

1 **Identification of Acute Respiratory Distress Syndrome subphenotypes**
2 **denovo using routine clinical data: a retrospective analysis of ARDS**
3 **clinical trials**

4
5 Abhijit Duggal MD^{1‡}, Rachel Kast PhD^{2‡}, Emily Van Ark PhD^{2‡}, Lucas Bulgarelli BSc^{2‡},
6 Matthew T. Siuba DO¹, Jeff Osborn², Diego Rey PhD², Fernando G Zampieri MD PhD³,
7 Alexandre B Cavalcanti MD PhD³, Israel S Maia MD³, Denise M Paisani PhD³, Ligia N
8 Laranjeira³, Ary Serpa Neto MD MSc PhD,^{4,5,6,7,8} Rodrigo Octávio Deliberato MD PhD²

9
10 ‡ Authors contributed equally

11
12 1. Department of Critical Care Medicine, Respiratory Institute, Cleveland Clinic,
13 Cleveland, Ohio, USA.

14 2. Department of Clinical Data Science, Endpoint Health Inc, Palo Alto, California, USA.

15 3. HCor Research Institute, São Paulo, Brazil

16 4. Australian and New Zealand Intensive Care Research Centre (ANZIC-RC), School of
17 Public Health and Preventive Medicine, Monash University, Melbourne, Australia.

18 5. Department of Critical Care, Melbourne Medical School, University of Melbourne,
19 Austin Hospital, Melbourne, Australia.

20 6. Department of Intensive Care, Austin Hospital, Melbourne, Australia.

21 7. Data Analytics Research and Evaluation (DARE) Centre, Austin Hospital, Melbourne,
22 Australia.

23 8. Department of Critical Care Medicine, Hospital Israelita Albert Einstein, São Paulo,
24 Brazil.

25
26 **Correspondence:**

27 Abhijit Duggal MD

28 Address: 9500 Euclid Ave, L2-330, Cleveland, Ohio, 44195

29 E-mail: duggala2@ccf.org

31 **Word count (Abstract):** 222 words

32 **Word count (Text):** 2814 words

33 **Number of figures:** 2 figures

34 **Number of tables:** 3 tables

35 **Supplementary Material:** 01

36 **Key words:** Subphenotype, machine learning, ARDS, critical care, clinical data,
37 clustering

38 **ABSTRACT**

39 **Rationale:** The acute respiratory distress syndrome (ARDS) is a heterogenous
40 condition, and identification of subphenotypes may help in better risk stratification.

41 **Objectives:** Identify ARDS subphenotypes using new simpler methodology and readily
42 available clinical variables.

43 **Design:** Retrospective Cohort Study of ARDS trials.

44 **Setting:** Data from the U.S. ARDSNet trials and from the international ART trial.

45 **Participants:** 3763 patients from ARDSNet datasets and 1010 patients from the ART
46 dataset.

47 **Primary and secondary outcome measures:** The primary outcome was 60-day or 28-
48 day mortality, depending on what was reported in the original trial. K-means cluster
49 analysis was performed to identify subgroups. For feature selection, sets. Sets of
50 candidate variables were tested to assess their ability to produce different probabilities
51 for mortality in each cluster. Clusters were compared to biomarker data, allowing
52 identification of subphenotypes.

53 **Results:** Data from 4,773 patients was analyzed. Two subphenotypes (A and B)
54 resulted in optimal separation in the final model, which included nine routinely collected
55 clinical variables, namely: heart rate, mean arterial pressure, respiratory rate, bilirubin,
56 bicarbonate, creatinine, PaO₂, arterial pH, and FiO₂. Participants in subphenotype B
57 showed increased levels of pro-inflammatory markers, had consistently higher mortality,
58 lower number of ventilator-free days at day 28, and longer duration of ventilation
59 compared to patients in the subphenotype A.

60 **Conclusions:** Routinely available clinical data can successfully identify two distinct
61 subphenotypes in adult ARDS patients. This work may facilitate implementation of
62 precision therapy in ARDS clinical trials.

63

64 **ARTICLE SUMMARY**

65 **Strengths and limitations of this study**

- 66 • Largest cohort of patients used to identify subphenotypes of ARDS patients.
- 67 • Subphenotypes were validated in the population of a large international ARDS
68 randomized controlled trial.
- 69 • Subphenotypes were identified by using only routinely collected clinical data.
- 70 • Our use of data exclusively from randomized controlled trials does not prove
71 generalizability to unselected ARDS populations.
- 72 • The clinical utility of the subphenotypes have to be validated in a prospective study.

