Rethinking Acute Respiratory Distress Syndrome after COVID-19: If a "Better" Definition Is the Answer, What Is the Question?

V. Marco Ranieri[1]*, Gordon Rubenfeld[2]*, and Arthur S. Slutsky[3,4]*

[1]Alma Mater Studiorum – Università di Bologna, Dipartimento di Scienze Mediche e Chirurgiche, IRCCS Policlinico di Sant'Orsola, Anesthesia and Intensive Care Medicine, Bologna, Italy; [2]Department of Critical Care, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada; [3]Keenan Research Centre for Biomedical Science, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada; and [4]Department of Medicine, University of Toronto, Toronto, Ontario, Canada

Abstract

The definition of acute respiratory distress syndrome (ARDS) has a somewhat controversial history, with some even questioning the need for the term "ARDS." This controversy has been amplified by the coronavirus disease (COVID-19) pandemic given the marked increase in the incidence of ARDS, the relatively new treatment modalities that do not fit neatly with the Berlin definition, and the difficulty of making the diagnosis in resource-limited settings. We propose that attempts to revise the definition of ARDS should apply the framework originally developed by psychologists and social scientists and used by other medical disciplines to generate and assess definitions of clinical syndromes that do not have gold standards. This framework is structured around measures of reliability, feasibility, and validity. Future revisions of the definition of ARDS should contain the purpose, the methodology, and the framework for empirically testing any proposed definition. Attempts to revise critical illness syndromes' definitions usually hope to make them "better"; our recommendation is that future attempts use the same criteria used by other fields in defining what "better" means.

Keywords: definition; acute respiratory distress syndrome; framework; validity

> "When I use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean—neither more nor less." "The question is," said Alice, "whether you can make words mean so many different things."
>
> —Lewis Carroll, Through the Looking Glass and What Alice Found There (1)

Ashbaugh and coworkers introduced the term "acute respiratory distress syndrome" (ARDS) in a seminal article published in 1967 describing 12 patients with acute hypoxemic respiratory failure (AHRF) and a constellation of signs, symptoms, and laboratory abnormalities that distinguished them from 272 other ventilated patients (2). Pathologic examination in seven patients revealed atelectasis, vascular congestion and hemorrhage, severe pulmonary edema, and hyaline membranes. Shortly afterward, Petty and Ashbaugh called this constellation "ARDS" (3). Over the years, there have been many refinements of the definition of ARDS.

The purpose of this article is to propose a change in how the critical care community has historically defined syndromes such as ARDS, from an approach based on a "consensus of experts" to a "scientific system of categorization," using approaches adopted from other fields to develop syndromic definitions for constructs that lack gold standards (4). Although these approaches have already been partially introduced in the most recent definition of ARDS (5), the pressing request for its revision (6, 7) requires the adoption of a rigorous and transparent methodology, widely shared by the entire clinical and scientific critical care medicine community.

Historical Perspective

Within a decade of its description, a debate ensued about the term "ARDS." On one side, lumpers argued that the term "ARDS" was useful and represented a distinct clinical entity (8). On the other side, splitters argued that the diagnosis served no purpose and proposed multiple individual diagnoses related to underlying lung injury mechanisms (9). By 1988, the splitters had proposed a definition incorporating the risk factors for ARDS and a "lung injury score" that assigned points to chest radiography

(Received in original form June 3, 2022; accepted in final form September 23, 2022)

*These authors contributed equally to this work.

Correspondence and requests for reprints should be addressed to Arthur S. Slutsky, M.D., St. Michael's Hospital, 30 Bond Street, Toronto, ON M5B 1W8, Canada. E-mail: arthur.slutsky@unityhealth.to.

Am J Respir Crit Care Med Vol 207, Iss 3, pp 255–260, Feb 1, 2023

Originally Published in Press as DOI: 10.1164/rccm.202206-1048CP on September 23, 2022
Internet address: www.atsjournals.org

(CXR), the ratio of arterial partial pressure of oxygen to $F_{I_{O_2}}$ (P:F), positive end-expiratory pressure (PEEP), and compliance (10).

In 1994, the American–European Consensus Conference standardized the definition (11). In that paper, relatively little attention was devoted to the development or justification of the definition, instead reviewing the science on acute lung injury. In 2004, to deal with poor reliability of CXR (12, 13) and poor validity of the definition compared with autopsy findings (14), Ferguson and coworkers used a Delphi process to formally elicit agreement on the definition (15).

