

RESEARCH

Open Access



# Clustering COVID-19 ARDS patients through the first days of ICU admission. An analysis of the CIBERESUCICOVID Cohort

Adrian Ceccato<sup>1,2,3\*</sup>, Carles Forne<sup>4,5</sup>, Lieuwe D. Bos<sup>6</sup>, Marta Camprubí-Rimblas<sup>1,2</sup>, Aina Areny-Balaguero<sup>1,2</sup>, Elena Campaña-Duel<sup>1,2</sup>, Sara Quero<sup>1,2</sup>, Emili Diaz<sup>1,2</sup>, Oriol Roca<sup>1,2</sup>, David De Gonzalo-Calvo<sup>2,7</sup>, Laia Fernández-Barat<sup>2,8</sup>, Anna Motos<sup>2,8</sup>, Ricard Ferrer<sup>9</sup>, Jordi Riera<sup>9</sup>, Jose A. Lorente<sup>10,11,2</sup>, Oscar Peñuelas<sup>10,2</sup>, Rosario Menendez<sup>12,2</sup>, Rosario Amaya-Villar<sup>13</sup>, José M. Añón<sup>14,2</sup>, Ana Balan-Mariño<sup>15</sup>, Carme Barberà<sup>16</sup>, José Barberán<sup>17</sup>, Aaron Blandino-Ortiz<sup>18,19</sup>, Maria Victoria Boado<sup>20</sup>, Elena Bustamante-Munguira<sup>2,21</sup>, Jesús Caballero<sup>22</sup>, Cristina Carbajales<sup>23</sup>, Nieves Carbonell<sup>24</sup>, Mercedes Catalán-González<sup>25</sup>, Nieves Franco<sup>26</sup>, Cristóbal Galbán<sup>27</sup>, Víctor D. Gumucio-Sanguino<sup>28</sup>, Maria del Carmen de la Torre<sup>2,29</sup>, Ángel Estella<sup>30</sup>, Elena Gallego<sup>31</sup>, José Luis García-Garmendia<sup>32</sup>, José Garnacho-Montero<sup>33</sup>, José M. Gómez<sup>34</sup>, Arturo Huerta<sup>35</sup>, Ruth Noemí Jorge-García<sup>36</sup>, Ana Loza-Vázquez<sup>37</sup>, Judith Marin-Corral<sup>38</sup>, Amalia Martínez de la Gándara<sup>39</sup>, María Cruz Martín-Delgado<sup>40</sup>, Ignacio Martínez-Varela<sup>41</sup>, Juan Lopez Messa<sup>42</sup>, Guillermo Muñoz-Albaiceta<sup>2,43</sup>, María Teresa Nieto<sup>44</sup>, Mariana Andrea Novo<sup>45</sup>, Yhivian Peñasco<sup>46</sup>, Juan Carlos Pozo-Laderas<sup>47</sup>, Felipe Pérez-García<sup>48,49</sup>, Pilar Ricart<sup>50</sup>, Ferran Roche-Campo<sup>51</sup>, Alejandro Rodríguez<sup>2,52</sup>, Victor Sagredo<sup>53</sup>, Angel Sánchez-Miralles<sup>54</sup>, Susana Sancho-Chinesta<sup>55</sup>, Lorenzo Socias<sup>56</sup>, Jordi Solé-Violan<sup>2,57</sup>, Fernando Suarez-Sipmann<sup>2,58</sup>, Luis Tamayo-Lomas<sup>2,59</sup>, José Trenado<sup>60</sup>, Alejandro Úbeda<sup>61</sup>, Luis Jorge Valdivia<sup>62</sup>, Pablo Vidal<sup>63</sup>, Jesus Bermejo<sup>2,64,65</sup>, Jessica Gonzalez<sup>2,7</sup>, Ferran Barbe<sup>2,7</sup>, Carolyn S. Calfee<sup>66</sup>, Antonio Artigas<sup>1,2,\*†</sup> and Antoni Torres<sup>2,8†</sup> on behalf of CIBERESUCICOVID Project

## Abstract

**Background** Acute respiratory distress syndrome (ARDS) can be classified into sub-phenotypes according to different inflammatory/clinical status. Prognostic enrichment was achieved by grouping patients into hypoinflammatory or hyperinflammatory sub-phenotypes, even though the time of analysis may change the classification according to treatment response or disease evolution. We aimed to evaluate when patients can be clustered in more than 1 group, and how they may change the clustering of patients using data of baseline or day 3, and the prognosis of patients according to their evolution by changing or not the cluster.

<sup>†</sup>Antonio Artigas and Antoni Torres have contributed equally.

\*Correspondence:

Adrian Ceccato  
aaceccato@tauli.cat  
Antonio Artigas  
aartigas@tauli.cat

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Methods** Multicenter, observational prospective, and retrospective study of patients admitted due to ARDS related to COVID-19 infection in Spain. Patients were grouped according to a clustering mixed-type data algorithm (k-prototypes) using continuous and categorical readily available variables at baseline and day 3.

**Results** Of 6205 patients, 3743 (60%) were included in the study. According to silhouette analysis, patients were grouped in two clusters. At baseline, 1402 (37%) patients were included in cluster 1 and 2341 (63%) in cluster 2. On day 3, 1557 (42%) patients were included in cluster 1 and 2086 (57%) in cluster 2. The patients included in cluster 2 were older and more frequently hypertensive and had a higher prevalence of shock, organ dysfunction, inflammatory biomarkers, and worst respiratory indexes at both time points. The 90-day mortality was higher in cluster 2 at both clustering processes (43.8% [ $n=1025$ ] versus 27.3% [ $n=383$ ] at baseline, and 49% [ $n=1023$ ] versus 20.6% [ $n=321$ ] on day 3). Four hundred and fifty-eight (33%) patients clustered in the first group were clustered in the second group on day 3. In contrast, 638 (27%) patients clustered in the second group were clustered in the first group on day 3.

**Conclusions** During the first days, patients can be clustered into two groups and the process of clustering patients may change as they continue to evolve. This means that despite a vast majority of patients remaining in the same cluster, a minority reaching 33% of patients analyzed may be re-categorized into different clusters based on their progress. Such changes can significantly impact their prognosis.

**Keywords** ARDS, Clustering, Mortality, Precision medicine

## Background

Acute respiratory distress syndrome (ARDS) is a heterogeneous condition characterized by respiratory failure and diffuse pulmonary noncardiogenic edema [1]. ARDS is associated with poor short- and long-term outcomes including high mortality rates, disability, and poor quality of life indicators. Several trials have failed to identify treatments for ARDS, and only corticosteroids [2], prone position [3], and limiting the damage caused by mechanical ventilation [4, 5] have shown improved outcomes.

The COVID-19 pandemic hit Europe in late winter/early spring 2020, probably resulting in the highest number of ARDS patients ever treated concomitantly. COVID-19 has increased the knowledge about the pathophysiology and treatment of ARDS, highlighting the importance of anti-inflammatory treatment over antivirals [6–9], thromboembolic events [10], and endothelial damage [11].

Sub-phenotypes of ARDS have been described as hypoinflammatory and hyperinflammatory, using clinical data and plasma inflammatory biomarkers in latent class analysis [12]. These sub-phenotypes have been associated with different outcomes, and reanalyses of several clinical trials found that the hyperinflammatory sub-phenotype may have better outcomes when receiving simvastatin [13], or high-positive end-expiratory pressure (PEEP) levels [12]. In addition, these sub-phenotypes responded differently to fluid management [14]. Models for a parsimonious identification of sub-phenotypes using a few variables and machine-learning strategies are being developed. However, during the COVID-19 pandemic, fewer patients could be classified into the hyperinflammatory sub-phenotype [15]. A possible limitation

of clustering is that the correct identification of sub-phenotypes is needed, and only one cross-sectional evaluation may underestimate the impact of the changes or evolutions of markers in the first days of admission. Bos et al. failed to identify sub-phenotypes in patients with COVID-19 by cross-sectional analysis with data from the respiratory system but two sub-phenotypes were identified through longitudinal analysis [16]. Few data evaluating clustering using inflammation data and common variables measuring organ failure are available [17, 18].

We hypothesized that patients with COVID-19-associated ARDS may be described by more than one discrete cluster and the clustering process and its result may change during the first days of their admission, which may have an impact on their prognosis. To evaluate this, we aimed to study a cohort of patients and analyze the clustering of patients at baseline and day 3 using readily available data measuring organ failure, respiratory variables, and inflammation.

## Methods

### Study design

CIBERESUCICOVID is a multicenter, observational, prospective, retrospective cohort study that enrolled patients with COVID-19 admitted to Spanish intensive care units (ICUs) [19]. The study was approved by the institution's Internal Review Board (Comité Ètic d'Investigació Clínica, registry number HCB/2020/0370), and written informed consent was obtained from either the patients or their relatives. Local researchers were contacted by a member of the study team, and the participating hospitals obtained approval from the local ethics committee. De-identified patient data were collected

and stored via the REDCap electronic data capture tool [20, 21] hosted at the Centro de Investigación Biomédica en Red (CIBER), Spain. Trained local researchers incorporated data from the patient's medical records into a separate database. Before statistical analyses, the data were checked by three independent experienced data collectors trained in critical care, and site investigators were contacted for any queries. Missing analyses were performed, and site investigators were contacted to obtain as much reliable and complete data as possible. Results were reported following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [22].

