Differences between Patients in Whom Physicians Agree and Disagree about the Diagnosis of Acute Respiratory Distress Syndrome

Michael W. Sjoding^{1,2}, Timothy P. Hofer^{1,3}, Ivan Co^{1,4}, Jakob I. McSparron¹, and Theodore J. Iwashyna^{1,3,5,6*}

¹Department of Internal Medicine, ²Center for Computational Medicine and Bioinformatics, and ⁴Department of Emergency Medicine, University of Michigan, Ann Arbor, Michigan; ³VA Center for Clinical Management Research, Ann Arbor, Michigan; ⁵Institute for Social Research, Ann Arbor, Michigan; and ⁶Section Editor, *AnnalsATS*

Abstract

Rationale: Because the Berlin definition of acute respiratory distress syndrome (ARDS) has only moderate reliability, physicians disagree about the diagnosis of ARDS in some patients. Understanding the clinical differences between patients with agreement and disagreement about the diagnosis of ARDS may provide insight into the epidemiology and pathophysiology of this syndrome, and inform strategies to improve the reliability of ARDS diagnosis.

Objectives: To characterize patients with diagnostic disagreement about ARDS among critical-care-trained physicians and compare them with patients with a consensus that ARDS developed.

Methods: Patients with acute hypoxemic respiratory failure (arterial oxygen tension/pressure [PaO_2]/fraction of inspired oxygen [FiO_2] < 300 during invasive mechanical ventilation) were independently reviewed for ARDS by multiple critical-care physicians and categorized as consensus-ARDS, disagreement about the diagnosis, or no ARDS.

Results: Among 738 patients reviewed, 110 (15%) had consensus-ARDS, 100 (14%) had disagreement, and 528 (72%) did not have ARDS. ARDS diagnosis rates ranged from 9% to 47% across clinicians. Patients with disagreement had baseline comorbidity

rates similar to those of patients with consensus-ARDS, but lower rates of ARDS risk factors and less severe measures of lung injury. Mean days of severe hypoxemia ($\text{PaO}_2/\text{FiO}_2 < 100$) were 3.2 (95% confidence interval [CI], 2.6–3.9), 2.0 (95% CI, 1.5–2.4), and 0.8 (95% CI, 0.7–0.9) among patients with consensus-ARDS, disagreement, and no ARDS, respectively. Hospital mortality was 37% (95% CI, 28–46%), 35% (95% CI, 26–44%), and 19% (95% CI, 15–22%) across groups. Simple combinations of specific ARDS risk factors and lowest $\text{PaO}_2/\text{FiO}_2$ value could effectively discriminate patients (area under the receiver operating characteristic curve = 0.90; 95% CI, 0.88–0.92). For example, 63% of patients with pneumonia, shock, and $\text{PaO}_2/\text{FiO}_2 < 110$ had consensus-ARDS, whereas 100% of patients without pneumonia or shock and $\text{PaO}_2/\text{FiO}_2 > 180$ did not have ARDS.

Conclusions: Disagreement about the diagnosis of ARDS is common and can be partly explained by the difficulty of dichotomizing patients along a continuous spectrum of ARDS manifestations. Considering both the presence of key ARDS risk factors and hypoxemia severity can help guide clinicians in identifying patients with diagnosis of ARDS agreed upon by a consensus of physicians.

Keywords: acute lung injury; diagnosis; epidemiology; risk factors

(Received in original form June 28, 2018; accepted in final form October 12, 2018)

*Participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

Supported by grants from the U.S. National Heart, Lung, and Blood Institute K01HL136687 (M.W.S.) and the Department of Veterans Affairs Health Services Research and Development Services IIR 13-079 (T.J.I.).

Author Contributions: M.W.S. contributed to the study design, analysis and interpretation of data, writing and revising the manuscript, and approval of the final manuscript. T.P.H., I.C., J.I.M., and T.J.I. contributed to the study design, analysis and interpretation of data, revising the manuscript for important intellectual content, and approval of the final manuscript.

