) 18:169–72. doi: 10.1136/jamia.2010.007229
- Tien M, Kashyap R, Wilson GA, Hernandez-Torres V, Jacob AK, Schroeder DR, et al. Retrospective derivation and validation of an automated electronic search algorithm to identify post operative cardiovascular and thromboembolic complications. *Appl. Clin. Inform.* (2015) 6:565– 76. doi: 10.4338/ACI-2015-03-RA-0026
- Rishi MA, Kashyap R, Wilson G, Hocker S. Retrospective derivation and validation of a search algorithm to identify extubation failure in the intensive care unit. *BMC Anesthesiol.* (2014) 14:41. doi: 10.1186/1471-225 3-14-41
- Dhungana P, Serafim LP, Ruiz AL, Bruns D, Weister TJ, Smichney NJ, et al. Machine learning in data abstraction: a computable phenotype for sepsis and septic shock diagnosis in the intensive care unit. *World J. Crit. Care Med.* (2019) 8:120–6. doi: 10.5492/wjccm.v8.i7.120
- Ahmed A, Vairavan S, Akhoundi A, Wilson G, Chiofolo C, Chbat N, et al. Development and validation of electronic surveillance tool for acute kidney injury: a retrospective analysis. J. Crit. Care. (2015) 30:988– 93. doi: 10.1016/j.jcrc.2015.05.007
- Herasevich V, Pickering BW, Dong Y, Peters SG, Gajic O. Informatics infrastructure for syndrome surveillance, decision support, reporting, and modeling of critical illness. *Mayo Clin. Proc.* (2010) 85:247–54. doi: 10.4065/mcp.2009.0479
- Singh B, Singh A, Ahmed A, Wilson GA, Pickering BW, Herasevich V, et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. *Mayo Clin. Proc.* (2012) 87:817–24. doi: 10.1016/j.mayocp.2012.04.015

- Weister T, Singhal A, Marquez A, Smischney N, Kashyap R. Refinement of a computable phenotype for initiation of mechanical ventilation in intensive care unit. Am. J. Respir. Crit. Care Med. (2020) 201:A1454.
- Kashyap R, Sarvottam K, Wilson GA, Jentzer JC, Seisa MO, Kashani KB. Derivation and validation of a computable phenotype for acute decompensated heart failure in hospitalized patients. *BMC Med. Inform. Decis. Mak.* (2020) 20:85. doi: 10.1186/s12911-020-1092-5
- Amra S, O'Horo JC, Singh TD, Wilson GA, Kashyap R, Petersen R, et al. Derivation and validation of the automated search algorithms to identify cognitive impairment and dementia in electronic health records. J. Crit. Care. (2017) 37:202–5. doi: 10.1016/j.jcrc.2016.09.026
- Guru PK, Singh TD, Passe M, Kashani KB, Schears GJ, Kashyap R. Derivation and validation of a search algorithm to retrospectively identify CRRT initiation in the ECMO patients. *Appl. Clin. Inform.* (2016) 7:596– 603. doi: 10.4338/ACI-2015-12-RA-0183
- Siew ED, Davenport A. The growth of acute kidney injury: a rising tide or just closer attention to detail? *Kidney Int.* (2015) 87:46–61. doi: 10.1038/ki.2014.293

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Song, Weister, Dong, Kashani and Kashyap. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.