### Study population and data collection

All consecutive patients admitted to the ICU at a participating center from February 25, 2020, to September 30, 2021, were enrolled if they fulfilled the following criteria:  $\geq 18$  years of age, admission to the ICU, and laboratory-confirmed SARS-CoV-2 infection. For this study, we selected those patients with a diagnosis of ARDS based on the Berlin definition [23] who needed invasive mechanical ventilation during the first day of admission to the ICU and who remained ventilated at least 3 days later. Patients were excluded if they had a non-confirmed SARS-CoV-2 infection, had no data at baseline, or were admitted to the ICU for other reasons.

After enrollment, prior epidemiological data, including demographics, comorbidities, clinical symptoms, disease chronology, and treatment administered upon hospital admission, were collected. The site researchers subsequently collected data acquired at hospital admission, ICU admission, start of mechanical ventilation (MV), 72–96 h after ICU admission, weaning, ICU discharge, and hospital discharge, including vital signs, respiratory support devices (i.e., oxygen mask, high-flow nasal cannula, and noninvasive and invasive mechanical ventilation), use of adjunctive therapies (i.e., neuromuscular blockade, prone position, and recruitment maneuvers), laboratory findings, arterial blood gases, and mechanical ventilation settings, if appropriate. Hemodynamic parameters and organ dysfunction were assessed using the Sequential Organ Assessment Failure Score (SOFA) [24] at ICU admission. Pharmacological treatments administered upon and during ICU admission until discharge from the ICU or hospital or death were also collected.

Specific data regarding MV since the start of invasive mechanical ventilation and on day three were analyzed. MV parameters related to ventilation-induced lung injury (VILI), such as tidal volume, respiratory rate, end-inspiratory plateau, peak inspiratory pressures, positive end-expiratory pressure (PEEP), driving pressure, and static compliance of the respiratory system (Crs) were

collected. Oxygenation impairment was analyzed using the PaO<sub>2</sub>/FiO<sub>2</sub> ratio, and abnormalities in CO<sub>2</sub> metabolism were studied using the ventilatory ratio (VR), a surrogate parameter of Vd/Vt. The worst event values were recorded preferentially. We did not analyze data from inflammatory cytokines.

### Definitions

The diagnosis of ARDS was based on the Berlin definition [23]. Tidal volume was reported in mL/kg of predicted body weight (PBW). The driving pressure was defined as plateau pressure minus PEEP. Crs was calculated as tidal volume/plateau pressure—PEEP. The ventilatory ratio was defined as follows:  $(\text{minute ventilation} \times \text{PaCO}_2) / (\text{PBW} \times 100 \times 37.5)$ .

### Outcome

The primary outcome was the 90-day mortality.

### Statistical analysis

To derive the clusters, we first evaluated the distributions, missingness, and correlation of the candidate variables (Additional file 1: Fig. S1). Multiple imputation with chained equations was used to account for missing data, the results from the first dataset are presented, and the additional datasets were evaluated for consistency. After evaluating correlation, we excluded highly correlated variables ( $|\rho| > 0.5$ ) and those variables with more than 60% of missing values. Ordering Points To Identify the Clustering Structure plots were used to determine the optimal clustering strategy [25]. Based on these plots, we applied the k-prototypes clustering to 44 variables (age, gender, height, weight, chronic heart failure, diabetes mellitus, obesity, hypertension, hematologic disease, cancer, immunosuppression, systolic blood pressure, mean blood pressure, shock, temperature, heart rate, respiratory rate, pH, Hco<sub>3</sub>, hemoglobin, white blood counts, lymphocytes, platelet, D-dimer, prothrombin time, C-reactive protein, bilirubin, lactate, LDH, sodium, potassium, creatinine, procalcitonin, albumin, ventilatory ratio, tidal volume/ predicted body weight, driving pressure, Pao<sub>2</sub>/FiO<sub>2</sub>, PEEP, peak pressure, neuromuscular blocking agents use, prone position, ECMO) collected at baseline and replicated with data collected at day 3 (33 of 44 variables changed their values according to variable evolution) using a partitioning approach. The optimal number of clusters was determined based on the Silhouette index [26, 27]. Once the optimal number of clusters was determined, patterns of clinical variables were visualized in two ways: (1) chord plots (showing how clusters differ based on categorical variables) and (2) standardized values of each continuous variable by cluster.

To understand the implications of clusters and changes of clusters on 90-day mortality, Kaplan–Meier survival curves and hazard ratios adjusted by SOFA score were estimated, with 95% confidence intervals (CI).

In order to assess whether the clusters were explained by well-established measures of disease severity, we fitted logistic regression models including the SOFA score and/or the most influential variables in clustering as independent variables, and the C-indexes were obtained.

Data were presented as means and standard deviations (SD) or medians and percentiles of 25% and 75%. For comparisons between clusters, we used t tests and the Mann–Whitney U test for continuous data and the  $\chi^2$  test for categorical data. The statistical significance was set at 0.05 for two-sided tests (without adjustment for multiple comparisons; therefore, statistically significant differences between clusters should be considered exploratory). Analysis was performed with the R statistical package version 4.3.1 (R Foundation for Statistical Computing).

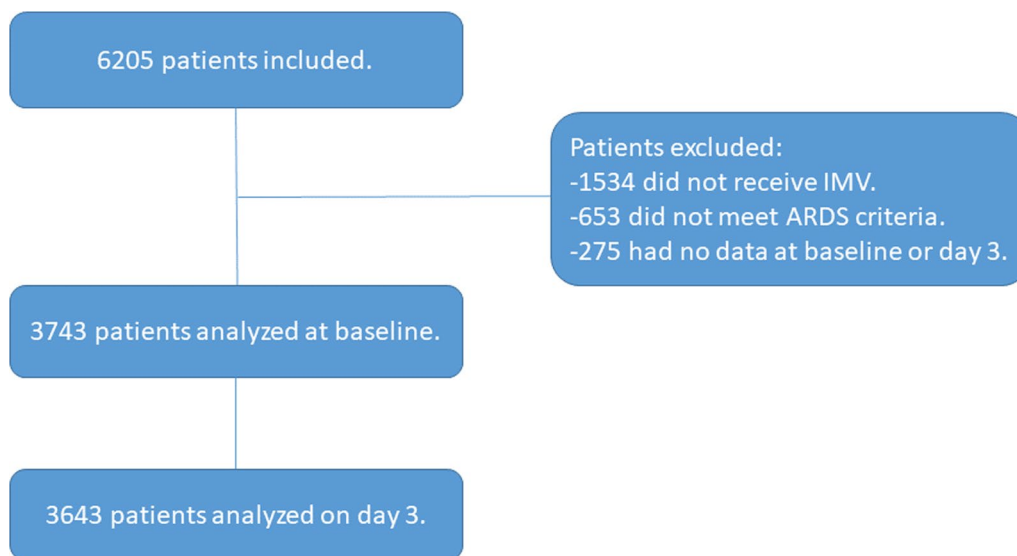
**Results**

Out of 6205 eligible patients, 3743 (60%) patients were included in the study (Fig. 1). Following the Silhouette statistic (Additional file 1: Fig. S3), patients were grouped into two clusters using the k-prototypes algorithm. At baseline, 1402 (37%) patients were included in cluster 1 and 2341 (63%) in cluster 2. The baseline characteristics of the patients are presented in Table 1. The patients included in cluster 2 were older, had more hypertension, shock, inflammation measured by CRP, and better respiratory indexes (including the

requirement of the prone position, NMBA requirement, oxygenation by Pao2/Fio2, ventilatory ratio, and driving pressure) (Table 1). On day 3, 3643 patients were analyzed, 1557 (42%) patients were included in cluster 1 and 2086 (57%) in cluster 2. One hundred patients (3%) died or were discharged before day 3. The weight of each continuous variable in the clusters is shown in Fig. 2 as standardized variable values and counts for categorical variables. In the clustering process between baseline and day 3, several variables changed their influence with an evolution of the values for each variable, but at both moments the clinical characteristics of the clustering were similar.

The 90-day mortality for patients included in cluster 2 at baseline was 43.8% ( $n=1025$ ) versus 27.3% in cluster one ( $n=383$ ) (adjusted by SOFA score hazard ratio [aHR] 1.44, 95% confidence interval [95% CI] 1.21 to 1.70). The 90-day mortality was 49% ( $n=1023$ ) for patients included in cluster 2 on day 3 versus 20.6% ( $n=321$ ) for patients in cluster 1 (aHR 2.18, 95% CI 1.80–2.63). Unadjusted Kaplan–Meier survival curves are shown in Fig. 3.