Correspondence and requests for reprints should be addressed to Michael W. Sjoding, M.D., M.Sc., G027W Building 16 NCRC, 2800 Plymouth Road, SPC 2800, 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 2, pp 258–264, Feb 2019

Copyright © 2019 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201806-434OC

Internet address: www.atsjournals.org

Although the clinical manifestations of acute respiratory distress syndrome (ARDS) were described over 50 years ago (1), efforts to codify these findings into a set of valid and reliable diagnostic criteria have had more limited success (2, 3). The current Berlin definition of ARDS has only moderate interrater reliability when applied to patients with respiratory failure (4). In addition, only 45% of patients who met the Berlin ARDS definition were found to have the histopathological hallmarks of ARDS in an autopsy series (5). The lack of a valid and reliable ARDS definition is often cited as a major reason for the underrecognition of ARDS in clinical practice and slow progress in advancing ARDS research (6–9).

Because the current Berlin ARDS definition has only moderate reliability, two critical-care-trained physicians reviewing the same group of patients with acute respiratory failure are likely to disagree about whether some of the patients have ARDS (4). Little is known about the patients with disagreement among physicians about the diagnosis of ARDS, or how these patients differ from those with a consensus among physicians that ARDS has developed. Understanding the differences in clinical characteristics among patients with agreement and disagreement about the diagnosis of ARDS may provide new insight into ARDS epidemiology and pathophysiology, and also might suggest strategies for improving the reliability of diagnosis in practice. If patients with diagnostic disagreement have uniquely identifiable clinical characteristics, such patients could be targeted for further observation or evaluation before diagnosis. In contrast, if patients with disagreement lack unique characteristics, but rather have a more intermediate presentation, this would suggest that difficulty in dichotomizing patients along a continuous spectrum of ARDS manifestations is the major reason for disagreement. If the latter is the case, understanding which physiologic parameters (e.g., arterial oxygen tension/pressure [PaO_2]/fraction of inspired oxygen [FiO_2]) or combinations are best aligned with this spectrum could help improve stratification of patients moving forward.

In this study, we sought to characterize patients with diagnostic disagreement about ARDS among critical-care-trained physicians and compare them with patients

with a consensus that ARDS developed. We hypothesized that these groups would have distinct clinical characteristics at the time when ARDS was first suspected, and that available clinical data could help differentiate them from other patients with hypoxemic respiratory failure.

Methods

Patient Cohort and ARDS Reviews

Consecutive patients were independently reviewed for the development of ARDS by two to four intensive-care-trained physicians if they were admitted between January and March 2016 or between October 2016 and June 2017, and had a $\text{PaO}_2/\text{FiO}_2 < 300$ while receiving invasive mechanical ventilation at a single tertiary-care center. Patients in the medical, surgical, cardiac, and trauma/burn intensive-care units were included. Thirteen clinicians (six pulmonary critical-care faculty and seven advanced pulmonary critical-care fellows) reviewed the patients for ARDS.

Clinicians retrospectively reviewed each patient's hospitalization records, including all chest imaging studies (chest X-rays and computed tomography scans), to determine whether ARDS developed during the first week of hospitalization according to the Berlin definition (2). They evaluated whether patients had specific ARDS risk factors using standard definitions (10) and considered all information available during the hospitalization, including a patient's evolution over time and response to therapy. Clinicians also identified the specific time when all ARDS criteria were met. Clinicians used a standardized ARDS evaluation tool to guide each review, which prompted them to determine whether a patient met each Berlin ARDS criterion, followed by an overall assessment of ARDS. Further details of the ARDS review process have been previously described (4).

Rates of ARDS diagnosis across individual clinicians, agreement rates among clinicians, and clinician characteristics potentially associated with differing rates of ARDS diagnosis and agreement were evaluated. Overall variation in the diagnosis of ARDS attributable to clinicians, rather than to differences between patients, was also quantified. These analyses are further described in the online supplement.

ARDS Diagnosis Categories: Consensus-ARDS, Disagreement, and No ARDS

We used a multilevel modeling strategy to estimate each patient's probability of ARDS based on the results of physician reviews, and then categorized the patients as consensus-ARDS, disagreement, or no ARDS based on the probability derived from this model. Technical details of the modeling approach and sample code are described in the online supplement. This approach allowed us to account for differences in the total number of reviews performed for each patient and systematic differences between clinicians in their reviews. Patients with greater disagreement among clinicians had a more intermediate probability of ARDS in the model. Patients with greater agreement that ARDS developed had a higher estimated ARDS probability, whereas patients with agreement that ARDS did not develop had low probability. Patients with a probability of ARDS above 80% were defined as consensus-ARDS, patients with a probability of 20–80% were defined as disagreement, and patients with a probability below 20% were defined as no ARDS.