In 2012, a consensus conference revised the definition of ARDS (5). The "Berlin definition" offered relatively small changes to the American–European Consensus Conference definition. Two definitions of severe ARDS were presented: one similar to the definition proposed by Ferguson and coworkers (15) and the second based solely on P:F. Using data from large databases, both definitions generated groups with similar mortality and other markers of lung injury, but the latter included more patients. Using prespecified rules, the simpler definition of severe ARDS was retained.

Despite all this work, the definition of ARDS has a number of limitations in both resource-rich and resource-poor settings (16), many of which were amplified by coronavirus disease (COVID-19). These include the fact that patients on high-flow nasal oxygen cannot (by definition) meet ARDS criteria (17), decreasing use of CXR as lung ultrasound use flourishes (18), decreasing use of arterial blood gas analysis with increasing use of arterial O_2 saturation (19), and the role of standardized ventilator settings (20).

Matthay and coworkers proposed an expanded definition of ARDS that includes high-flow nasal oxygen, lung ultrasound, and the ratio of oxygen saturation as measured by pulse oximetry to $F_{I_{O_2}}$ (6); an international consensus conference group will soon make formal recommendations for expanding the definition of ARDS (7). Conversely, others have proposed abandoning the term "ARDS" in favor of risk factor–associated diseases (9, 21). For example, in the original lumper/splitter debate, Murray noted that although "the pulmonary manifestations are similar in patients with fat embolism, smoke inhalation, and acute pancreatitis, the basic mechanisms of lung injury in these three members of the ARDS family appear to be different" and may detract "from important and distinctive differences in pathogenesis, therapy and prognosis" (9).

Why Are Formal Definitions of ARDS Needed, and How Should We Create Them?

A formal definition of ARDS is helpful because it allows the inclusion of similar patients in research and provides therapeutic and/or prognostic information. A major challenge of achieving these goals is the lack of gold standards (22); even diffuse alveolar damage (DAD) is not consistently accepted as a gold standard (14). However, the challenge of using rigorous methods to validate diagnostic criteria for syndromes that lack gold standards is not insurmountable (4), as demonstrated by psychologists and social scientists (4).

Critical for addressing this challenge is understanding the similarity between a clinical syndrome and what social scientists call a "construct." According to Binning, a construct is "derived from the general scientific process: observing natural phenomena, inferring common features, and constructing a label for the observed commonality or the underlying cause of the commonality" (23). The key, as Binning argued, is that constructs derive their scientific value from the shared meaning they represent for different people (23).

A clearly defined construct is one that different people think similarly about and thus is helpful in facilitating common understanding (23). According to Streiner and colleagues, "many of what physicians call 'syndromes' would be called 'hypothetical constructs' by psychologists" (4). ARDS fits into this schema, as it is closer to a construct than an actual disease, because the causal mechanisms are heterogeneous and poorly understood and there is no gold standard diagnostic test (4). Clinicians from all fields diagnose, treat, and study constructs daily: heart failure, frailty, irritable bowel syndrome, dementia, and chronic fatigue are some examples.

In 1987 Feinstein, a physician/epidemiologist, published *Clinimetrics*, a book that addressed how to incorporate the rigorous methods used to measure "constructs" into the measurement/definition of clinical phenomena (24). An enormous body of literature exists for developing tools to measure these complex constructs such as intelligence, racism, and quality of life (25, 26). The key insight is that diagnostic criteria for a syndrome should be developed and evaluated as instruments to measure a construct; this may not be intuitive, but it is not novel. Psychiatrists adopted this methodology to develop rigorous definitions that evolved "from the great professor principle, to the consensus of experts, to a scientific system of categorization" (27).

We suggest that to redefine ARDS, the critical care community should use the existing framework to measure constructs, which is structured around measures of reliability, feasibility, and validity (4).

Reliability

We intentionally place reliability first. Because a primary goal of definitions is to facilitate research, syndromic diagnostic criteria should identify the same patients at different centers. Reliability quantifies the agreement between observers on the diagnosis, or a measurement of one of the diagnostic criteria (28) using measurements of intraobserver and interobserver reliability.