Four hundred and fifty-eight (33%) patients clustered in the first group were clustered in the second group on day 3. In contrast, 638 (27%) patients clustered in the second group, were clustered in the first group on day 3. Lower survival probability at 90 days was observed in those patients grouped in cluster 2 at both time points compared to those who were grouped at cluster 1 at both time points (aHR 2.45, 95%CI 1.89–3.19), followed by those who changed from cluster 1 to cluster 2 over time (aHR 2.11, 95% CI 1.56–2.87). Those who changed from



**Fig. 1** Flowchart of patient screening and enrollment

**Table 1** Demographic and clinical characteristics of patients by clusters at baseline and day 3

	Missing values, n (%) at baseline/n (%) at day 3	Clustering at baseline		Clustering on day 3	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
Age, years	4 (0.11%)	60 [50,69]	67 [60,72]	61 [51,69]	67 [60,73]
Sex, female	4 (0.11%)	494 (35%)	614 (26%)	574 (37%)	505 (24%)
Active smoker	730 (19.5)	68 (6.09%)	134 (7.06%)	81 (6.40%)	115 (6.93%)
Hypertension	15 (0.40%)	306 (22%)	1633 (70%)	463 (30%)	1420 (68%)
Diabetes mellitus	19 (0.51%)	221 (16%)	761 (33%)	296 (19%)	659 (32%)
Obesity	14 (0.37%)	459 (33%)	890 (38%)	529 (34%)	790 (38%)
Chronic heart failure	359 (9.59%)	112 (9%)	380 (18%)	145 (10%)	327 (18%)
Chronic kidney disease	11 (0.29%)	49 (3%)	171 (7%)	50 (3%)	157 (7%)
Chronic respiratory disease	11 (0.29%)	92 (7%)	280 (12%)	113 (7%)	246 (12%)
Hematologic disease	402 (10.7%)	48 (4%)	108 (5%)	66 (5%)	85 (5%)
Malignant neoplasm	394 (10.5%)	40 (3%)	78 (4%)	44 (3%)	72 (4%)
Immunosuppression	411 (10.9%)	24 (2%)	42 (2%)	20 (2%)	42 (2%)
APACHE II score at ICU admission	1743 (46.6%)	12 [9, 16]	14 [10, 19]	–	–
SOFA score	1411 (37.7%)/1685 (45%)	5 [4, 7]	7 [6, 8]	4 [3, 6]	7 [6, 8]
Temperature, °C	649 (17.3%)/875 (23.4%)	36.9 [36.1,37.7]	36.8 [36.0,37.6]	36.7 [36.0,37.2]	36.8 [36.0,37.5]
Respiratory rate, bpm	765 (20.4%)/1011 (27%)	24 [20, 31]	25 [20, 30]	20 [18, 24]	23 [20, 26]
Systolic blood pressure, mmHg	589 (15.7%)/925 (24.7%)	122 [105,140]	120 [101,140]	130 [113,151]	125 [109,146]
Mean blood pressure, mmHg	1376 (36.8%)/1510 (40.3%)	85 [72,98]	83 [70,96]	90 [75,103]	82 [71,96]
Heart rate, bpm	620 (16.6%)/876 (23.4%)	85 [70,100]	87 [70,102]	70 [55,89]	80 [60,98]
PaO <sub>2</sub> /FIO <sub>2</sub> ratio, mmHg	310 (8.28%)/359 (9.59%)	119 [83,177]	110 [80.6,154]	198 [156,250]	161 [122,210]
PaO <sub>2</sub> mmhg	204 (5.45%)/328 (8.76%)	84 [67.0,114]	84[66.9,110]	88 [75,110]	86 [73,107]
Vasopressor use	419 (11.2%)/555 (14.8%)	352 (28%)	1718 (82%)	444 (32%)	1434 (79%)
pH	107 (2.86%)/310 (8.28%)	7.40 [7.34,7.45]	7.34 [7.27,7.41]	7.43 [7.40,7.46]	7.37 [7.31,7.42]
PaCO <sub>2</sub> , mmHg	117 (3.13%)/305 (8.15%)	41 [35.0,47.9]	45 [37.1,53.8]	43 [39.0,47.3]	48 [42.6,55.4]
Lactate, mmol/L	817 (21.8%)/1001 (26.7%)	1.40 [1.10,1.90]	1.50 [1.10,2.10]	1.62 [1.20,2.10]	1.80 [1.36,2.30]
Hemoglobin, g/dl	546 (14.6%)/684 (18.3%)	13.0 [11.8,14.1]	13.2 [11.9,14.4]	11.8 [10.7,13.0]	12.1 [10.9,13.3]
White blood cell count, 10 <sup>9</sup> /L	52 (1.39%)/393 (10.5%)	8.96 [6.39,12.8]	10.0 [7.07,13.9]	8.30 [6.30,10.9]	9.82 [7.20,13.5]
Lymphocyte count, 10 <sup>9</sup> /L	146 (3.90%)/407 (10.9%)	0.64 [0.42,0.91]	0.62 [0.41,0.90]	0.70 [0.48,1.05]	0.60 [0.40,0.90]
Neutrophil count, 10 <sup>9</sup> /L	265 (7.08%)/468 (12.5%)	7.6 [5.20,11.3]	8.62 [5.96,12.5]	6.8 [4.92,9.20]	8.41 [5.90,11.9]
Platelet count, 10 <sup>9</sup> /L	58 (1.55%)/331 (8.84%)	236 [181,312]	230 [174,302]	277 [212,350]	242 [179,316]
D-dimers, mg/L	705 (18.8%)/1191 (31.8%)	1197 [600,3398]	1371 [698,4110]	2000 [940,5174]	2498 [1102,7323]
Ferritin, ng/mL	2123 (56.7%)/2456 (65.6%)	1109 [561,1995]	1248 [678,1976]	994 [526,1690]	1214 [722,1928]
IL6, pg/mL	2747 (73.4%)/3112 (83.1%)	119 [38.3,299]	116 [40.1,272]	60.8 [15.4,242]	140 [35.5,555]
CRP, mg/dL	332 (8.87%)/700 (18.7%)	126 [58.1,215]	155 [72.3,249]	42.0 [17.1,95.1]	77.4 [26.4,193]
Procalcitonin, µg/l	1137 (30.4%)/1830 (48.9%)	0.20 [0.09,0.50]	0.27 [0.13,0.80]	0.12 [0.07,0.32]	0.36 [0.13,1.20]
Lactate dehydrogenase, U/l	612 (16.2%)/1105 (29.2%)	490 [376,652]	542 [414,740]	379 [304,492]	434 [338,585]
Troponin T, ng/l	3047 (81.4%)/3244 (86.7%)	0.01 [0.01,0.04]	0.02 [0.01,0.07]	0.02 [0.01,0.07]	0.02 [0.01,0.0]
NT-proBNP, pg/mL	3132 (83.1%)/3330 (89%)	314 [123,785]	642 [245,1763]	225 [108,567]	676 [239,2156]
Prothrombin time, seg	1239 (33.1%)/1474 (39.4%)	13.0 [12.1,14.2]	13.3 [12.2,14.7]	12.7 [11.9,13.9]	13.0 [11.9,14.4]
Bilirubin, mg/dL	476 (12.7%)/822 (22.0%)	0.60 [0.40,0.90]	0.60 [0.40,0.90]	0.50 [0.31,0.80]	0.60 [0.38,1.10]
Albumin, g/dL	1751 (46.8%)/1825 (48.8%)	3.10 [2.70,3.40]	3.10 [2.70,3.40]	2.88 [2.59,3.10]	2.64 [2.37,3.00]
Serum creatinine, mg/dL	42 (1.12%)/326 (8.71%)	0.75 [0.62,0.94]	0.93 [0.72,1.26]	0.70 [0.56,0.90]	0.99 [0.71,1.5]
Sodium, mmol/L	467 (12.5%)/633 (16.9%)	139 [136,141]	138 [135,141]	142 [139,145]	141 [139,145]
Potassium, mmol/L	486 (13%)/640 (17.1%)	4 [3.66,4.40]	4.10 [3.70,4.50]	4.06 [3.70,4.40]	4.30 [3.90,4.70]
Tidal volume/PBW (mL/kg)	1264 (33.8%)/1155 (30.9%)	7.09 [6.38,7.91]	7.03 [6.38,7.81]	7.19 [6.46,8.0]	7.08 [6.39,7.83]
PEEP, cmH <sub>2</sub> O	537 (14.3%)/366 (9.78%)	12 [10, 14]	12 [10, 14]	12 [10, 13]	12 [10, 14]
FiO <sub>2</sub> , %	100 (2.67%)/219 (5.85%)	80 [60,100]	90 [65,100]	45 [40,55]	55 [45,70]
Peak inspiratory pressure, cmH <sub>2</sub> O	1869 (49.9%)/1921 (51.0%)	30 [27, 34]	31 [28, 35]	28 [25, 32]	32 [28, 36]