Clinical Characteristics and Outcome Variables

Baseline demographic information and selected comorbidities were collected for all patients. Physiologic data were extracted from the electronic health record during the first 24-hour period after ARDS onset for those suspected of developing ARDS by any physician, or during a randomly matched 24-hour period of mechanical ventilation for those without ARDS (matched based on mechanical ventilation day). The ARDS onset time was defined as the earliest time a clinician believed all ARDS criteria were met. Physiologic characteristics abstracted during the 24-hour window included the lowest calculated $\text{PaO}_2/\text{FiO}_2$ or $\text{PaO}_2/\text{FiO}_2$ derived from arterial oxygen saturation/ FiO_2 (11), highest measured plateau pressure, lowest quasi-static lung compliance, and highest estimated dead space fraction (12). Because data can be incorrectly recorded in the electronic health record, outlier values were dropped if they were three standard deviations from the mean. Missing data on plateau pressure (9%), lung compliance (11%), and dead space fraction (8%) during the 24-hour window

were imputed using multiple imputation (13). Outcome variables examined included days with severe hypoxemia ($\text{PaO}_2/\text{FiO}_2 < 100$), ventilator-free days through day 28, and in-hospital mortality.

Statistical Analysis

First, we compared demographics, comorbidities, risk factors, chest imaging findings, and outcomes across patients grouped by ARDS diagnosis (consensus-ARDS, disagreement, or no ARDS) using ANOVA for continuous variables, chi-square for categorical variables, and Poisson regression for days of severe hypoxemia or ventilator-free days. Patients were assigned to these groups based on their ARDS probability as described above. To determine whether differences were statistically significant, we used the Bonferroni correction to the *P* value, which lowers the *P* value threshold necessary to conclude that a result is statistically significant based on the number of comparisons performed to minimize false discovery (14). A *P* value of 0.05/3 was considered significant because three comparisons were performed across groups.

Second, to understand whether the degree of hypoxemia was a surrogate for ARDS diagnosis, we performed a cross-tabulation of patients grouped by ARDS diagnosis (consensus-ARDS, disagreement, or no ARDS) and categories of hypoxemia (mild: >200 ; moderate: $100\text{--}200$; severe: < 100) and examined the overlap between these two methods of categorizing patients. We compared mortality rates across ARDS diagnosis and hypoxemia categories, and calculated the mortality discrimination of each. Mortality discrimination was compared using the area under the receiver operating characteristic curve (AUC).

Third, we compared measures of lung injury severity ($\text{PaO}_2/\text{FiO}_2$, plateau pressure, lung compliance, and dead space fraction) among patients grouped by ARDS diagnosis (consensus-ARDS, disagreement, or no ARDS). We generated area plots to illustrate the proportion of patients in each group over ranges of each physiologic measure. These plots were generated by first fitting a multinomial logistic regression model, where the ARDS diagnosis category (consensus-ARDS, disagreement, or no ARDS) was the outcome and the physiologic measure was the sole predictor. Restricted cubic splines with three knots

were used to model potential nonlinear relationships between the predictor and outcome. After each model was fitted, the proportion of patients in each group was estimated across ranges of values for each physiologic measure.

Fourth, because of the significant relationship among $\text{PaO}_2/\text{FiO}_2$, pneumonia, shock, and ARDS diagnosis identified by the current and previous studies (10, 15), we further examined how the relationship between $\text{PaO}_2/\text{FiO}_2$ and ARDS diagnosis was modified by the presence of these other factors. For this analysis, the unit of analysis was the individual ARDS reviews performed by clinicians. We fit a multilevel logistic regression of reviews nested within patient, where ARDS diagnosis was the outcome. Model covariates included the $\text{PaO}_2/\text{FiO}_2$, a squared $\text{PaO}_2/\text{FiO}_2$ term, interaction terms with pneumonia and shock, and an adjustment for the clinical reviewer. We then estimated the mean marginal probability of ARDS diagnosis across ranges of $\text{PaO}_2/\text{FiO}_2$ in the presence or absence of pneumonia and shock.