Empiric research on the reliability of most ARDS diagnostic criteria is lacking. For example, P:F is dependent on ventilator settings. Therefore, similar patients will not necessarily be diagnosed the same way unless PEEP and volume history are standardized. Even more disconcerting, the same patient with ARDS may not have ARDS after an increase in PEEP if the increase leads to P:F $>$ 300 mm Hg. As another example, CXR criteria are unreliable and attempts to standardize them have not been successful (29). Lung ultrasound may be more reliable; however, measuring reliability across multiple sonographers may be more difficult than similar studies with CXR given the interactive nature of ultrasound. Given the poor reliability of imaging criteria to diagnose ARDS, we have no reason to believe that similar patients, particularly those at the less severe end of the spectrum, will be identified the same way by CXR and lung ultrasound or similarly at different centers.

Future attempts to improve the reliability of the definition of ARDS should focus on empiric studies of proposed criteria. Standard approaches to improve reliability include protocols, training materials, and incorporating multiple measurements. Of note, poor agreement on components of a

definition may not be fatal for the overall definition. This can occur if the domain (e.g., chest imaging or exclusion of heart failure) has no effect on validity, particularly response to treatments. In situations in which reliability is poor, definition panels should consider dropping that part of the definition.

Feasibility

Feasibility is the "real-world practicality of obtaining diagnostic criteria in clinical and research settings" (22). Ideally, feasibility questions are answered with pilot studies addressing availability, complexity, time, cost, consent, and the consequences of false positives and false negatives (4) in different settings.

Let's address a few examples. First, the ratio of oxygen saturation as measured by pulse oximetry to F_{IO_2} is clearly a more feasible measure of hypoxemia than P:F, as pulse oximetry is safer, less painful, and more readily available. A similar argument can be made in favor of lung ultrasound rather than CXR. These examples are particularly important in resource-poor regions (16). Second, standardized ventilator settings (20) or special imaging to diagnose ARDS (30) might require consent if not part of routine care, which can pose a significant barrier to research. Third, biomarkers to diagnose ARDS also pose feasibility questions (31). If a new definition requires biomarker measurements, it will be important that the biomarker(s) be routinely available and inexpensive to measure.

Validity

The most challenging aspect of evaluating a proposed definition of ARDS is assessing its validity. In 1927 Kelley stated that "a test is valid if it measures what it says it measures" (32). The problem remains, how do we know whether the definition identifies patients who "really" have ARDS if there is no gold standard? Here, too, there is a large body of empiric, theoretical, and statistical work from the social sciences to rely on. Unfortunately, the terminology for validity testing is complex, evolving, and often inconsistently used (4, 26).

Feinstein noted that "validity is probably the most difficult word encountered in the metrics of clinical and psychosocial indexes. When the jargon adds an array of prefixes to differentiate various types of validity such as face validity, content validity, criterion validity, and construct validity, substantial effort may be needed simply to distinguish and remember all the different connotations" (24). To further complicate matters terms such as "sensibility," "accuracy," "suitability," and "consistency" are used as synonyms for some types of validation.

Recent recommendations tend to avoid classification schemes for validity; rather they focus on forming a validation hypothesis and testing it (4, 26). For our purposes, we will separate validity into face validity, predictive validity, and other tests of validation. Readers with an interest in the various validity schemes are referred to texts of psychometric testing (4, 26).

Face validity. Face validity describes whether the proposed definition captures all the features of a syndrome. This can be assessed empirically using surveys or with formal expert consensus techniques (e.g., Delphi, nominal group). More often, it is based on the informal best shared opinion of the group writing the paper. Although this does not seem rigorous, it is essential, as only clinical experts can reasonably judge whether the definition fully captures the construct (4, 26).

It is important to identify disagreements over face validity, as these are difficult to resolve empirically (4, 26). For example, some have argued that patients with rapid resolution of hypoxemia do not have ARDS. In essence this addresses face validity, as it follows directly from the original description by Ashbaugh, who noted that patients with ARDS "did not respond to usual and ordinary methods of respiratory therapy" (2). As one would expect from the fact that people who get better do better, patients with rapid resolution of hypoxemia have markedly lower mortality than those who do not (20). Therefore, whether these patients have ARDS cannot be based on their lower mortality, as the lower mortality is expected on the basis of incorporating "rapid resolution" in the definition. The fundamental questions here are whether these patients fit our construct of ARDS and whether the diagnostic delay imposed by incorporating a nonresponder component in the definition ARDS would impede clinical research (4).