**Table 1** (continued)

	Missing values, <i>n</i> (%) at baseline/ <i>n</i> (%) at day 3	Clustering at baseline		Clustering on day 3	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
End-inspiratory plateau pressure, cmH <sub>2</sub> O	2158 (57.7%)/2322 (61.3%)	24 [21, 27]	25 [22, 28]	23 [20, 26]	25 [22, 28]
Driving pressure, cmH <sub>2</sub> O <sup>a</sup>	2172 (58%)/2334(60%)	12 [10, 15]	12.5 [10, 15]	12 [9, 14]	12 [10, 15]
Compliance, mL/cmH <sub>2</sub> O <sup>b</sup>	2284 (61%)/2422 (64.7%)	35.8 [28.2,46.0]	35.7 [29.1,47]	37.8 [30,50]	36.4 [28.7,46.4]
Ventilatory ratio <sup>c</sup>	1346 (36%)/1279 (34.2%)	1.59 [1.28,1.95]	1.75 [1.40,2.24]	1.66 [1.40,1.99]	2.05 [1.69,2.50]
Neuromuscular blocking agent use	425 (11.4%)/555 (14.8%)	430 (35%)	1683 (81%)	551 (40%)	1394 (77%)
Position	681 (18.2%)/769 (20.5%)				
Supine		778 (70%)	1135 (58%)	946 (75%)	1048 (61%)
Prone		319 (28%)	770 (40%)	282 (23%)	631 (37%)
Other		18(2%)	42(2%)	29 (2%)	38 (2%)

Continuous variables are expressed as median (IQR) and categorical variables as numbers (percentages)

CRP C-reactive protein; FiO<sub>2</sub> fraction of inspired oxygen; IL Interleukin; MV mechanical ventilation; NT-proBNP N-terminal pro-brain natriuretic peptide; PaCO<sub>2</sub> arterial partial pressure of carbon dioxide; PaO<sub>2</sub> partial pressure of arterial oxygen; PBW predicted body weight; SOFA sequential organ failure assessment score

<sup>a</sup> Defined as plateau pressure—PEEP

<sup>b</sup> Defined as tidal volume/ (Plateau pressure – PEEP)

<sup>c</sup> Defined as (minute ventilation × PaCO<sub>2</sub>) / (PBW × 100 × 37.5)

cluster 2 to cluster 1 overtime had similar survival probability (aHR 1.18, 95% CI 0.85–1.62) to those patients grouped at cluster 1 at both moments (Table 2, Figs. 3 and 4).

We conducted an evaluation to determine if we could cluster patients using parsimonious variables. To do this, we assessed the ability to group patients using C-index for SOFA score at baseline and on the third day. The results showed that the C-index was 0.74 (95% CI 0.72–0.76) for baseline and 0.79 (95% CI 0.77–0.81) for the third day. We further analyzed the results by selecting the four most differential variables of each clustering process (age, glucose, sodium bicarbonate and pH at baseline and ventilator ratio, peak inspiratory pressure, PaO<sub>2</sub>/FiO<sub>2</sub>, and pH on day 3). The C-index was 0.74 (95% CI 0.72–0.76) for baseline and 0.82 (95% CI 0.80–0.85) for the third day.

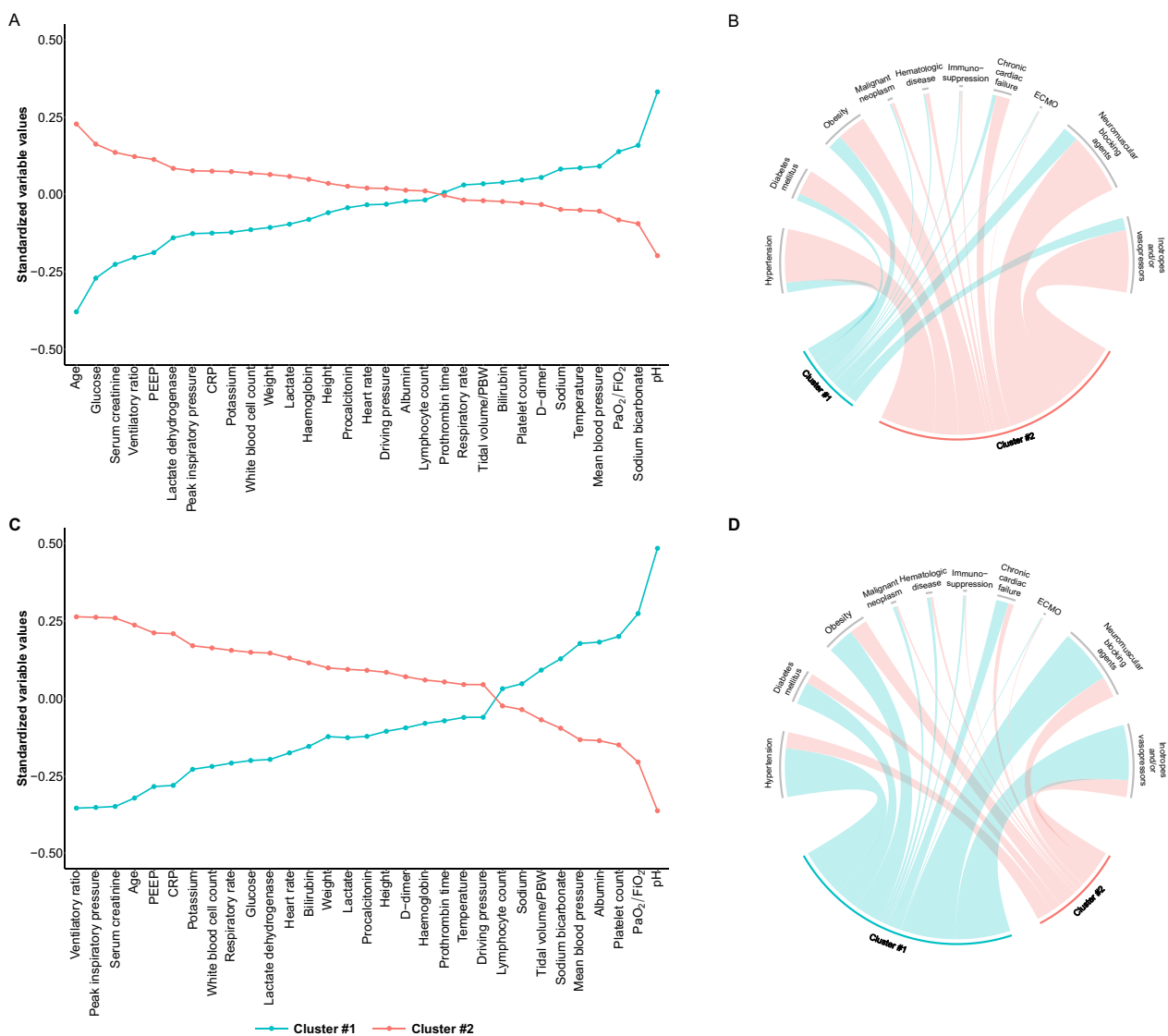
### Discussion

In our study, we found that patients with COVID-19-related ARDS could be clustered into two groups by k-prototypes at baseline and on day 3. Patients with more severe organ dysfunction, older age, worse oxygenation, respiratory mechanics, higher ventilatory ratio, pH, bicarbonate, and inflammatory markers were clustered in cluster 2 and had the highest mortality at both times. It is noteworthy that patients may change the cluster between the analyses at baseline or during day 3, improving or worsening their prognosis.

Strategies based on personalized medicine involve the identification of groups or clusters of patients with similar clinical and biological characteristics in order to tailor

treatment to populations where it is most likely to be beneficial [28, 29]. This is usually not easy, as some conditions common to critically ill patients, such as sepsis or ARDS, are heterogeneous, share common causes and inflammatory pathways, and may differ in other ways. Clustering based on common clinical and biological characteristics has been proposed as a strategy to identify homogeneous populations to which treatments can be applied [30–32]. Post hoc analysis of negative clinical trials has shown that the same treatments may be beneficial in certain clusters. The clustering of ARDS patients was developed by cross-sectional analysis in the first hours of development; however, clinical characteristics and biological responses are dynamic and may change. In sepsis, a hyperinflammatory phase followed by a regulatory hypoinflammatory response has been described [33]. In pneumonia, markers of hyperinflammation may be followed by signs of immunosuppression, such as lymphocytopenia, with a more dysregulated response for both phases being associated with the worst prognosis [34, 35].