Finally, we derived simple decision rules that could be used at the bedside to approximate the collective judgment of multiple clinicians and identify patients with agreement or disagreement about the diagnosis of ARDS. We derived rules using the ARDS risk factors pneumonia and shock, and the most common measure of lung injury severity ($\text{PaO}_2/\text{FiO}_2$ ratio). For this analysis, patients were again analyzed based on ARDS diagnosis categories (consensus-ARDS, disagreement, no ARDS). Decision rules were derived using a classification and regression tree, which attempts to find rules that split the data into subsets that maximize the purity of each subset, i.e., maximizing the number of patients in one group while minimizing the number of those from other groups. Gini impurity was used to derive the best splits. The decision rule's performance was evaluating using the AUC, and its generalization performance was estimated by drawing 1,000 bootstrap samples of the data to determine 95% confidence intervals.

All analyses were performed in Stata 14.0 and Python 3.6. The institutional review board of the University of Michigan approved this study with a waiver of informed consent from study participants.

Results

The cohort included 738 patients with a $\text{PaO}_2/\text{FiO}_2 < 300$ during invasive mechanical ventilation reviewed by multiple clinicians, with 1,863 total reviews performed.

Among these patients, 110 (14.9%) had consensus-ARDS, 100 had disagreement about the diagnosis (13.6%), and 528 did not have ARDS (71.5%) (Table 1). There was notable variation among the 13 clinicians in ARDS diagnosis rates, ranging from 8% to 47% (Figure E1 in the online supplement), although diagnostic agreement rates among clinicians ranged from 69% to 83%, and only 7% of the overall variation in ARDS diagnosis was attributable to clinicians rather than patients. ARDS diagnosis rates were 20% among fellows ($n = 7$), 30% among clinical faculty ($n = 3$), and 34% among faculty involved in ARDS clinical trials ($n = 3$), although these results were not statistically different (Table E1).

On hospital presentation, patients with disagreement about the diagnosis of ARDS had no significant differences in comorbidities that could potentially mask or confuse the diagnosis, including rates of obesity, congestive heart failure, and chronic kidney disease (Table 1). However, rates of specific ARDS risk factors were significantly different across groups. Pneumonia was present in 75% of patients with consensus-ARDS, 49% of patients with disagreement, and 19% of patients without ARDS (significantly different across pairwise comparisons; see Table E2 for *P* values). Noncardiogenic shock was present in 62% of patients with consensus-ARDS, 43% of patients with disagreement, and 23% of patients without ARDS, and also significantly different across groups. Rates of nonpulmonary sepsis were higher in patients with consensus-ARDS (35%) or disagreement (32%) than in patients without ARDS (17%), but not significantly different between patients with consensus-ARDS and disagreement.

Chest imaging studies more clearly demonstrated bilateral airspace disease in patients with consensus-ARDS compared with other groups (Table 1): 70% of the chest imaging studies showed bilateral airspace disease in patients with consensus-ARDS, 23% showed bilateral airspace disease in patients with disagreement, and 8% showed bilateral airspace disease in patients without ARDS. There was no difference in the rates of chest computed tomography scans across groups.

Table 1. Clinical characteristics and outcomes of patients with hypoxemic respiratory failure

Patient Categories	Consensus-ARDS	Disagreement	No ARDS
Number (%)	<i>n</i> = 110 (14.9)	<i>n</i> = 100 (13.6)	<i>n</i> = 528 (71.5%)
Age, mean	54	58	58
Male, %	59	55	64
Caucasian, %	86	79	78
Comorbidities, %			
Morbid obesity	16	21	25
Congestive heart failure	29	32	33
Chronic kidney disease	27	34	33
Chronic liver disease	13	14	15
ARDS risk factors, %			
None	2	8	36
Pneumonia	75	49	19
Aspiration	18	18	14
Sepsis (nonpulmonary source)	35	32	17
Noncardiogenic shock	62	43	23
Major trauma	8	5	11
Chest imaging			
Number of chest images, mean	7	6	5
Positive, %*	70	23	8
Chest computed tomography scan performed	39	36	30
Clinical outcomes			
Days of severe hypoxemia, mean	3.2	2.0	0.8
Ventilator-free days, mean	14	16	21
Hospital mortality, %	37	35	19

Definition of abbreviation: ARDS = acute respiratory distress syndrome.