73

74 **INTRODUCTION**

75 The Berlin definition of acute respiratory distress syndrome (ARDS) encompasses acute
76 hypoxemic respiratory failure due to a wide variety of etiologies [1]. Due to this inclusion
77 of heterogeneous conditions within the syndrome, there are significant clinical and
78 biological differences that makes ARDS challenging to treat [2,3]. These differences
79 amongst ARDS patients are associated with variation in risk of disease development
80 and progression [3,4], potentially generating differential responses to treatments and
81 interventions [5–10]. In spite of those evidences, clinical risk stratification of ARDS
82 patients still solely depends on $\text{PaO}_2/\text{FiO}_2$ ratios [11,12], possibly misleading the
83 interpretation of results in clinical trials and clinicians when evaluating treatment options
84 for patients [13].

85 Therefore, identifying groups of patients who have similar clinical, physiologic, or
86 biomarker traits becomes relevant [6,14] as it can help with stratification of patients
87 producing better targeted therapies and interventions [15]. These different groups can
88 be defined as ARDS subphenotypes [4,14]. Two ARDS subphenotypes have been
89 consistently identified in previous studies [6–10,16–18]. However, these models are
90 complex, and significant barriers exist in their implementation and use in clinical
91 practice. Existing models use up to 40 predictor variables, including biomarkers and
92 other variables that are not readily available at the bedside [6–10,16–18]. These
93 limitations explain the current status quo of ARDS care, where clinicians must depend
94 on the limited prognostic value of $\text{PaO}_2/\text{FiO}_2$ ratios instead of biologically distinct
95 subphenotypes.

96 We hypothesized that the use of a simpler methodology and a small number of
97 easily available clinical variables could identify new ARDS subphenotypes and thus
98 provide the means to allow future implementation of bedside stratification.

99

100 **METHODS**

101 **Data source and participants**

102 We performed a retrospective study using a de-identified dataset pooling data from six
103 randomized clinical trials in patients with ARDS, namely: ARMA, ALVEOLI, FACTT,
104 EDEN, SAILS, and ART [19–24]. Patients in ARMA, ALVEOLI, FACTT, EDEN and
105 SAILS trials were eligible if they met the American-European consensus for ARDS,
106 including patients with a $\text{PaO}_2 / \text{FiO}_2$ ratio < 300 up to 48 hours before enrollment. From
107 1996 to 2013, these trials enrolled 902, 549, 1000, 1000, and 745 patients, respectively,
108 and tested a variety of interventions [19–23]. Between 2011 and 2017 the international
109 ART study enrolled 1010 adult patients diagnosed with moderate to severe ARDS
110 according to the Berlin definition ($\text{PaO}_2 / \text{FiO}_2$ ratio < 200) for less than 72 hours of
111 duration and assessed two different ventilatory strategies [24]. To avoid biases due to
112 high mortality in the high tidal volume group of the ARMA study [19], which has not
113 been standard of care since the beginning of 2000, only 473 patients receiving low tidal
114 volume in that study were included.

115 **Predictors**

116 Six clinical trials were assessed to identify a set of clinical variables recorded closest to
117 time of randomization which were most commonly available across all datasets. The list
118 of potential candidates was then further refined to include only those that are frequently
119 observed in the routine care of ARDS patients at the time of its diagnosis. In order to
120 develop a clustering algorithm for potential rapid translation into clinical use, elements
121 which would not be commonly found in the electronic health records (EHR) at the time
122 of ARDS diagnosis, such as biomarker levels, ARDS risk factors, organ support apart

123 from mechanical ventilation settings, and severity scores, were excluded from model
124 development. The treatment assignment in the original trials, and clinical outcomes
125 were not considered in the model development.

126 After all assessment, 16 variables that are routinely collected as part of the usual
127 care and which were uniformly present in all the trials were considered, including: age,
128 gender, arterial pH, PaO₂, PaCO₂, bicarbonate, creatinine, bilirubin, platelets, heart rate,
129 respiratory rate, mean arterial pressure, positive end-expiratory pressure (PEEP),
130 plateau pressure, FiO₂, and tidal volume adjusted for predicted body weight (mL/kg
131 PBW). The PBW was calculated as equal to $50 + 0.91$ (centimeters of height – 152.4) in
132 males, and $45.5 + 0.91$ (centimeters of height – 152.4) in females [18]. These variables
133 were grouped into five domains named demographics, arterial blood gases, laboratory
134 values, vital signs, and ventilatory variables. Plateau pressure was excluded due to a
135 high rate of missingness across the trials included in the training set. Amount of missing
136 data in the training datasets is reported in **eTable 1**.