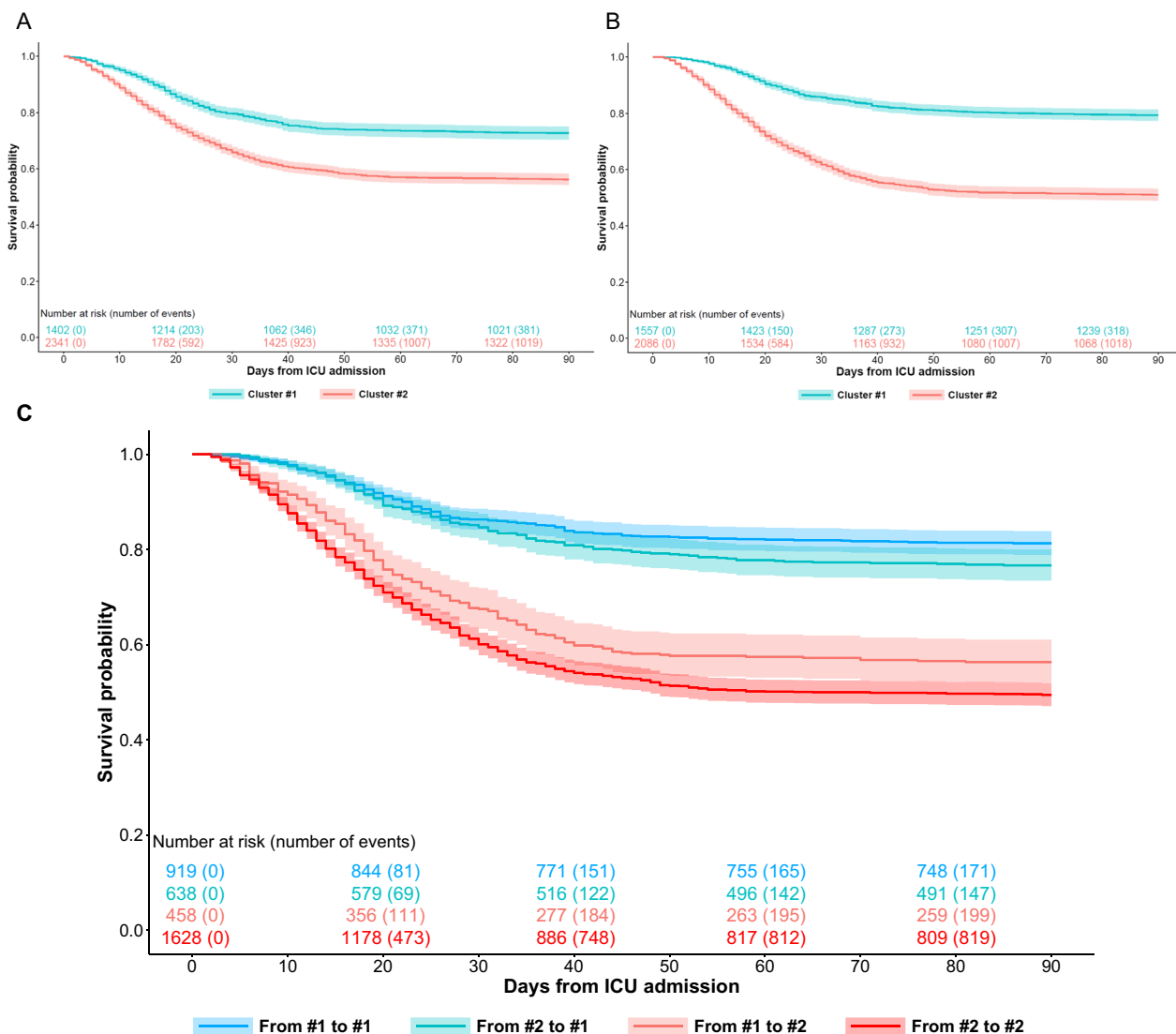
Typically, clinical trials aim to enroll patients within the first 24–48 h of admission. However, there may be a delay before treatment is administered during which the sub-phenotype classification of patients may change. In real-life situations, this delay may be even longer. Therefore, it is critical to continuously evaluate readily available measures that can dynamically classify patients and account for changes in sub-phenotypes. This is necessary because it affects prognosis. A previous study showed that sub-phenotypes analyzed by latent class analysis showed stability from baseline



**Fig. 2** Comparison of variables that contribute to clusters. **A** Standardized values of each continuous variable by cluster at baseline. **B** Chord plots (showing how clusters differ based on categorical variables) at baseline. **C** Standardized values of each continuous variable by cluster at day 3. **D** Chord plots (showing how clusters differ based on categorical variables) at day 3. CRP C-reactive protein; FIO<sub>2</sub> fraction of inspired oxygen; IL Interleukin; MV mechanical ventilation; NT-proBNP N-terminal pro-brain natriuretic peptide; PaCO<sub>2</sub> arterial partial pressure of carbon dioxide; PaO<sub>2</sub>, partial pressure of arterial oxygen; PBW, predicted body weight; SOFA: sequential organ failure assessment score. <sup>a</sup>Defined as plateau pressure—PEEP. <sup>b</sup>Defined as tidal volume/ (Plateau pressure — PEEP). <sup>c</sup>Defined as (minute ventilation × PaCO<sub>2</sub>) / (PBW × 100 × 37.5)

to day 3 and 94% remained in the same class [36]. Bos et al. did discover two trajectories related to ventilatory ratio and mechanical power that could potentially predict the duration of mechanical ventilation and the risk of death [16]. More recently, Chen and colleagues [37] described 3 longitudinal phenotypes with several changes between phenotypes during the first 4 days. Lu et al. analyzed the incidence of sub-phenotypes and their trajectories during 10 days according to respiratory support [38]. Sinha et al. observed that

transcriptional profiling of the phenotypes reveals divergent biological signatures and changes in gene expression over time [39]. These results are in line with our findings and emphasize the need to classify patients just before planning an intervention taking into account the changes in clinical and analytical variables. According to our results, the probability of survival at 90 days is better predicted by clustering at day 3 (Fig. 3b, c), and the differences between clusters are also greater at day 3, but as expected with lower values compared



**Fig. 3** Unadjusted survival curves for clusters

to baseline (Table 1). Whether prognostic enrichment could be modified over time is also an important point to resolve.

Previous clustering processes were based on the inflammatory status, a fundamental key in the pathogenesis of ARDS. However, it is not well known how much inflammation is enough to damage the lungs. Whether or not a cytokine storm is associated with critical COVID-19 has been discussed in relation to anti-inflammatory treatment. It is recognized that in critically COVID-19 there is less inflammation than other diseases [40], although the impact on organ failure is large. It seems that the clustering process in our study is mostly influenced by organ failure, as suggested

by the individual SOFA score components (as shown in Fig. 2). This is significant because the SOFA score is commonly used and can accurately predict how the subjects are assigned to clusters in our cohort. Changes in SOFA scores have been used to measure the outcome of several studies and as an indicator of poor prognosis. However, the prognosis of clustering in our cohort is independent of the SOFA score, as demonstrated by the adjusted HR.

We decided to use k-prototypes, an unsupervised machine-learning method for clustering. This method has shown good performance in identifying clusters for heterogeneous data [41] and allowed us to include continuous and categorical variables that we considered