Patients were grouped by whether there was agreement among physicians that ARDS developed (consensus-ARDS), disagreement about the diagnosis, or agreement that ARDS did not develop.

*Positive: rate at which all physicians reviewing a chest imaging study believed bilateral airspace opacities were present consistent with a diagnosis of ARDS.

Clinical outcomes, including days with severe hypoxemia, ventilator-free days, and mortality, were different across groups, but only days with severe hypoxemia and ventilator-free days were statistically significantly different across all pairwise comparisons. Mean days of severe hypoxemia were 3.2 (95% CI, 2.6–3.9), 2.0 (95% CI, 1.5–2.4), and 0.8 (95% CI, 0.7–0.9) among patients with consensus-ARDS, disagreement, and no ARDS, respectively. Hospital mortality was 37% (95% CI, 28–46%), 35% (95% CI, 26–44%), and 19% (95% CI, 15–22%) across groups.

There was minimal concordance between a patient's degree of hypoxemia and the diagnosis of ARDS (Table 2). When patients were categorized as consensus-ARDS, disagreement, or no ARDS, or categorized as mild, moderate, or severe hypoxemia based on $\text{PaO}_2/\text{FiO}_2$, there was 31% agreement between the categorizations. For example, among patients with severe hypoxemia ($\text{PaO}_2/\text{FiO}_2 < 100$), 26% of patients had consensus-ARDS, 19% had disagreement, and 55% had no ARDS.

Neither the ARDS diagnosis categories nor the hypoxemia categories provided good discrimination of mortality (AUC = 0.60 and 0.59, respectively; Figures E2 and E3).

Physiologic parameters during the first 24 hours after ARDS onset and matched periods during mechanical ventilation for patients without ARDS revealed that patients with consensus-ARDS had worse

lung injury than other patients (Figure 1). However, there were no clear cutoff values that could distinguish patients with consensus-ARDS from patients without ARDS. For example, when the dead space fraction was above 80%, 26% of the patients had consensus-ARDS and 58% did not have ARDS.

The relationship between the $\text{PaO}_2/\text{FiO}_2$ ratio and ARDS diagnosis was significantly modified based on the presence or absence of the ARDS risk factors pneumonia and shock (Figure 2). Accounting for this interaction yielded an ARDS model with good discrimination (AUC = 0.84) and reasonable calibration across ARDS risk deciles (Figure E4). In the setting of pneumonia and shock, the likelihood of ARDS diagnosis increased rapidly as $\text{PaO}_2/\text{FiO}_2$ worsened. However, when pneumonia or shock was absent, the strength of the relationship between $\text{PaO}_2/\text{FiO}_2$ and ARDS diagnosis was minimal.

Simple decision rules derived using the two most common ARDS risk factors and $\text{PaO}_2/\text{FiO}_2$ (Table 3) had excellent discriminative performance overall (AUC = 0.90; 95% CI, 0.88–0.92). However, the rules better discriminated patients with consensus-ARDS (AUC = 0.86; 95% CI, 0.82 – 0.92) and patients without ARDS (AUC = 0.83; 95% CI, 0.80–0.86), compared with patients with disagreement (AUC = 0.73; 95% CI, 0.68–0.78) (Figure E3). For example, 63% of patients with pneumonia, shock, and a $\text{PaO}_2/\text{FiO}_2$ below 120 had consensus-ARDS. On the other hand, 100% of patients without pneumonia or shock, and a $\text{PaO}_2/\text{FiO}_2 > 180$ did not have ARDS. Decision rules that identified patients with the highest rates of disagreement included patients with pneumonia and $\text{PaO}_2/\text{FiO}_2 < 120$ but without

Table 2. Concordance and hospital mortality in patients with hypoxemic respiratory failure

$\text{PaO}_2/\text{FiO}_2$	ARDS Diagnosis			Totals	Hospital Mortality
	Consensus-ARDS	Disagreement	No ARDS		
<100	81	59	172	312	31%
100–200	27	36	243	306	19%
>200	2	5	113	120	17%
Totals	110	100	528	738	
Hospital mortality	37%	35%	19%		23%

Definition of abbreviations: ARDS = acute respiratory distress syndrome; FiO_2 = fraction of inspired oxygen; PaO_2 = arterial oxygen tension/pressure.

