ARDs: progress unlikely with non-biological definition

S. Frohlich1,2*, N. Murphy1,2 and J. F. Boylan1

1 Department of Anaesthesia and Intensive Care Medicine, St Vincent’s University Hospital, Dublin, Ireland
2 Lung Biology Group, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland
* Corresponding author: Department of Anaesthesia and Intensive Care Medicine, St Vincent’s University Hospital, Dublin 4, Ireland.
E-mail: frohlics@yahoo.co.uk

Accurate disease definitions are essential for clinical decision-making, trial enrolment, and mechanistic research. Since it was first recognized over four decades ago multiple attempts have been made to adequately define the acute respiratory distress syndrome (ARDS). However, up to half of the patients captured by definitions to date do not have the disease. This poor diagnostic accuracy may, in part, explain how over 200 randomized control trials have, with the exception of a low tidal volume ventilation strategy, failed to identify any mortality-reducing therapies. The recently published Berlin definition of ARDS was introduced to address the poor accuracy and shortcomings associated with previous models; however, a recent validation study found the diagnostic reliability of the Berlin definition to be no superior to its predecessor. In the absence of accurate, objectively validated criteria for diagnosing this condition, clinical trials will continue to include large numbers of patients without the disease.

ARDS is a non-pneumonic, non-cardiogenic condition characterized by increased vascular permeability, pulmonary oedema, and severe arterial hypoxaemia—a clinical triad found in many critically ill patients. It is precipitated by both direct and indirect causes, resulting in a variable clinical pattern of presentation and progression. ARDS is frequently associated with other organ failures, resulting in both pulmonary and extra-pulmonary factors influencing mortality. Unlike most medical conditions, this complex syndrome does not yet benefit from an accessible in vivo reference standard (e.g. diagnosis of pulmonary embolism), or the existence of clear biomarkers (e.g. acute coronary syndrome). As a result, separating true ‘lung injury’ from other conditions, such as heart failure or pneumonia, which display similar clinical signs, is challenging.

Although the reference standard has so far remained constant, lack of clarity about the clinical definition of ARDS has been problematic. Since 1967, when Ashbaugh and Petty first described the syndrome, there have been attempts to refine its definition. Initially, Murray proposed a definition comprising four variables—chest X-ray (CXR) findings, PaO₂/FIO₂ ratio, positive end expiratory pressure (PEEP), and respiratory compliance. The resultant lung injury score (LIS) predicts the need for prolonged mechanical ventilation and has good sensitivity and specificity (0.74 and 0.77, respectively) as a predictor of diffuse alveolar damage (DAD) at autopsy. Despite this, the LIS has not entered routine practice, and a more pragmatic approach that targets ease of use has prevailed.

In 1994, the American-European Consensus Conference (AECC) redefined ARDS using criteria based on hypoxaemia, CXR infiltrates, and absence of left atrial hypertension; abnormal lung mechanics were dropped as a criterion. The AECC authors also urged diagnostic caution ‘in order to minimize the chance of including non-ARDS-related illnesses’.3 Their criteria are accessible, were adopted in widespread clinical practice, and were used as entry criteria in major clinical trials, without prior validation against pathological data. However, specificity is low (Table 1), and when lung biopsy is undertaken in seriously ill patients fulfilling the criteria, a diagnosis other than ARDS that necessitates a change in treatment is found in 60% of cases.5 If a sensitivity of 0.80 and specificity of 0.50—as reported by Ferguson and colleagues—are typical, then in critically ill patients with an ARDS prevalence of 30%, the AECC criteria have a positive predictive value (PPV) and negative predictive value (NPV) of 0.41 and 0.85, respectively.6 Although the definition may be useful in a screening role, 50% or more of patients diagnosed with ARDS by AECC criteria have another condition; AECC-based diagnosis exposes patients to another rate fallacy, in which pretest probability and test likelihood ratios are unknown, the distinction between sensitivity and PPV is blurred, and diagnosis is based on pattern recognition. This may result in inability to translate basic science findings, underpowered trials, and misalignment of trial results with clinical practice.