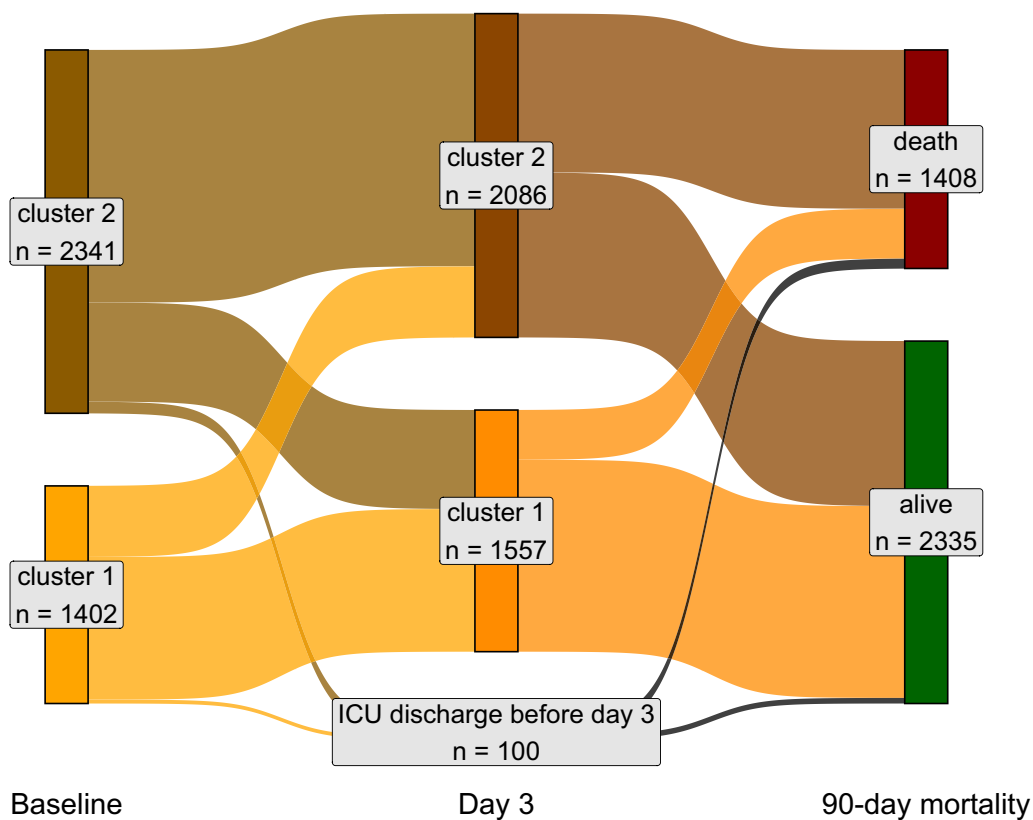


**Table 2** Unadjusted hazard ratio (95% CI) for 90-day survival probability by clusters

	90-day mortality	Hazard ratio	95% CI
At baseline			
Cluster 1 (n = 1402)	383 (27.3%)	Reference	
Cluster 2 (n = 2341)	1025 (43.8%)	1.82	1.62–2.05
Day 3			
Cluster 1 (n = 1557)	321 (20.6%)	Reference	
Cluster 2 (n = 2086)	1023 (49.0%)	2.97	2.62–3.37
Change from baseline to day 3			
From 1 to 1 (n = 919)	172 (18.7%)	Reference	
From 1 to 2 (n = 458)	200 (43.7%)	2.77	2.26–3.40
From 2 to 1 (n = 638)	149 (23.4%)	1.26	1.02–1.58
From 2 to 2 (n = 1628)	823 (50.6%)	3.45	2.93–4.07

clinically relevant. Our results are similar to the clusters developed by latent class analysis in the order of inflammation, severity, respiratory parameters, and prognosis, and we found a significant difference in mortality between clusters at both time points evaluated.

It is important to acknowledge both the strengths and limitations of our study. On the positive side, we were able to collect data from a large and well-defined group of patients. However, it is important to note that our study focused only on patients with COVID-19-associated ARDS, and some results may be specific to COVID-19 rather than ARDS. However, most ARDS studies show pneumonia as the primary cause, which may help to mitigate this limitation. Other limitations are that we did not collect a complete biological, physiological or radiological assessment with cytokines, esophageal pressure or tomographic measurements for all patients. However, we were able to include several standard analytical variables and found clusters with similar characteristics as previously described. Furthermore, we included several variables that could vary according to the management of each center and also used all data as a derivation cohort and did not validate our results in another cohort. Finally, we found that the clustering process was different at baseline and day 3, with several variables changing their weight and the differences between clusters becoming more pronounced at day 3.



**Fig. 4** Sankey plot for ARDS population

## Conclusion

During the first few days, patients can be clustered into two groups and the process of clustering patients may change as they continue to evolve. This means that despite a vast majority of patients remaining in the same cluster a minority reaching 33% of patients analyzed may be re-categorized into different clusters based on their progress. Such changes can significantly impact their prognosis. Such changes can significantly impact their prognosis, either positively or negatively. To ensure better classification, the clustering process must take into account the evolving condition of patients.

## Abbreviations

ARDS	Acute respiratory distress syndrome
HR	Hazard ratio
ICUs	Intensive care units
MV	Mechanical ventilation
PEEP	Positive end-expiratory pressure
SOFA	Sequential Organ Assessment Failure Score

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13054-024-04876-5>.

**Additional file 1.** Clustering COVID-19 ARDS patients through the first days of ICU admission. An analysis of the CIBERESUCICOVID Cohort.

## Acknowledgements

We are indebted to all our medical and nursing colleagues for their assistance and cooperation in this study.

**CIBERESUCICOVID Collaborators:** Berta Adell-Serrano, María Aguilar Cabello, Luciano Aguilera, Victoria Alcaraz-Serrano, Cesar Aldecoa, Cynthia Alegre, Raquel Almansa, Sergio Álvarez, Antonio Álvarez Ruiz, Rosario Amaya Villar, Ruth Andrea, Mariana Andrea Novo, José Ángel, Jose Manuel Añon, Marta Arrieta, J Ignacio Ayestarán, Joan Ramon Badia, Mariona Badía, Orville Báez Pravia, Ana Balan Mariño, Begoña Balsera, Carme Barberà, José Barberán, Laura Barbena, Enric Barbeta, Tommaso Bardi, Patricia Barral Segade, Marta Barroso, José Ángel Berezo García, Jesús F Bermejo-Martin, Belén Beteré, Judit Bigas, Aaron Blandino Ortiz, Rafael Blancas, María Luisa Blasco Cortés, María Boado, María Bodi Saera, Neus Bofill, María Teresa Bouza Vieiro, Leticia Bueno, Elena Bustamante-Munguira, Juan Bustamante-Munguira, Cecilia del Busto Martínez, Jesús Caballero, David Campi Hermoso, Sandra Campos Fernández, Cristina Carbajales, Iosune Cano, María Luisa Cantón-Bulnes, Nieves Carbonell, Pablo Cardina Fernández, Laura Carrión García, Sulamita Carvalho, Núria Casacuberta-Barberà, Manuel Castellà, Andrea Castellví, Pedro Castro, Mercedes Catalán-González, Ramon Ciuendez Ávila, Catia Cillóniz, Luisa Clar, Cristina Climent, Jordi Codina, Pamela Conde, Sofía Contreras, María Cruz Martín, Raul de Pablo Sánchez, Diego De Mendoza, Emili Díaz, Yolanda Díaz, María Digna Rivas Vilas, Cristina Dólera Moreno, Irene Dot, Pedro Enríquez Giraudo, Inés Esmoris Arijón, Angel Estella, Teresa Farre Monjo, Javier Fernández, Carlos Ferrando, Albert Figueras, Eva Forcadell-Ferreres, Lorena Forcelledo Espina, Nieves Franco, Enric Franquesa, Àngels Furro, Albert Gabarrus, Cristóbal Galbán, Elena Gallego, Felipe García, Beatriz García, José Luis García Garmendia, Dario Garcia-Gasulla, Emilio García Prieto, Carlos García Redruello, Amaia García Sagastume, José Garnacho-Montero, María Luisa Gascón Castillo, Gemma Gomà, José M. Gómez, Vanesa Gómez Casal, Silvia Gómez, Carmen Gómez Gonzalez, David de Gonzalo-Calvo, Jessica González, Federico Gordo, María Pilar Gracia, Víctor D. Gumucio-Sanguino, Alba Herraiz, Rubén Herrán-Monge, Arturo Huerta, Mercedes Ibarz, Silvia Iglesias, María Teresa Janer, Gabriel Jiménez, Ruth Noemí Jorge García, Mar Juan Díaz, Karsa Kiarostami, Juan I Lazo Álvarez, Miguel León, Alexandre López-Gavín, Ana López Lago, Juan Lopez Messa, Esther López-Ramos, Ana Loza-Vázquez, Desire Macías Guerrero, Nuria Mamolar Herrera, Rafael Mañez Mendiluce, Cecilia L Mantellini, Gregorio Marco Naya, Pilar Marcos, Judith Marin-Corral, Enrique Marmol Peis, Paula Martín Vicente,

María Martínez, Carmen Eulalia Martínez Fernández, Amalia Martínez de la Gándara, María Dolores Martínez Juan, Basilia Martínez Palacios, Ignacio Martínez Varela, Juan Fernando Masa Jimenez, Joan Ramon Masclans, Emilio Maseda, Eva María Menor Fernández, Mar Miralbés, Josman Monclou, Juan Carlos Montejo-González, Neus Montserrat, María Mora Aznar, Dulce Morales, Sara Guadalupe Moreno Cano, David Mosquera Rodríguez, Rosana Muñoz-Bermúdez, Guillermo Muñiz Albaiceta, José María Nicolás, María Teresa Nleto, Ramon Nogue Bou, Rafaela Noguerras Salinas, Marta Ocón, Ana Ortega, Sergio Ossa, Pablo Pagliarani, Francisco Parrilla, Josep-Pedregosa-Díaz, Yhivian Peñasco, Oscar Peñuelas, Leire Pérez Bastida, Purificación Pérez, Felipe Pérez-García, Gloria Pérez Planelles, Eva Pérez Rubio, David Pestaña Laguna, Àngels Piñol-Tena, Javier Prados, Andrés Pujol, Juan Carlos Pozo, Núria Ramon Coll, Gloria Renedo Sanchez-Giron, Jordi Riera, Pilar Ricart, Ferran Roche-Campo, Alejandro Rodríguez, Laura Rodríguez, Felipe Rodríguez de Castro, Silvia Rodríguez, Covadonga Rodríguez Ruiz, Jorge Rubio, Alberto Rubio López, Ángela Leonor Ruiz-García, Miriam Ruiz Miralles, Pablo Ryan Murúa, Eva Saborido Paz, Victor Sagredo, Ana Salazar Degracia, Inmaculada Salvador-Adell, Miguel Sanchez, Ana Sánchez, Angel Sánchez-Miralles, Susana Sancho Chinesta, Bitor Santacoloma, Miguel Sanchez, María Teresa Sariñena, Marta Segura Pensado, Lidia Serra, Mireia Serra-Fortuny, Ainhoa Serrano Lázaro, Lluís Servià, Lorenzo Socias, Laura Soliva, Jordi Solé-Violan, Fernando Suarez Sijmann, Carla Speziale, Luis Tamayo Lomas, Adrián Tormos, María del Carmen de la Torre, Gerard Torres, Mateu Torres, Sandra Trefler, Josep Trenado, Javier Trujillano, Alejandro Úbeda, Luis Urrelo-Cerrón, Estela Val, Manuel Valledor, Luis Valdivia Ruiz, Montserrat Vallverdú, María Van der Hofstadt Martin-Montalvo, Sabela Vara Adrio, Nil Vázquez, Javier Vengoechea, Pablo Vidal, Clara Vilà-Vilardel, Judit Vilanova, Tatiana Villada Warrington, Hua Yang, Minlan Yang, Ana Zapatero.