137 **Outcomes**

138 The primary outcome was 60-day mortality for all ARDSnet trials, and 28-day mortality
139 for ART trial. Secondary outcomes included 90-day mortality, number of ventilator free
140 days at day 28 [25], and the duration of mechanical ventilation in survivors within the
141 first 28 days post enrollment.

142 **Data preparation**

143 Data preprocessing was performed before modeling, and the pooled dataset was
144 assessed for completeness and consistency. Patients with values out of the plausible
145 physiological range for a specific variable were excluded from the final analysis

146 (described in **eTable 2**). The training dataset was constructed using data from the two
147 largest ARDSnet trials, EDEN and FACTT. The validation dataset was sourced from the
148 four remaining trials: ALVEOLI, ARMA, SAILS, and ART. Means and standard
149 deviations for z-scoring variables were calculated from the training dataset and
150 subsequently applied to the validation data.

151 **Statistical analysis**

152 Baseline and outcome data were presented according to the assigned cluster.
153 Continuous variables were presented as medians with their interquartile ranges and
154 categorical variables as total number and percentage. Proportions were compared
155 using Fisher exact tests and continuous variables were compared using the Wilcoxon
156 rank-sum test. Study outcomes were further compared using the median and mean
157 absolute differences for continuous and categorical values, respectively.

158 **Model development and validation**

159 For the model development, the K-means clustering algorithm was used. K-means is
160 one of the simplest and most commonly used classes of clustering algorithms. In critical
161 care research, unsupervised machine learning techniques have already been used in
162 several studies, attempting to find homogeneous subgroups within a broad
163 heterogeneous population [26]. This specific algorithm identifies a K number of clusters
164 in a dataset by finding K centroids within the n-dimensional space of clinical features
165 [26].

166 For feature selection, different sets of candidate variables were tested to assess
167 their ability to produce significantly different mortality probabilities in each cluster using
168 the minimum amount of readily available clinical data. For each set of candidate

169 variables, the optimal number of clusters was determined by comparing models with
170 between 2 and 5 clusters, using the Elbow method [27] and the Calinski-Harabasz index
171 [28]. Information about the methods for selecting number of clusters are provided in the
172 supplemental material.

173 The following steps were performed for the final model selection: 1) all predictors
174 were assessed for correlation (**eTable 3**); and 2) ten different combinations of the
175 proposed variables were investigated. These combinations were developed based on
176 the perceived clinical importance of each variable and its combinations. All 10 models
177 were tested for the optimal number of clusters and based on both the Elbow method
178 and the Calinski-Harabasz index, as described above. The models were then
179 compared, aiming for the minimum set of variables with high 60-day mortality
180 separation. The description of each model is show in **eTable 4**.

181 Biological and clinical characteristics of the clusters were evaluated using clinical,
182 laboratory, and (when available) biomarker data to establish subphenotypes [4]. All
183 iterations in model development were done on the training set and the generalizability of
184 the final model was assessed using the validation dataset. K-means clustering analysis
185 is structured to ignore cases with missing data. No assumption was made for
186 missingness and we therefore conducted a complete case analysis. Model development
187 and evaluation was performed using Python version 3.8 and scikit-learn 0.23.1.

188 **Data availability**

189 Data from the ARDSnet studies (EDEN, FACTT, ARMA, ALVEOLI and SAILS) is
190 publicly available from the NHLBI ARDS Network and data from the ART trial can be
191 requested from study authors.

193 **RESULTS**

194 **Participants**

195 Data from 4777 clinical trial patients were considered for inclusion. In total, 4 patients
 196 were excluded for having clinical measurements outside plausible range. The remaining
 197 1998 patients from EDEN and FACTT trials were included in the training set, while the
 198 2775 patients from ARMA, ALVEOLI, SAILS, and ART were included in the validation
 199 cohort.

200 Baseline characteristics of the patients in the training and validation sets are
 201 presented in **Table 1**. Pneumonia was the prevailing etiology followed by sepsis and
 202 aspiration in all trials. Between 29.3% to 72.7% of the patients were receiving
 203 vasopressors at the time of randomization. At randomization, PaO₂ / FiO₂ ratio ranged
 204 from 112 (75 - 158) to 134 (96 - 185) mmHg, and PEEP from 8 (5 - 10) to 12 (10 - 14)
 205 cmH₂O across trials. Mortality at 60 days for the ARDSnet trials ranged from 22.7% to
 206 30.1%, while in the ART trial mortality at 28 days was 58.8%.