A new, improved definition of ARDS was required. The consensus-based Berlin definition proposed the subdivision of ARDS into three categories based on degree of hypoxaemia. Some AECC criteria—timing of insult, CXR pattern, and evaluation of patients with left atrial hypertension—were refined, and two new variables added—an explicit ‘risk factor’ requirement and PEEP of ≥ 5 cm H₂O.7 Static compliance and expired minute ventilation were examined post hoc in a single subgroup, but not included in the final iteration, and the need to outlaw ‘non-ARDS-related illnesses’ was not discussed. In the validation dataset, the new definition did not identify a
visibly different group of patients compared with its predecessor: no ‘AECC-negative’ patients were included, and of the 4188 patients studied, 3670 satisfied the Berlin definition. Of the 518 excluded patients, all either lacked PEEP data or failed to meet Berlin PEEP criteria; the other new/modified variables had no diagnostic impact. With identical mortality in ‘PEEP-excluded’ vs ‘PEEP-included’ patients (180/518 vs 1256/3670; \( P = 0.84 \)), the Berlin definition’s improved mortality prediction appears not to come from variable selection, but from a trichotomized dataset with a new cut-off point (PFR < 100) that outperforms the dichotomized Berlin dataset (Fig. 1). In reality, without testing known ‘true positive’ and ‘true negative’ patients, and comparing the results independently against a reference standard, objective validation of a new diagnostic technique is impossible.8

DAD, incorporating histological features of hyaline membranes, oedema, cell necrosis, and fibrosis, remains the accepted morphological hallmark of ARDS.9 10 Other diseases clinically resembling ARDS (such as pneumonia or pulmonary oedema) are not associated with these histological features.9 Therefore, the presence of DAD can be used as an independent reference standard, against which clinical definitions of ARDS can be validated. Thille and colleagues recently evaluated the accuracy of the Berlin definition in 712 critically ill patients who underwent post-mortem between 1991 and 2010, using DAD as a clinico-pathological standard. Though a direct comparison with the AECC definition was not made, they found the Berlin definition to have a sensitivity of 0.89 and a specificity of 0.63 (PPV and NPV of 0.51 and 0.93, respectively); similar to the diagnostic accuracy of the AECC definition. Worryingly, if patients are only included after the introduction of a low tidal volume ventilation strategy in 2000 (\( n = 231 \)) specificity and PPV are worse, at 0.42 and 0.41, respectively. In patients meeting Berlin criteria but without morphological evidence of DAD, pneumonia was the predominant histological finding.10 The Berlin criteria also performed poorly in patients with ARDS risk factors compared with the group as a whole, with greater over diagnosis in ‘at risk’ patients compared with ‘all comers’, confirming that the definition is useful for screening—but not diagnostic—purposes (Table 2). When the reference standard was changed to ‘DAD plus pneumonia’, the Berlin model still overdiagnosed, with much lower specificity in the ‘at risk’ group (0.37) than in ‘all comers’ (0.77). Taken together, these findings confirm that the Berlin definition is unable to differentiate patients with hypoxaemic respiratory failure because of ‘true’ ARDS from those with pneumonia, pulmonary oedema, or other clinical conditions. Even when DAD and pneumonia are merged into a novel, unvalidated disease entity, the model’s poor discriminative capacity is still inadequate as an inclusion criterion for patient enrolment in clinical trials. At best, current iterations of the AECC or Berlin definitions might serve as a clinical decision rule, similar to the commonly used Wells criteria for stratifying DVT risk.11

ARDS is unusual in not having a ‘pretest probability – test – post-test probability’ diagnostic pathway; an approach that has become the norm for other serious illnesses. But behind the methodological debate, the real issue is that of pathophysiology. Although acute alveolar inflammation (the pathological process underlying ARDS) results in impaired lung function and radiographic infiltrates, these are consequences that are not unique to ARDS. Other conditions (e.g. pneumonia, pulmonary oedema, pulmonary embolus, interstitial lung disease, chronic obstructive pulmonary disease) may give rise to similar clinical findings.10 Recent iterations of the ARDS definition do not incorporate parameters that allow differentiation of hypoxemic respiratory failure as a result of alveolar inflammation (ARDS) from other pulmonary conditions. Such parameters would be reflective of the underlying pathological process (such as lung mechanics), or biomarkers that might reflect causes—or presence—of inflammatory lung processes. In this context, if we continue to apply a ‘definition’ that may be a good screening test but a poor diagnostic one, and that does not include relevant biological ‘markers’, it is difficult to see how the entity can be characterized more accurately by making small changes in the existing definition. This may explain why the vast majority of trials conducted to date have been negative: almost all have recruited patients using non-specific criteria,

---

**Table 1** Sensitivity and specificity values of ARDS definitions, when compared with autopsy evidence of diffuse alveolar damage (DAD). *Patients included from 1991 to 2002, before low tidal volume ventilation era. †Paediatric population. PPV, positive predictive value; NPV, negative predictive value