## Author contributions

AC, CF, LB, MCR, AAB, ECD, SQ, OR, CC, and AA participated in protocol development, study design, study management, statistical analysis, and data interpretation. AC and CF participated in statistical analysis. AC, DGC, AM, LFB, JAL, JR, RF, OP, RM, FB, and AT participated in study design, and study management. CIBERESUCICOVID consortium participated in data collection. All authors read and approved the final manuscript.

## Funding

Financial support was provided by the Instituto Carlos III de Madrid (COV20/00110, ISCIII) and by the Centro de Investigación Biomedica En Red – Enfermedades Respiratorias (CIBERES). AC acknowledges receiving financial support from Instituto de Salud Carlos III (ISCIII; Sara Borrell 2021: CD21/00087). DdGC has received financial support from Instituto de Salud Carlos III (Miguel Servet 2020: CP20/00041), co-funded by the European Union. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and in the decision to submit the manuscript for publication.

## Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

The study was approved by the institution's Internal Review Board (Comité Ético d'Investigació Clínica, registry number HCB/2020/0370), and written informed consent was obtained from either the patients or their relatives.

### Consent for publication

Not applicable.

### Competing interests

Dr. Roca reported a research grant from Hamilton Medical AG and Fisher & Paykel Healthcare Ltd., speaker fees from Hamilton Medical AG, Fisher & Paykel Healthcare Ltd., and Aerogen Ltd., and non-financial research support from Timpe; all outside the submitted work. Minority stakeholder of Tesai Care SL, a spin-off of Parc Taulí University Hospital. The other authors have no conflict of interest to report.

**Author details**

<sup>1</sup>Critical Care Center, Hospital Universitari Parc Taulí, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Department of Medicine, Universitat Autònoma de Barcelona, Plaça Torre de l'Aigua, S/N, 08208 Sabadell, Spain. <sup>2</sup>Centro de Investigación Biomédica en Red en Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain. <sup>3</sup>Intensive Care Unit, Hospital Universitari Sagrat Cor, Grupo Quironsalud, Barcelona, Spain. <sup>4</sup>Heorfy Consulting, Lleida, Spain. <sup>5</sup>Department of Basic Medical Sciences, University of Lleida, Lleida, Spain. <sup>6</sup>Intensive Care and Laboratory of Experimental Intensive Care and Anesthesiology (LEICA), Amsterdam UMC Location AMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands. <sup>7</sup>Translational Research in Respiratory Medicine, Respiratory Department, Hospital Universitari Aran de Vilanova and Santa Maria, IRBLleida, Lleida, Spain. <sup>8</sup>Department of Pneumology, Hospital Clinic of Barcelona, August Pi i Sunyer Biomedical Research Institute—IDIBAPS, University of Barcelona, Barcelona, Spain. <sup>9</sup>Intensive Care Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Institut de Recerca, Barcelona, Spain. <sup>10</sup>Hospital Universitario de Getafe, Universidad Europea, Madrid, Spain. <sup>11</sup>Department of Bioengineering, Universidad Carlos III, Madrid, Spain. <sup>12</sup>Pulmonary Department, University and Polytechnic Hospital La Fe, Valencia, Spain. <sup>13</sup>Intensive Care Clinical Unit, Hospital Universitario Virgen de Rocío, Seville, Spain. <sup>14</sup>Servicio de Medicina Intensiva, Hospital Universitario La Paz, IdiPAZ, Madrid, Spain. <sup>15</sup>Hospital Universitario San Agustín, Asturias, Spain. <sup>16</sup>Hospital Santa María, IRBLleida, Lleida, Spain. <sup>17</sup>Hospital Universitario HM Montepíncipe, Facultad HM Hospitales de Ciencias de La Salud, Universidad Camilo Jose Cela, Madrid, Spain. <sup>18</sup>Servicio de Medicina Intensiva, Hospital Universitario Ramón y Cajal, Madrid, Spain. <sup>19</sup>Intensive Care Unit, and Emergency Medicine, Universidad de Alcalá, Madrid, Spain. <sup>20</sup>Hospital Universitario de Cruces, Barakaldo, Spain. <sup>21</sup>Department of Intensive Care Medicine, Hospital Clínico Universitario Valladolid, Valladolid, Spain. <sup>22</sup>Critical Intensive Medicine Department, Hospital Universitari Arnau de Vilanova de Lleida, IRBLleida, Lleida, Spain. <sup>23</sup>Intensive Care Unit, Hospital Álvaro Cunqueiro, Vigo, Spain. <sup>24</sup>Intensive Care Unit, Hospital Clínico Universitario, Valencia, Spain. <sup>25</sup>Department of Intensive Care Medicine, Hospital Universitario, 12 de Octubre, Madrid, Spain. <sup>26</sup>Hospital Universitario de Móstoles, Madrid, Spain. <sup>27</sup>Department of Critical Care Medicine, CHUS, Complejo Hospitalario Universitario de Santiago, Santiago, Spain. <sup>28</sup>Department of Intensive Care, Hospital Universitari de Bellvitge, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain. <sup>29</sup>Hospital de Mataró de Barcelona, Barcelona, Spain. <sup>30</sup>Department of Medicine, Intensive Care Unit University Hospital of Jerez, University of Cádiz, INIBICA, Cádiz, Spain. <sup>31</sup>Unidad de Cuidados Intensivos, Hospital Universitario San Pedro de Alcántara, Cáceres, Spain. <sup>32</sup>Intensive Care Unit, Hospital San Juan de Dios del Aljarafe, Seville, Spain. <sup>33</sup>Intensive Care Clinical Unit, Hospital Universitario Virgen Macarena, Seville, Spain. <sup>34</sup>Hospital General Universitario Gregorio Marañón, Madrid, Spain. <sup>35</sup>Pulmonary and Critical Care Division, Emergency Department, Clínica Sagrada Família, Barcelona, Spain. <sup>36</sup>Intensive Care Department, Hospital Nuestra Señora de Gracia, Saragossa, Spain. <sup>37</sup>Unidad de Medicina Intensiva, Hospital Universitario Virgen de Valme, Seville, Spain. <sup>38</sup>Critical Care Department, Hospital del Mar-IMIM, Barcelona, Spain. <sup>39</sup>Department of Intensive Medicine, Hospital Universitario Infanta Leonor, Madrid, Spain. <sup>40</sup>Hospital Universitario Torrejón-Universidad Francisco de Vitoria, Madrid, Spain. <sup>41</sup>Critical Care Department, Hospital Universitario Lucus Augusti, Lugo, Spain. <sup>42</sup>Complejo Asistencial Universitario de Palencia, Palencia, Spain. <sup>43</sup>Departamento de Biología Funcional, Instituto Universitario de Oncología del Principado de Asturias, Instituto de Investigación Sanitaria del Principado de Asturias, Hospital Central de Asturias, Universidad de Oviedo, Oviedo, Spain. <sup>44</sup>Hospital General de Segovia, Segovia, Spain. <sup>45</sup>Servei de Medicina Intensiva, Hospital Universitari Son Espases, Palma, Illes Balears, Spain. <sup>46</sup>Servicio de Medicina Intensiva, Hospital Universitario Marqués de Valdecilla, Santander, Spain. <sup>47</sup>UGC-Medicina Intensiva, Hospital Universitario Reina Sofía, Instituto Maimonides IMIBIC, Córdoba, Spain. <sup>48</sup>Servicio de Microbiología Clínica, Facultad de Medicina, Departamento de Biomedicina y Biotecnología, Hospital Universitario Príncipe de Asturias - Universidad de Alcalá, Alcalá de Henares, Madrid, Spain. <sup>49</sup>Centro de Investigación Biomédica en Red en Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, Madrid, Spain. <sup>50</sup>Servei de Medicina Intensiva, Hospital Universitari Germans Trias, Badalona, Spain. <sup>51</sup>Institut d'Investigació Sanitària Pere Virgili (IISPV), Hospital Verge de la Cinta, Tortosa, Tarragona, Spain. <sup>52</sup>Critical Care Department, Hospital Universitario Joan XXIII, CIBERES, Rovira and Virgili University, IISPV, Tarragona, Spain. <sup>53</sup>Hospital Universitario de Salamanca, Salamanca, Spain. <sup>54</sup>Intensive Care Unit, Hospital Universitario Sant Joan d'Alacant, Sant Joan d'Alacant, Alicante,