207

Table 1 - Baseline Characteristics and Clinical Outcomes in the Included Trials						
	Training set (n = 1998)		Validation set (n = 2775)			
	EDEN (n = 1000)	FACTT (n = 998)	ALVEOLI (n = 549)	ARMA (n = 472)	ART (n = 1010)	SAILS (n = 744)
Age, year	52.0 (42.0 - 63.0)	49.0 (38.0 - 60.8)	50.0 (39.0 - 65.0)	50.0 (37.8 - 65.0)	52.0 (36.0 - 64.0)	55.0 (42.0 - 66.0)
Male gender - no. (%)	510 (51.0)	533 (53.4)	302 (55.0)	285 (60.4)	631 (62.5)	365 (49.0)
Etiology - no. (%)						
Pneumonia	650 (65.0)	471 (47.2)	221 (40.3)	145 (30.7)	555 (55.0)	526 (70.7)
Sepsis	147 (14.7)	231 (23.1)	120 (21.9)	125 (26.5)	196 (19.4)	147 (19.8)
Aspiration	96 (9.6)	149 (14.9)	84 (15.3)	72 (15.3)	58 (5.7)	49 (6.6)
Trauma	36 (3.6)	74 (7.4)	45 (8.2)	59 (12.5)	31 (3.1)	6 (0.8)
Other	71 (7.1)	73 (7.3)	79 (14.4)	71 (15.0)	170 (16.8)	16 (2.2)
Severity of Illness*	73.0 (59.0 - 89.0)	78.0 (62.0 - 94.0)	78.0 (64.0 - 93.0)	83.0 (70.0 - 97.0)	63.0 (50.2 - 75.0)	76.0 (61.0 - 92.0)
Vasopressors - no. (%)	489 (48.9)	397 (40.5)	156 (29.3)	147 (31.3)	734 (72.7)	395 (54.2)
Laboratory tests						

White blood cell count, 10 ⁹ /L	12.0 (7.8 - 16.7)	11.8 (7.2 - 17.1)	11.6 (7.7 - 15.7)	11.5 (7.5 - 16.2)	---	13.9 (8.7 - 20.0)
Platelets, 10 ⁹ /L	169 (108 - 241)	183 (106 - 258)	157 (83 - 247)	135 (80 - 211)	175 (106 - 263)	167 (96 - 247)
Creatinine, mg/dL	1.2 (0.8 - 2.0)	1.0 (0.7 - 1.5)	1.0 (0.7 - 1.7)	1.1 (0.8 - 1.7)	1.3 (0.8 - 2.2)	1.0 (0.7 - 1.7)
Bilirubin, mg/dL	0.8 (0.5 - 1.4)	0.8 (0.5 - 1.6)	0.8 (0.5 - 1.5)	1.0 (0.6 - 2.1)	0.8 (0.4 - 1.5)	0.8 (0.5 - 1.4)
Arterial blood gas						
pH*	7.36 (7.30 - 7.42)	7.37 (7.30 - 7.43)	7.40 (7.34 - 7.44)	7.41 (7.35 - 7.45)	7.28 (7.19 - 7.36)	7.37 (7.31 - 7.42)
PaO ₂ , mmHg	83 (68 - 108)	79 (67 - 100)	77 (67 - 93)	76.5 (67 - 93)	112 (81 - 155)	83 (69 - 103)
PaO ₂ / FiO ₂	125 (86 - 178)	118 (80 - 163)	134 (96 - 185)	112 (75 - 158)	112 (81 - 155)	133 (89 - 178)
PaCO ₂ , mmHg	38 (34 - 45)	39 (34 - 45)	38 (33 - 43)	36 (31 - 41)	50 (42 - 62)	39 (34 - 45)
Bicarbonate, mmol/L	21.0 (18.0 - 25.0)	21.0 (17.4 - 25.0)	22.0 (18.0 - 26.0)	22.0 (18.0 - 25.0)	22.9 (19.4 - 26.3)	22.0 (18.0 - 25.0)
Ventilatory variables						
Tidal volume, mL	410 (360 - 470)	450 (400 - 510)	500 (420 - 600)	700 (600 - 750)	350 (308 - 400)	400 (350 - 460)
Per PBW, mL/kg PBW	6.3 (6.0 - 7.3)	7.1 (6.1 - 8.1)	7.9 (6.6 - 9.4)	10.2 (9.0 - 11.3)	5.9 (5.1 - 6.1)	6.2 (6.0 - 7.1)
Plateau pressure, cmH ₂ O	24.0 (20.0 - 27.0)	26.0 (22.0 - 30.0)	26.0 (22.0 - 31.0)	29.0 (24.8 - 34.0)	26.0 (22.0 - 29.0)	24.0 (19.0 - 28.0)
PEEP, cmH ₂ O	10 (5 - 12)	10 (5 - 12)	10 (5 - 12)	8 (5 - 10)	12 (10 - 14)	10 (5 - 11)
FiO ₂	0.60 (0.50 - 0.80)	0.60 (0.50 - 0.80)	0.60 (0.50 - 0.80)	0.60 (0.50 - 0.74)	0.70 (0.60 - 1.00)	0.60 (0.40 - 0.70)
Clinical outcomes						
28-day mort. - no. (%)	---	---	---	---	594 (58.8)	---
60-day mort. - no. (%)	227 (22.7)	268 (26.9)	144 (26.2)	141 (30.1)	---	199 (26.7)
90-day mort. - no. (%)	233 (23.3)	283 (28.6)	148 (27.5)	143 (30.8)	---	204 (27.4)
Ventilator-free days, day 28	20.0 (0.0 - 24.0)	17.0 (0.0 - 23.0)	18.0 (0.0 - 24.0)	13.0 (0.0 - 23.0)	0.0 (0.0 - 13.0)	20.0 (0.0 - 25.0)
Ventilator days in survivors	7.0 (4.0 - 13.0)	8.0 (5.0 - 16.0)	8.0 (4.0 - 14.0)	8.0 (4.0 - 15.0)	13.0 (8.0 - 20.0)	6.0 (4.0 - 11.0)
Data are median (quartile 25 th - quartile 75 th) or N (%)						
Abbreviations: 28-day mort. is 28-day mortality, 60-day mort. is 60 days mortality, and 90-day mort. is 90-day mortality.						
* Except for ART, that uses SAPS-3, all studies use APACHE-IV						