In scenarios in which the definition panel finds itself unable to specify any validation hypotheses, face validity takes on unique importance. Here the panel is expressing its own gestalt of "what ARDS is." Under these circumstances, the makeup of the panel and its deliberative methods are particularly crucial. As such, the panel should ideally represent diverse opinions across a broad spectrum of clinicians, researchers, and patients.

Predictive validity. Predictive validity has received a great deal of attention since its use in the Berlin definition addressing the severity subcategories for ARDS (5). It is crucial to distinguish predictive validation from prediction. The predictive validity analysis in the Berlin consensus conference picked between two proposed definitions of severe ARDS (5). As a prediction tool for mortality, the Berlin definition with three P:F strata performs poorly. A definition for ARDS incorporating age, risk factors, and other organ failures would predict mortality significantly better than the Berlin definition. So predictive accuracy is not the goal of predictive validity.

Furthermore, just because P:F was selected using a predictive validity framework does not necessarily mean it is a "better" definition; it simply means it was picked using prespecified criteria to define "better." The alternative Berlin definition for severe ARDS, which included compliance and a marker for dead space, might have been a better choice if the goal were face validity or if the goal were to encourage physicians to consider physiology more in clinical management.

Some studies incorporate predictive validity hypotheses without acknowledging them, as for example the observation that patients with AHRF, regardless of whether they meet ARDS criteria, have similar mortality (33, 34). In these studies, the only criterion separating AHRF from ARDS is a compatible CXR. In the framework proposed here, we would say that these data demonstrate that CXR lacks predictive validity for mortality. Of course, this observation alone does not mean that CXR should be removed from the definition. It may be an important part of face validity, as ARDS is considered as a diffuse inflammatory process, and all of the clinical data used to arrive at this conclusion were obtained from patients with ARDS who were enrolled using a definition that incorporated the CXR criterion.

Other tests of validation. Validation studies should begin with a simple hypothesis: patients with the syndrome should have X and patients without the syndrome should not have X. For example,

people with depression should "miss more work," "have more suicide," and "have more divorces" compared with patients without this diagnosis (35). To address other ARDS validation questions, one should begin with a simple hypothesis, for example, patients with ARDS should have DAD and patients without ARDS should not have DAD (14, 36). If such a hypothesis cannot be stated, then empiric validation of the definition cannot be done, and attempts to refine the definition would be left with face validity, reliability, and feasibility.

We believe that one of the core purposes of the expert panels convened for the purpose of redefining syndromes is to reach consensus on the validation hypotheses. Although the terminology of validation may be new to some critical care researchers, the research questions are not. Studies that compare patients with ARDS and those without ARDS regarding different forms of X such as mortality, length of stay, biomarkers, lung water, and compliance all test validation hypotheses (4). However, the literature is limited because different forms of the definition of ARDS are not usually tested, and the hypotheses are more often framed around exploring mechanisms of ARDS rather than validation of the definition. In fact, it may be that there is no adequate validation variable, even pathology, that we can agree that all patients with ARDS share. As such, perhaps all that consensus-based definitions can do is improve reliability and face validity.

Of course, there are other details of validation studies that must be resolved. For example, it may add to the validation argument that there is a dose–response relationship between ARDS severity and X. Appropriate choice of a control group (patients without ARDS) is critical. The control group can include normal subjects, which would test the most extreme comparison, and such studies are probably valuable to pilot novel imaging or biomarker measures. Other studies may test patients at the margins of the definitions, for example, comparing patients who meet the definition of ARDS according to CXR findings, those who do not, and those in between (37, 38). The most important validation hypothesis is whether patients who meet the definition of ARDS respond to treatments designed to treat ARDS. Unfortunately, this is not a pragmatic tool for refining definitions, as most randomized trials fail; it is unclear if their failure is due to poor definitions or therapies that are simply ineffective.

What Are the Answers?

Although the Berlin definition included some features of the framework proposed here, there are several features it did not incorporate. A few examples are the fact that it did not explicitly state the process for selecting experts; it did not include a diverse panel; it did not state the existence of and rationale for trade-offs among the components of feasibility, reliability, and validity; and it did not use a methodology to achieve consensus, or indicate specific validation hypotheses for future research. We believe that the following steps should be taken into consideration in any revision process of the definition of ARDS.