<table>
<thead>
<tr>
<th>Author</th>
<th>n</th>
<th>Definition Tested</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>AECC definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferguson and colleagues</td>
<td>138</td>
<td>AECC</td>
<td>0.83</td>
<td>0.51</td>
<td>0.43</td>
<td>0.88</td>
</tr>
<tr>
<td>Esteban, 200419*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All comers</td>
<td>382</td>
<td>AECC</td>
<td>0.75</td>
<td>0.84</td>
<td>0.66</td>
<td>0.84</td>
</tr>
<tr>
<td>Risk factors</td>
<td>284</td>
<td>AECC</td>
<td>0.76</td>
<td>0.75</td>
<td>0.66</td>
<td>0.83</td>
</tr>
<tr>
<td>Pinheiro, 200720</td>
<td>22</td>
<td>AECC</td>
<td>0.71</td>
<td>0.67</td>
<td>0.5</td>
<td>0.83</td>
</tr>
<tr>
<td>Rodriguez Martinez, 200621†</td>
<td>34</td>
<td>AECC</td>
<td>0.81</td>
<td>0.71</td>
<td>0.91</td>
<td>0.50</td>
</tr>
<tr>
<td>Alternative definitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferguson and colleagues</td>
<td>138</td>
<td>LIS</td>
<td>0.74</td>
<td>0.77</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>Ferguson and colleagues</td>
<td>138</td>
<td>Delphi</td>
<td>0.69</td>
<td>0.82</td>
<td>0.63</td>
<td>0.86</td>
</tr>
</tbody>
</table>
and many participants will have had ‘non-ARDS-related illnesses’. Underpowered trials are likely to be ‘positive’ only if the effect on ARDS is bigger than hypothesized—or if the treatment works for patients with or without ARDS—while ‘negative’ trial results are suspect (because indeterminate).

Defining complex disease states such as ARDS by using simple variables that do not reflect specific processes leads to diminishing returns, with limited improvement in sensitivity or specificity. Variables can only be reliably selected and defended using an independent reference standard. There are currently two mechanisms by which researchers can validate candidate variables against a pathological standard. First, they can be tested against the presence of DAD at autopsy in patients who died with ARDS. Admittedly, this validates only a variable in the cohort who died with the condition; however, patients who died are those with the severest form of the disease, and accurate findings in this cohort are likely to be applicable to survivors also. Secondly, there are circumstances where an invasive standard, such as lung biopsy, is appropriate. Lung biopsy may be appropriate when correct intervention requires tissue diagnosis, the prognosis is poor without definitive therapy, and tissue can be safely obtained. Given that lung biopsy is safe to perform, and changes management in more than 60% of patients, this approach is justified in a subset of patients with a (provisional) diagnosis of ARDS.5 Matching histological data against clinical status and outcome could advance the science of defining, understanding, and prognosticating in patients with acutely injured...
lungs and provide a rigorous method to identify and test accessible variables in an enhanced ARDS definition.

There are numerous candidate variables that have a rational basis for being validated as components of an ARDS definition. Lung mechanics, discarded since 1994, are accessible, predict adverse outcome, and when included in diagnostic criteria (LIS score, Delphi method) predict DAD with far superior specificity than AECC criteria. Similarily, oxygenation index, right heart dysfunction, dead space fraction, and fibroproliferation on high-resolution CT are accessible and independent early predictors of outcomes in patients with AECC-based diagnosis of ARDS. Although no single biomarker has been identified, panels of biomarkers can predict outcomes with a high degree of accuracy. Testing these and other known variables against a fixed reference standard may help identify suitable components for a new, more accurate definition.

In summary, considering the current low yield in ARDS trials, inaccurate syndrome definition is unsustainable. Progress in ARDS depends on novel therapies that can be tested rigorously in the population for which they were intended. Minor changes to an inadequate existing definition, or ad hoc changes in the reference standard, will not deliver the required improvement in diagnostic accuracy. When the ARDS syndrome definition allows us to distinguish between a clinical picture 'compatible with ARDS' (pretest probability) and one 'diagnostic of ARDS' (post-test probability), we will have a rigorous—rather than a pragmatic—model.

Acknowledgement
The authors gratefully acknowledge the assistance of Prof. Brian Kavanagh in the preparation of this manuscript.

Declaration of interest
None declared.

References