208

209 Predictor variables and model selection

210 The correlation between the 15 variables selected for clustering is shown in **eTable 3**.

211 The strongest correlation was between PEEP and FiO₂ ($r = 0.49$). The comparison of
 212 the 10 models regarding the optimal number of clusters based on both the Elbow
 213 method and the Calinski-Harabasz index is shown in **eFigure 1**. In all models and
 214 methods, two clusters were a better fit than a higher number of clusters.

215 Across the ten models, absolute mortality difference between cluster 1 and
 216 cluster 2 ranged from 3.9% to 13.1% for the FACTT study and between 0.1% to 8.1%
 217 for EDEN (**Table 2**). The models with the highest 60-day absolute mortality separation
 218 between the clusters for each of the two trials in the training set were then further
 219 evaluated. Models 6, 5, and 8 were consistently amongst the models with highest
 220 separation (**Table 2**). Model 8 was selected for further investigation, as it the fewest
 221 variables (**eTable 4**).

222

FACTT trial (n = 998)			EDEN trial (n = 1000)		
Model	Patients scored	Mortality difference among clusters	Model	Patients scored	Mortality difference among clusters
6	93.5%	13.1%	7	77.7%	8.1%
2	57.4%	12.5%	8	77.7%	8.1%
5	65.5%	12.2%	6	84.1%	6.7%
8	70.2%	11.6%	5	71.7%	6.5%
7	70.2%	11.4%	9	84.7%	6.1%
1	57.4%	11.2%	3	77.7%	4.4%
4	70.2%	10.6%	4	77.7%	4.0%
9	93.5%	10.4%	2	57.7%	3.9%
3	70.2%	10.1%	10	87.3%	2.8%
10	98.8%	3.9%	1	57.7%	0.1%

223

224 **Clinical characteristics of each cluster**

225 Based on model 8, only nine clinical and laboratory variables were needed to identify
 226 the two distinct clusters in ARDS patients, namely: heart rate, mean arterial pressure,
 227 respiratory rate, bilirubin, bicarbonate, creatinine, PaO₂, arterial pH, and FiO₂. For each
 228 variable in the model, opposing measurements could be observed for each cluster
 229 (**Figure 1** and **eFigure 2**). For the ARDSnet trials, the incidence of cluster 1 patients

230 varied from 57.8% (EDEN) to 73.6% (ARMA), and 41.5% of ART patients were part of
 231 cluster 1. Across all trials, patients in cluster 2 had higher severity of illness, rate of
 232 vasopressor, heart rate, respiratory rate, creatinine, and bilirubin, as well as lower
 233 platelets, pH, BUN, and bicarbonate compared to patients in cluster 1 (**eTable 5, 6 and**
 234 **7**). In addition, 28-, 60-, and 90-day mortality rate was higher in patients in cluster 2 in
 235 all trials (**Table 3**). Likewise, for each trial, ventilator-free days at day 28 was lower in
 236 patients in cluster 2 compared to cluster 1, and duration of ventilation in survivors was
 237 longer in cluster 1.