First, every attempt to redefine ARDS should begin with two questions: "What is wrong with the current definition?" and "How will we know that the new definition is better?" We argue that the process of improving the definition must begin with understanding what is meant by "better" and must include the methods to quantify "better." We therefore need to explicitly address the language required to describe ARDS as a construct and to deal with the heterogeneous causal mechanisms and the consequences of the lack of gold-standard diagnostic tools (4, 23). Sharing a language on what a syndrome definition is will allow us to communicate on how to make it better.

Second, a formal methodology for developing the proposed definition with explicit elucidation of rationale, framework, and methods of evaluating it is required (Table 1). Multidisciplinary panels coordinated by professional societies are needed to prevent competing definitions. For example, consensus is required to address questions of face validity (i.e., whether the proposed changes in definition still capture the clinical construct of ARDS). The validation process should begin with the hypothesis that patients with ARDS should have X and patients without ARDS should not have X, and the panel should define the variables that would be persuasive in a validation study. If such a hypothesis cannot be stated, then empiric validation simply cannot be done, and attempts to refine the definition would best be left with reliability, feasibility, and consensus face validity (37, 38).

The panel should address explicit trade-offs that may arise when efforts to improve one domain of the definition (say, reliability) directly conflict with the requirements of another domain (say, feasibility). It is important to note that *1*) validity is a continuous process of evaluation, reevaluation, refinement, peer review, and development, and *2*) although the statistical methods for developing, refining, and validating multidimensional scores and instruments can be quite complex, the hypotheses around syndrome validation generally are not. Validation of a clinical syndrome definition is fundamentally a clinical question and not a statistical one.

Third, consensus methods and selection criteria of panel members should be explicitly declared and widely available. Panels should *1*) be diverse along multiple domains, *2*) define as "experts" clinical scientists and clinician practitioners, *3*) involve all healthcare professionals involved in the care of patients with ARDS, and *4*) include patients and family representatives.

Fourth, any new definition of ARDS should recognize that there are different kinds of ARDS on the basis of clinical and/or physiological characteristics (39), such as responsiveness to PEEP (40), and biological characteristics (31). Hierarchical approaches could be used starting with the AHRF phenotype, on the basis of a simple cluster of feasible and reliable observable characteristics, and then carve out different endotypes from this phenotype (41). Sophisticated clustering algorithms using biochemical and physiological parameters could be used to identify AHRF endotypes (42). These data-mining techniques rely on the hypothesis that clusters of patient markers that are statistically close to one another will behave similarly in ways that may not be obvious clinically. Which of these mathematically derived clusters would be called "ARDS" would be up to the investigators and could not be empirically derived or, perhaps, "ARDS" would be abandoned in favor of a mathematically derived endotype (37, 38).

In conclusion, the challenge of developing syndromic definitions for medical conditions that lack gold-standard diagnostic tests is common in medicine, but the methodology to develop and then test these definitions has been infrequently used

Table 1. Outline for Acute Respiratory Distress Syndrome and Other Critical Illness Syndrome Definition Papers

Introduction	• The clinical phenomenon the definition wants to capture (description of the clinical construct) • A review of prior definitions • What specific aspects of the previous definition need improvement and why • How the changes in the proposed definition will be assessed to determine if the needed improvement has been achieved • Justification of any changes to the definition that are proposed without evaluation
Methodology	• Consensus: Methods for reaching consensus and expressing minority opinion or an explicit consensus method should be stated. The selection and diversity of panel members are important, as in other consensus-based projects (15). The method for selecting and ensuring a diverse panel should be stated. • Feasibility: practicality of obtaining diagnostic criteria in clinical and research settings • Reliability: ability to provide the same diagnosis in the same patient by different clinicians in varying clinical scenarios • Validity: ability of the definition to capture what clinicians truly seek to identify ○ Face validity: The characteristics are obviously part of the syndrome and together distinguish patients with the syndrome from those without. ○ Predictive validity: The definition predicts outcomes that patients with the syndrome should have compared with those without. ○ Other tests of validation: The validation process of the revised definition begins by stating a validation hypothesis.
Potential limitations	• Need not present novel research as part of the definition if sufficient published evidence supports change • Can propose needed research, but changes in definition should not be based on anticipated research • Feasibility often relies on unpublished clinical perception, but rigorous studies, particularly in underresourced settings, are likely needed. • Issues that rely specifically on consensus, for example, face validity or trade-offs between aspects of the definition, should be stated; as well as minority opinions. • The statistical limitations required to prove that two populations are the same should be acknowledged in reviewing these studies (44).
Future research	• Where the panel identifies limitations of its definition, specific recommendations for reliability, feasibility, and validation studies should be made. These should be explicit enough for investigators to carry out.