Spain. <sup>55</sup>Servicio de Medicina Intensiva, Hospital Universitario y Politécnico La Fe, Valencia, Spain. <sup>56</sup>Intensive Care Unit, Hospital Son Llàtzer, Illes Balears, Palma, Spain. <sup>57</sup>Critical Care Department, Hospital Universitario de GC Dr. Negrín, Universidad Fernando Pessoa Canarias, Las Palmas, Gran Canaria, Spain. <sup>58</sup>Intensive Care Unit, Hospital Universitario La Princesa, Madrid, Spain. <sup>59</sup>Critical Care Department, Hospital Universitario Río Hortega de Valladolid, Valladolid, Spain. <sup>60</sup>Servicio de Medicina Intensiva, Hospital Universitario Mútua de Terrassa, Terrassa, Barcelona, Spain. <sup>61</sup>Servicio de Medicina Intensiva, Hospital Punta de Europa, Algeciras, Spain. <sup>62</sup>Hospital Universitario de León, León, Spain. <sup>63</sup>Complejo Hospitalario Universitario de Ourense, Ourense, Spain. <sup>64</sup>Hospital Universitario Río Hortega de Valladolid, Valladolid, Spain. <sup>65</sup>Instituto de Investigación Biomédica de Salamanca (IBSAL), Gerencia Regional de Salud de Castilla y León, Salamanca, Spain. <sup>66</sup>Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA.

Received: 25 January 2024 Accepted: 14 March 2024

Published online: 21 March 2024

**References**

- Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, et al. Acute respiratory distress syndrome. *Nat Rev Disease Prim*. 2019;5(1):1–22.
- Villar J, Ferrando C, Martínez D, Ambrós A, Muñoz T, Soler JA, et al. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir Med*. 2020;8(3):267–76.
- Guérin C, Reigner J, Richard JC, Beuret P, Gacouin A, Boulain T, et al. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med*. 2013;368(23):2159–68.
- Amato MBP, Barbas CSV, Medeiros DM, Magaldi RB, Schettino GP, Lorenzi-Filho G, et al. Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. *N Engl J Med*. 1998;338(6):347–54.
- Acute Respiratory Distress Syndrome Network; Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, et al. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med*. 2000;342(18):1301–8.
- RECOVERY Collaborative Group; Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med*. 2021;384(8):693–704.
- Salama C, Han J, Yau L, Reiss WG, Kramer B, Neidhart JD, et al. Tocilizumab in patients hospitalized with Covid-19 pneumonia. *N Engl J Med*. 2021;384(1):20–30.
- Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the treatment of Covid-19—final report. *N Engl J Med*. 2020;383:NEJMoa2007764.
- Arabi YM, Gordon AC, Derde LPG, Nichol AD, Murthy S, Beidh FA, et al. Lopinavir-ritonavir and hydroxychloroquine for critically ill patients with COVID-19: REMAP-CAP randomized controlled trial. *Intensive Care Med*. 2021;47(8):867–86.
- Ceccato A, Campubí-Rimblas M, Campaña-Duel E, Areny-Balaguero A, Morales-Quinteros L, Artigas A. Anticoagulant treatment in severe ARDS COVID-19 patients. *J Clin Med*. 2022;11(10):2695.
- Perico L, Benigni A, Casiraghi F, Ng LFP, Renia L, Remuzzi G. Immunity, endothelial injury and complement-induced coagulopathy in COVID-19. *Nat Rev Nephrol*. 2021;17(1):46–64.
- Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA. Latent class analysis of ARDS subphenotypes: analysis of data from two randomized controlled trials. *Lancet Respir Med*. 2014;2(8):611–20.
- Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, et al. ARDS subphenotypes and differential response to simvastatin: secondary analysis of a randomized controlled trial. *Lancet Respir Med*. 2018;6(9):691–8.
- Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med*. 2017;195(3):331–8.

15. Sinha P, Calfee CS, Cherian S, Brealey D, Cutler S, King C, et al. Prevalence of phenotypes of acute respiratory distress syndrome in critically ill patients with COVID-19: a prospective observational study. *Lancet Respir Med.* 2020;8(12):1209–18.
16. Bos LDJ, Sjoding M, Sinha P, Bhavani SV, Lyons PG, Bewley AF, et al. Longitudinal respiratory subphenotypes in patients with COVID-19-related acute respiratory distress syndrome: results from three observational cohorts. *Lancet Respir Med.* 2021;9(12):1377–86.
17. Verhoeve PA, Spicer AB, Lopez-Espina C, Bhargava A, Schmalz L, Sims MD, et al. Analysis of protein biomarkers from hospitalized COVID-19 patients reveals severity-specific signatures and two distinct latent profiles with differential responses to corticosteroids\*. *Crit Care Med.* 2023;51(12):1697.
18. López-Martínez C, Martín-Vicente P, Gómez de Oña J, López-Alonso I, Gil-Peña H, Cuesta-Llavona E, et al. Transcriptomic clustering of critically ill COVID-19 patients. *Eur Respir J.* 2023;61(1):220.
19. Torres A, Motos A, Ceccato A, Bermejo-Martin J, de Gonzalo-Calvo D, Pérez R, et al. Methodology of a large multicenter observational study of patients with COVID-19 in Spanish intensive care units. *Arch Bronconeumol.* 2022;58(Suppl 1):22–31.
20. The REDCap consortium: Building an international community of software platform partners - ScienceDirect. 2023. Available from: <https://www.sciencedirect.com/science/article/pii/S1532046419301261>
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–81.
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573–7.
23. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA.* 2012;307(23):2526–33.
24. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707–10.
25. Ankerst M, Breunig MM, Kriegel HP, Sander J. OPTICS: ordering points to identify the clustering structure. *SIGMOD Rec.* 1999;28(2):49–60.
26. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987;20:53–65.
27. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. *Bioinformatics.* 2005;21(15):3201–12.
28. Matthay MA, Arabi YM, Siegel ER, Ware LB, Bos LDJ, Sinha P, et al. Phenotypes and personalized medicine in the acute respiratory distress syndrome. *Intensive Care Med.* 2020;46(12):2136–52.
29. Seymour CW, Gomez H, Chang CCH, Clermont G, Kellum JA, Kennedy J, et al. Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness. *Crit Care.* 2017;21(1):257.
30. Bos LDJ, Ware LB. Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. *Lancet.* 2022;400(10358):1145–56.
31. Wilson JG, Calfee CS. ARDS Subphenotypes: Understanding a Heterogeneous Syndrome. *Crit Care.* 2020;24(1):102.
32. Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for Sepsis. *JAMA.* 2019;321(20):2003–17.
33. Hotchkiss RS, Monneret G, Payen D. Immunosuppression in sepsis: a novel understanding of the disorder and a new therapeutic approach. *Lancet Infect Dis.* 2013;13(3):260–8.
34. Bermejo-Martin JF, Cilloniz C, Mendez R, Almansa R, Gabarrus A, Ceccato A, et al. Lymphopenic Community Acquired Pneumonia (L-CAP), an Immunological phenotype associated with higher risk of mortality. *EBioMedicine.* 2017;24:231–6.
35. Ceccato A, Panagiotarakou M, Ranzani OT, Martín-Fernández M, Almansa-Mora R, Gabarrus A, et al. Lymphocytopenia as a predictor of mortality in patients with ICU-acquired pneumonia. *J Clin Med.* 2019;8(6):843.
36. Delucchi K, Famous KR, Ware LB, Parsons PE, Thompson BT, Calfee CS, et al. Stability of ARDS subphenotypes over time in two randomised controlled trials. *Thorax.* 2018;73(5):439–45.
37. Chen H, Yu Q, Xie J, Liu S, Pan C, Liu L, et al. Longitudinal phenotypes in patients with acute respiratory distress syndrome: a multi-database study. *Crit Care.* 2022;26(1):340.
38. Lu M, Drohan C, Bain W, Shah FA, Bittner M, Evankovich J, et al. Trajectories of host-response subphenotypes in patients with COVID-19 across the spectrum of respiratory support. *CHEST Crit Care.* 2023;1(3):100018.
39. Sinha P, Neyton L, Sarma A, Wu N, Jones C, Zhuo H, et al. Molecular phenotypes of ARDS in the ROSE trial have differential outcomes and gene expression patterns that differ at baseline and longitudinally over time. *Am J Respir Crit Care Med.* 2024. <https://doi.org/10.1164/rccm.202308-1490OC>.
40. Leisman DE, Ronner L, Pinotti R, Taylor MD, Sinha P, Calfee CS, et al. Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes. *Lancet Respir Med.* 2020;8(12):1233–44.
41. Preud'homme G, Duarte K, Dalleau K, Lacomblez C, Bresso E, Smail-Tabbone M, et al. Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark. *Sci Rep.* 2021;11(1):4202.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.