	Cluster 1	Cluster 2	Difference (95% CI)	p value
Training set				
FACTT	<i>n</i> = 407	<i>n</i> = 294		
60-day mortality - no. (%)	94 (23.1)	102 (34.7)	11.6% (4.9% to 18.3%)	0.001
90-day mortality - no. (%)	103 (25.4)	106 (36.3)	10.9% (4.1% to 17.8%)	0.002
Ventilator-free days at day 28	19.0 (0.0 - 24.0)	10.0 (0.0 - 21.0)	-9.0 (-11.9 to -6.1)	< 0.001
Duration of ventilation in survivors, days	8.0 (4.0 - 13.0)	10.0 (7.0 - 19.0)	2.0 (0.5 to 3.5)	< 0.001
EDEN				
	<i>n</i> = 449	<i>n</i> = 328		
60-day mortality - no. (%)	87 (19.4)	90 (27.4)	8.1% (2.1% to 14.0%)	0.010
90-day mortality - no. (%)	90 (20.0)	93 (28.4)	8.3% (2.3% to 14.3%)	0.009
Ventilator-free days at day 28	21.0 (0.0 - 25.0)	15.0 (0.0 - 22.2)	-6.0 (-8.1 to -3.9)	< 0.001
Duration of ventilation in survivors, days	6.0 (4.0 - 11.0)	8.0 (6.0 - 18.0)	2.0 (0.9 to 3.1)	< 0.001
Validation set				
ALVEOLI				
	<i>n</i> = 336	<i>n</i> = 157		
60-day mortality - no. (%)	59 (17.6)	68 (43.3)	25.8% (17.7% to 33.8%)	< 0.001
90-day mortality - no. (%)	60 (18.1)	70 (45.5)	27.3% (19.2% to 35.5%)	< 0.001
Ventilator-free days at day 28	21.0 (4.8 - 25.0)	2.0 (0.0 - 19.0)	-19.0 (-20.8 to -17.2)	< 0.001
Duration of ventilation in survivors, days	7.0 [4.0,13.0]	11.0 (6.0 - 22.2)	4.0 (2.1 to 5.9)	< 0.001
ARMA				
	<i>n</i> = 279	<i>n</i> = 100		
60-day mortality - no. (%)	69 (24.8)	42 (42.0)	17.2% (6.9% to 27.5%)	0.002
90-day mortality - no. (%)	70 (25.5)	42 (42.0)	16.5% (6.0% to 26.9%)	0.003
Ventilator-free days at day 28	17.0 (0.0 - 24.0)	2.0 (0.0 - 19.0)	-15.0 (-18.6 to -11.4)	< 0.001
Duration of ventilation in survivors, days	7.0 (4.0 - 13.8)	11.0 (5.0 -18.0)	4.0 (1.5 to 6.5)	0.018
SAILS				
	<i>n</i> = 319	<i>n</i> = 188		
60-day mortality - no. (%)	80 (25.1)	60 (31.9)	6.8% (-1.2% to 14.9%)	0.119
90-day mortality - no. (%)	81 (25.4)	63 (33.5)	8.1% (0.0% to 16.3%)	0.063
Ventilator-free days at day 28	21.0 (0.0 - 25.0)	16.0 (0.0 - 23.0)	-5.0 (-7.3 to -2.7)	< 0.001

Duration of ventilation in survivors, days	6.0 (3.0 - 10.0)	8.0 (5.0 - 14.0)	2.0 (0.7 to 3.3)	< 0.001
ART	<i>n</i> = 211	<i>n</i> = 298		
28-day mortality - no. (%)	81 (38.4)	180 (60.4)	22.0% (13.4% to 30.7%)	< 0.001
Ventilator-free days at day 28	0.0 (0.0 - 17.0)	0.0 (0.0 - 7.8)	-0.0 (-1.0 to 1.0)	< 0.001
Duration of ventilation in survivors, days	12.0 (8.0 - 20.0)	13.5 (8.0 - 20.0)	2.0 (-0.3 to 4.2)	0.570
Data are median (quartile 25 th - quartile 75 th) or N (%). Difference is mean difference with (95% CI) for binomial variables and median difference with (95% CI) for continuous variables Abbreviations: CI is confidence interval.				