Patients were categorized by whether physicians agreed with the diagnosis of ARDS and the severity of their hypoxemia. Overall agreement between these two categorizations was 31%.

Figure 1. Estimated proportions of patients categorized as consensus-ARDS, disagreement, or no ARDS when reviewed by multiple physicians across ranges of individual physiologic variables. Physiologic variables analyzed were the worst values recorded during the 24 hours after ARDS onset or matched 24-hour periods during mechanical ventilation for patients without ARDS. Missing values were imputed using multiple imputation and the estimates were combined. (A) $\text{PaO}_2/\text{FiO}_2$. (B) Static lung compliance. (C) Plateau pressure. (D) Dead space fraction. ARDS = acute respiratory distress syndrome; FiO_2 = fraction of inspired oxygen; PaO_2 = arterial oxygen tension/pressure.

shock, and patients with shock and $\text{PaO}_2/\text{FiO}_2 < 110$ but without pneumonia.

Discussion

In the current study, we found a high prevalence of patients with disagreement about the diagnosis of ARDS among those with hypoxemic respiratory failure. Patients with diagnostic disagreement appeared to be more borderline cases, with intermediate rates of common ARDS risk factors and measurements of lung injury severity. Yet, there were no clear cutoff values among single physiologic parameters (e.g., $\text{PaO}_2/\text{FiO}_2$) that could distinguish patients in any one group. We also found significant variation in the rate of ARDS diagnosis across clinicians: in half of the patients who received a diagnosis of ARDS from some clinicians, there was disagreement. Overall,

these results suggest that differences in how clinicians dichotomize patients along a continuous spectrum of ARDS manifestations are an important driver of disagreement about the diagnosis.

This analysis may have practical implications for how the diagnosis of ARDS can be improved in clinical practice. Given concerns that the low reliability of chest X-ray interpretation is a major driver of the low reliability of ARDS diagnosis (4, 16), and that standardized training in chest imaging interpretation does not improve the reliability of chest X-ray interpretation (17), the current analysis identified other important clinical characteristics that can help guide ARDS diagnosis. Specific combinations of clinical features can either strongly support the diagnosis of ARDS or potentially prompt physicians to perform further evaluation in patients in whom the diagnosis is likely to be unclear. The specific

combination of pneumonia, shock, and severe hypoxemia strongly indicated ARDS. The diagnosis was less clear in patients with pneumonia and $\text{PaO}_2/\text{FiO}_2 < 120$ but no shock, and patients with shock and $\text{PaO}_2/\text{FiO}_2 < 110$ but without pneumonia. Patients without pneumonia or shock and with a $\text{PaO}_2/\text{FiO}_2 > 180$ uniformly did not have ARDS, although these criteria should be validated in additional populations.

The finding in this study that patients with pneumonia and shock were most likely to receive a diagnosis of ARDS by a consensus of physicians aligns with the current understanding of the pathophysiology of this syndrome. In contrast to previous research that attempted to place patients in mutually exclusive categories of pulmonary and nonpulmonary ARDS (18, 19), we found that patients with risk factors from both groups had the highest rate of consensus-ARDS.

Figure 2. Relationship between the severity of hypoxemia and the likelihood of an ARDS diagnosis, and how this relationship is modified by the presence or absence of pneumonia and shock. The figure was generated after fitting multilevel logistic regression of ARDS reviews nested within patient, where ARDS diagnosis was the outcome, and model covariates included the $\text{PaO}_2/\text{FiO}_2$, a squared $\text{PaO}_2/\text{FiO}_2$ term, interaction terms with pneumonia and shock, and clinical reviewer adjustment. ARDS = acute respiratory distress syndrome; FiO_2 = fraction of inspired oxygen; PaO_2 = arterial oxygen tension/pressure.

Inflammatory fluid filling the alveoli during ARDS may arise after injury to both the lung epithelium and endothelium (20). Injury to one without the other fails to cause clear manifestations of ARDS in a classic sheep model of this syndrome (21). Patients presenting with pneumonia and shock may

have both direct lung epithelial injury and indirect endothelial injury from the systemic inflammatory response, leading to the more unambiguous manifestations of ARDS.