in critical care (43). Future attempts to revise critical care syndromic definitions should contain limitations of the current definition, purpose of the revision, methodology for its derivation, and the framework for empirically testing reliability, feasibility, and validity. We believe that it is time to abandon the concept of a single definition for all purposes. It may be useful to develop different definitions for use in randomized trials in which consent is obtainable or for use in less resourced settings. However, it is essential that we understand any differences in the patients identified with different versions of the definition. ■

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

References

1. Carroll L. Through the looking glass and what Alice found there. London: Macmillan; 1871.
2. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE. Acute respiratory distress in adults. *Lancet* 1967;2:319–323.
3. Petty TL, Ashbaugh DG. The adult respiratory distress syndrome: clinical features, factors influencing prognosis and principles of management. *Chest* 1971;60:233–239.
4. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford, UK: Oxford University Press; 2015.
5. 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.
6. Matthay MA, Thompson BT, Ware LB. The Berlin definition of acute respiratory distress syndrome: should patients receiving high-flow nasal oxygen be included? *Lancet Respir Med* 2021;9:933–936.
7. Ware LB. Go with the flow: expanding the definition of acute respiratory distress syndrome to include high-flow nasal oxygen. *Am J Respir Crit Care Med* 2022;205:380–382.
8. Petty TL. Editorial: the adult respiratory distress syndrome (confessions of a "lumper"). *Am Rev Respir Dis* 1975;111:713–715.
9. Murray JF. Editorial: the adult respiratory distress syndrome (may it rest in peace). *Am Rev Respir Dis* 1975;111:716–718.
10. Murray JF, Matthay MA, Luce JM, Flick MR. An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis* 1988;138: 720–723.
11. 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.
12. 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.
13. Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. *Chest* 1999;116:1347–1353.
14. Esteban A, Fernández-Segoviano P, Frutos-Vivar F, Aramburu JA, Nájera L, Ferguson ND, *et al.* Comparison of clinical criteria for the acute respiratory distress syndrome with autopsy findings. *Ann Intern Med* 2004;141:440–445.
15. Ferguson ND, Davis AM, Slutsky AS, Stewart TE. Development of a clinical definition for acute respiratory distress syndrome using the Delphi technique. *J Crit Care* 2005;20:147–154.
16. Riviello ED, Kiviri W, Twagirumugabe T, Mueller A, Banner-Goodspeed VM, Officer L, *et al.* Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition. *Am J Respir Crit Care Med* 2016;193:52–59.