238

239 Identification of Subphenotypes

240 After comparing the clinical characteristics of the clusters, each cluster was assigned to
241 represent a distinct subphenotype of ARDS, with patients in cluster 1 assigned to
242 subphenotype A, and patients in cluster 2 assigned to subphenotype B. Using blood
243 biomarker information available for a subset of patients from both ARMA and ALVEOLI,
244 subphenotype B showed increased levels of pro-inflammatory markers when compared
245 to subphenotype A (**Figure 2** and **eTables 8 and 9**).

246

247 **DISCUSSION**

248 This study successfully demonstrated that nine easily obtainable clinical variables can
249 identify two distinct ARDS subphenotypes with different clinical and biologic
250 characteristics as well as outcomes across the test and validation cohorts. There was
251 good generalizability amongst diverse populations from multiple validation datasets with
252 temporal and geographical differences.

253 It is understandable that researchers feel compelled to use as much information
254 as possible to build robust models. This is supportable for two main reasons: (1) the
255 well-known heterogeneity of complex syndromes such as ARDS and (2) the abundance
256 of highly granular clinical data generated by electronic health records (EHRs). It is
257 anticipated that analyzing this vast amount of data will provide new knowledge
258 regarding disease mechanisms by enabling researchers to find plausible hidden
259 patterns within the data [29]. However, this data-heavy approach has the potential
260 drawback of using predictors which are not generally obtained in a time window prior to
261 intervention, or worse yet, using variables that are not part of the routine standard of
262 care for patients. The rationale of using fewer and easy to collect clinical variables is not
263 new in the field of critical care. Prognostic models have already shown that it is indeed
264 feasible to create meaningful models using fewer predictors [30,31].

265 Our initial choices to define variables commonly found in the EHR at ARDS
266 diagnosis was inspired by a recent report from the World Health Organization (WHO)
267 which showed an enormous discrepancy of medical devices availability in a survey
268 across 135 countries [29]. Recognizing this inconsistency is essential for widespread
269 implementation of machine learning models regardless of varying availability of

270 resources across countries and health systems [29]. The aim is to provide clinically
271 relevant information within a defined and short time period that might impact the delivery
272 of effective interventions to the right patient population and to as many patients as
273 possible [29].

274 Recently, Sinha *et al.* developed supervised-learning gradient boosted classifier
275 models trained using 24 or 14 readily available clinical data elements to reproduce
276 biomarker-derived subphenotypes which were previously identified by Calfee *et al.* [17].
277 Unlike Sinha *et al.*, who predicted previously identified subphenotypes, our study has
278 identified two subphenotypes *de novo* using a small set of clinical variables.

279 Although the subphenotypes that we have identified and those that have been
280 previously published look similar, our work is distinct from previous studies in several
281 ways. We employed different training and validation datasets and also utilized a
282 different and well-established unsupervised learning technique. Moreover, we utilized a
283 process for selecting predictors which is not comparable to previous studies.
284 Acknowledging these differences is crucial. It would not be unexpected to assume that
285 these deviations would be relevant enough to produce different subphenotypes [32].
286 However, the clinical, laboratory characteristics, and the clinical outcomes of our
287 subphenotypes show that they are remarkably similar to subphenotypes found in
288 previous papers, regardless of methodological differences.

289 At this point it is not possible to go beyond this comparative analysis, as there is
290 no gold standard definition of ARDS subphenotypes [32]. Nonetheless, our work does
291 provide robust evidence that ARDS does indeed have two subphenotypes that can be
292 systematically identified, despite major differences in population assessed and

293 methodological approach used compared with previous studies. It also reinforces that
294 we should continue to explore the underlying biological pathways of such
295 subphenotypes to find responders to new or previously tested therapies.

296 Our study has several strengths. First, it is the largest cohort of patients that has
297 been studied to develop distinct subphenotypes of ARDS patients. Moreover, our
298 validation cohort included patients from the ART trial, allowing us to validate our model
299 in the contemporaneous population of a large international randomized clinical trial in
300 addition to the ARDSnet studies used in other subphenotyping studies. Second, our
301 subphenotyping model was developed exclusively on the training set and then validated
302 across multiple separate datasets. Nevertheless, similar separation in mortality was
303 seen between the two subphenotypes across all trials. Third, we used the K-means
304 algorithm to identify our subphenotypes, and the results obtained with this technique
305 can be easily interpreted by clinicians and implemented in clinical practice. Lastly, this is
306 the first phenotyping study that has used easily available clinical variables to identify
307 ARDS phenotypes *de novo*, which allows for early identification of these patients in the
308 clinical care at the bedside. Using this algorithm with a small number of routinely
309 collected variables could enable our model to be applied in trials that either
310 retrospectively or prospectively assess interventions targeted to each subphenotype.
311 Future work should analyze previous trials to identify possible differential treatment
312 response for the subphenotypes of ARDS patients identified in this study.