The current analysis suggests that rather than being a deterministic phenomenon, patients with features of ARDS have a

spectrum of clinical manifestations and severity. Differences in where physicians choose to dichotomize patients along this spectrum seemed to be a primary driver of disagreement. Considering ARDS as a spectrum of clinical manifestations—in the same way that acute kidney injury is a spectrum—may be useful when examining the strength of associations with ARDS risk factors or the heterogeneity of treatment effects on outcomes in patients with ARDS. For example, in a study by Shah and colleagues, excluding patients in whom it was difficult to establish a diagnosis because they lacked clear manifestations of ARDS led to findings of stronger associations between molecular markers and ARDS development (22). Rather than excluding such patients, an alternative approach used in recent study aiming to develop an algorithm to diagnosis ARDS weighted patients used to train the algorithm based on the clinical reviewer's confidence in the diagnosis (23). For example, if patients exhibited clear manifestations of ARDS, their diagnoses were assigned higher confidence, and these patients were weighted more heavily in the algorithm.

Although the current Berlin definition does conceptualize ARDS as having a

Table 3. Simple decision rules to approximate the collective judgment of multiple clinicians reviewing patients for ARDS

Decision Rules	Probability of Each Group		
	Consensus-ARDS	Disagreement	No ARDS
Pneumonia, shock, $\text{PaO}_2/\text{FiO}_2 < 120$	63%	16%	20%
Pneumonia, no shock, $\text{PaO}_2/\text{FiO}_2 < 120$	31%	30%	38%
Pneumonia, no shock, $\text{PaO}_2/\text{FiO}_2 = 120\text{--}180$	18%	10%	71%
Pneumonia, no shock, $\text{PaO}_2/\text{FiO}_2 > 180$	4%	26%	70%
No pneumonia, shock, $\text{PaO}_2/\text{FiO}_2 < 110$	20%	30%	50%
No pneumonia, shock, $\text{PaO}_2/\text{FiO}_2 > 110$	7%	7%	86%
No pneumonia, no shock, $\text{PaO}_2/\text{FiO}_2 < 180$	4%	10%	87%
No pneumonia, no shock, $\text{PaO}_2/\text{FiO}_2 > 180$	0%	0%	100%

Definition of abbreviation: ARDS = acute respiratory distress syndrome; FiO_2 = fraction of inspired oxygen; PaO_2 = arterial oxygen tension/pressure.

A decision tree classifier was used to derive decision rules for classifying patients as consensus-ARDS, disagreement about the diagnosis, or no ARDS.

spectrum of severity defined by the $\text{PaO}_2/\text{FiO}_2$, we did not find hypoxemia severity to be well calibrated with the likelihood of an ARDS diagnosis or with patient outcomes. Even among patients with a $\text{PaO}_2/\text{FiO}_2 < 100$, the majority did not have ARDS. This finding argues against using low $\text{PaO}_2/\text{FiO}_2$ alone as a surrogate for ARDS. Identifying clinical markers in patients with hypoxemic respiratory failure that are better aligned with both the spectrum of ARDS manifestations and clinical outcomes would improve stratification of patients for enrichment in clinical studies (24) and enable better tailoring of treatments in clinical practice.

Our study has some limitations. Because one of our interests was in understanding differences in the degree of lung injury across patient groups, we focused on patients receiving mechanical ventilation, enabling the calculation of lung compliance or the

dead space fraction. With the increasing use of high-flow nasal cannula, more patients with manifestations of ARDS may not receive mechanical ventilation. We suspect our findings would also apply to this population, but this requires confirmation. Additionally, the analysis was performed using data recorded in the electronic health record as part of routine care. Such data may have a higher level of error rates than prospectively collected research data, which may have blunted any differences between patient groups (6). Diagnostic agreement was evaluated among clinicians who reviewed patients retrospectively, so whether similar results would be obtained from a prospective evaluation is not known. Finally, although our study included 13 critical-care physicians and a large cohort of patients drawn from medical, surgical, cardiac, and trauma patient populations, further validation of the

findings in additional patient cohorts is warranted.