17. Ranieri VM, Tonetti T, Navalesi P, Nava S, Antonelli M, Pesenti A, *et al.* High-flow nasal oxygen for severe hypoxemia: oxygenation response and outcome in patients with COVID-19. *Am J Respir Crit Care Med* 2022;205:431–439.
18. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. *Anesthesiology* 2004;100:9–15.
19. Rice TW, Wheeler AP, Bernard GR, Hayden DL, Schoenfeld DA, Ware LB; National Institutes of Health, National Heart, Lung, and Blood Institute ARDS Network. Comparison of the $Sp_{O_2}/F_{I_{O_2}}$ ratio and the $Pa_{O_2}/F_{I_{O_2}}$ ratio in patients with acute lung injury or ARDS. *Chest* 2007; 132:410–417.
20. Villar J, Perez-Mendez L, Blanco J, Anon JM, Blanch L, Belda J, *et al.*; Spanish Initiative for Epidemiology, Stratification, and Therapies for ARDS (SIESTA) Network. A universal definition of ARDS: the $Pa_{O_2}/F_{I_{O_2}}$ ratio under a standard ventilatory setting—a prospective, multicenter validation study. *Intensive Care Med* 2013;39: 583–592.
21. Gattinoni L, Marini JJ. Isn't it time to abandon ARDS? The COVID-19 lesson. *Crit Care* 2021;25:326.
22. Rubenfeld GD. Epidemiology of acute lung injury. *Crit Care Med* 2003;31: S276–S284.
23. Binning JF. Construct. Chicago: The Britannica Group; 2016 [updated 2016 Feb 22; accessed 2022 Jul 19]. Available from: https://www.britannica.com/science/construct.
24. Feinstein AR. Clinimetrics. New Haven, CT: Yale University Press; 1987.
25. Allen MJ, Yen WM. Introduction to measurement theory. Long Grove, IL: Waveland Press; 2001.
26. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
27. Kendler KS. Toward a scientific psychiatric nosology: strengths and limitations. *Arch Gen Psychiatry* 1990;47:969–973.
28. Green CE, Chen CE, Helms JE, Henze KT. Recent reliability reporting practices in *Psychological Assessment*: recognizing the people behind the data. *Psychol Assess* 2011;23:656–669.
29. Goddard SL, Rubenfeld GD, Manoharan V, Dev SP, Laffey J, Bellani G, *et al.* The Randomized Educational Acute Respiratory Distress Syndrome Diagnosis study: a trial to improve the radiographic diagnosis of acute respiratory distress syndrome. *Crit Care Med* 2018; 46:743–748.
30. Pesenti A, Musch G, Lichtenstein D, Mojoli F, Amato MBP, Cinnella G, *et al.* Imaging in acute respiratory distress syndrome. *Intensive Care Med* 2016;42:686–698.
31. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA; NHLBI ARDS Network. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014;2:611–620.
32. Kelley TL. Interpretation of educational measurements. New York: Macmillan; 1927.
33. Luhr OR, Karlsson M, Thorsteinsson A, Rylander C, Frostell CG. The impact of respiratory variables on mortality in non-ARDS and ARDS patients requiring mechanical ventilation. *Intensive Care Med* 2000;26: 508–517.
34. Villar J, Mora-Ordoñez JM, Soler JA, Mosteiro F, Vidal A, Ambrós A, *et al.* The PANDORA study: prevalence and outcome of acute hypoxemic respiratory failure in the pre-COVID-19 era. *Crit Care Explor* 2022;4:e0684.
35. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, *et al.* Cross-national epidemiology of *DSM*-IV major depressive episode. *BMC Med* 2011;9:90.
36. Thille AW, Esteban A, Fernández-Segoviano P, Rodriguez 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.
37. 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.
38. Shah FA, Meyer NJ, Angus DC, Awdish R, Azoulay É, Calfee CS, *et al.* A research agenda for precision medicine in sepsis and acute respiratory distress syndrome: an official American Thoracic Society research statement. *Am J Respir Crit Care Med* 2021;204:891–901.
39. Nuckton TJ, Alonso JA, Kallet RH, Daniel BM, Pittet JF, Eisner MD, *et al.* Pulmonary dead-space fraction as a risk factor for death in the acute respiratory distress syndrome. *N Engl J Med* 2002;346:1281–1286.
40. Gattinoni L, Caironi P, Cressoni M, Chiumello D, Ranieri VM, Quintel M, *et al.* Lung recruitment in patients with the acute respiratory distress syndrome. *N Engl J Med* 2006;354:1775–1786.
41. Agache I, Akdis CA. Precision medicine and phenotypes, endotypes, genotypes, regiotypes, and theratypes of allergic diseases. *J Clin Invest* 2019;129:1493–1503.
42. Beitler JR, Thompson BT, Baron RM, Bastarache JA, Denlinger LC, Esserman L, *et al.* Advancing precision medicine for acute respiratory distress syndrome. *Lancet Respir Med* 2022;10:107–120.
43. Angus DC, Seymour CW, Coopersmith CM, Deutschman CS, Klompas M, Levy MM, *et al.* A framework for the development and interpretation of different sepsis definitions and clinical criteria. *Crit Care Med* 2016;44:e113–e121.
44. Shankar-Hari M, Rubenfeld GD. Just because things are not different, does not mean they are the same: biomarker patterns in acute respiratory distress syndrome. *Crit Care Med* 2017;45:1955–1957.