Conclusions

Patients with disagreement about the diagnosis of ARDS had more intermediate rates of ARDS risk factors and lung injury severity. Overall, these findings suggest that ARDS has a spectrum of clinical manifestations, and the difficulty of dichotomizing patients along this spectrum may explain disagreements about the diagnosis of ARDS. Considering both the presence of key ARDS risk factors and hypoxemia severity can help guide clinicians in identifying patients with a diagnosis of ARDS agreed upon by a consensus of physicians. ■

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

References

- Ashbaugh DG, Bigelow DB, Petty TL, Levine BE. Acute respiratory distress in adults. *Lancet* 1967;2:319–323.
- Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al.; ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin definition. *JAMA* 2012;307:2526–2533.
- Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149:818–824.
- 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.
- Thille AW, Esteban A, Fernández-Segoviano P, Rodríguez JM, Aramburu JA, Peñuelas O, et al. Comparison of the Berlin definition for acute respiratory distress syndrome with autopsy. *Am J Respir Crit Care Med* 2013;187:761–767.
- 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.
- Fröhlich S, Murphy N, Boylan JF. ARDS: progress unlikely with non-biological definition. *Br J Anaesth* 2013;111:696–699.
- Sjoding MW, Hyzy RC. Recognition and appropriate treatment of the acute respiratory distress syndrome remains unacceptably low. *Crit Care Med* 2016;44:1611–1612.
- Rubenfeld GD. Confronting the frustrations of negative clinical trials in acute respiratory distress syndrome. *Ann Am Thorac Soc* 2015;12:S58–S63.
- Gajic O, Dabbagh O, Park PK, Adesanya A, Chang SY, Hou P, et al.; U.S. Critical Illness and Injury Trials Group: Lung Injury Prevention Study Investigators (USCIITG-LIPS). Early identification of patients at risk of acute lung injury: evaluation of lung injury prediction score in a multicenter cohort study. *Am J Respir Crit Care Med* 2011;183:462–470.
- Brown SM, Duggal A, Hou PC, Tidswell M, Khan A, Exline M, et al.; National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute (NHLBI) Prevention and Early Treatment of Acute Lung Injury (PETAL) Network. Nonlinear imputation of $\text{PaO}_2/\text{FiO}_2$ from $\text{SpO}_2/\text{FiO}_2$ among mechanically ventilated patients in the ICU: a prospective, observational study. *Crit Care Med* 2017;45:1317–1324.
- Siddiki H, Kojicic M, Li G, Yilmaz M, Thompson TB, Hubmayr RD, et al. Bedside quantification of dead-space fraction using routine clinical data in patients with acute lung injury: secondary analysis of two prospective trials. *Crit Care* 2010;14:R141.
- Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
- Cao J, Zhang S. Multiple comparison procedures. *JAMA* 2014;312:543–544.
- Walkey AJ, Summer R, Ho V, Alkana P. Acute respiratory distress syndrome: epidemiology and management approaches. *Clin Epidemiol* 2012;4:159–169.
- Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. *Chest* 1999;116:1347–1353.
- Peng JM, Qian CY, Yu XY, Zhao MY, Li SS, Ma XC, et al.; China Critical Care Clinical Trial Group (CCCCTG). Does training improve diagnostic accuracy and inter-rater agreement in applying the Berlin radiographic definition of acute respiratory distress syndrome? A multicenter prospective study. *Crit Care* 2017;21:12.
- Gattinoni L, Pelosi P, Suter PM, Pedoto A, Vercesi P, Lissoni A. Acute respiratory distress syndrome caused by pulmonary and extrapulmonary disease. Different syndromes? *Am J Respir Crit Care Med* 1998;158:3–11.
- Thille AW, Richard JC, Maggiore SM, Ranieri VM, Brochard L. Alveolar recruitment in pulmonary and extrapulmonary acute respiratory distress syndrome: comparison using pressure-volume curve or static compliance. *Anesthesiology* 2007;106:212–217.
- Matthay MA, Zemans RL. The acute respiratory distress syndrome: pathogenesis and treatment. *Annu Rev Pathol* 2011;6:147–163.
- Wiener-Kronish JP, Albertine KH, Matthay MA. Differential responses of the endothelial and epithelial barriers of the lung in sheep to *Escherichia coli* endotoxin. *J Clin Invest* 1991;88:864–875.
- Shah CV, Lanken PN, Localio AR, Gallop R, Bellamy S, Ma SF, et al. An alternative method of acute lung injury classification for use in observational studies. *Chest* 2010;138:1054–1061.
- Reamaroon 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* [online ahead of print] 28 Feb 2018; DOI: 10.1109/JBHI.2018.2810820.
- Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX. Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. *Am J Respir Crit Care Med* 2016;194:147–155.