313 This study also has limitations. First, we have developed our models exclusively
314 on patients enrolled in clinical trials. Due to the strict inclusion and exclusion criteria of
315 these clinical trials, the generalizability of these results needs to be evaluated in

316 unselected ARDS populations. Although there are clear clinical and biomarker
317 differences between the identified subphenotypes, the model's clinical utility needs to be
318 prospectively validated and further investigated. Additionally, our biomarker analysis is
319 limited to those patients in which the data was made publicly available by the study
320 authors. Lastly, K-means clustering does not handle missing data, and no approach
321 was used to impute missing values. However, the extremely low rate of missingness in
322 our study makes this issue less relevant.

323

324 **CONCLUSIONS**

325 This study confirms the existence of two distinct subphenotypes in ARDS patients using
326 a novel clustering model on routinely collected clinical data. This work may allow for
327 easier identification of ARDS subphenotypes to facilitate implementation of precision
328 clinical trial enrollment and development of targeted therapies in a variety of settings
329 without the added burdens of biomarker evaluation.

330

331 **DECLARATIONS**

332 **Funding:** This research received no specific grant from any funding agency in the
333 public, commercial or not-for-profit sectors.

334 **Competing Interest:** AD, MS, FGZ, ABC, ISM, DMP, LNL declare no relevant financial
335 conflicts of interest. RK, EVA, LB, JO, DR and ROD are employees of Endpoint Health,
336 Inc. ASN reported receiving personal fees from Dräger unrelated to the submitted work.

337 **Ethics Approval:** All patients provided informed consent in the original trials. This
338 secondary analysis study was determined to be not human subject research and was
339 this exempt from IRB review by WIRB. WIRB Work Order #1-1228617-1

340 **Availability of data and material:** Data from the ARDSnet studies (EDEN, FACTT,
341 ARMA, ALVEOLI) is publicly available from the NHLBI ARDS Network (NHLBI ARDS
342 Network) and data from the ART trial can be requested from study authors.

343 **Author Contributions:** AD, RK, EVA, LB participated in study design and analysis,
344 drafted, and revised the manuscript, and are the guarantor of the document. MS, DR,
345 JO, FGZ, ABC, ISM, DMP, LNL, and ASN participated in interpretation of data analysis,
346 drafted the manuscript, and revised it for critically important intellectual content. ROD
347 participated in the study design, analysis, interpretation of data analysis, and final
348 revision of the manuscript content.

349 **Twitter:** @msiuba, @f_g_zampieri, @rod_deliberato, @a_serpaneto, @I_bulgarelli,
350 @endpointhealth

351

352 **REFERENCES**

- 353 1 ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, *et al.* Acute respiratory
354 distress syndrome: the Berlin Definition. *JAMA* 2012;**307**:2526–33.
355 doi:10.1001/jama.2012.5669
- 356 2 Thille AW, Esteban A, Fernández-Segoviano P, *et al.* Comparison of the Berlin
357 Definition for Acute Respiratory Distress Syndrome with Autopsy. *Am J Respir Crit*
358 *Care Med* 2013;**187**:761–7. doi:10.1164/rccm.201211-1981OC
- 359 3 Reilly J, Calfee C, Christie J. Acute Respiratory Distress Syndrome Phenotypes.
360 *Semin Respir Crit Care Med* 2019;**40**:019–30. doi:10.1055/s-0039-1684049
- 361 4 Reddy K, Sinha P, O’Kane CM, *et al.* Subphenotypes in critical care: translation into
362 clinical practice. *The Lancet Respiratory Medicine* 2020;**8**:631–43.
363 doi:10.1016/S2213-2600(20)30124-7
- 364 5 Shankar-Hari M, Fan E, Ferguson ND. Acute respiratory distress syndrome (ARDS)
365 phenotyping. *Intensive Care Med* 2019;**45**:516–9. doi:10.1007/s00134-018-5480-6
- 366 6 Calfee CS, Delucchi K, Parsons PE, *et al.* Subphenotypes in acute respiratory
367 distress syndrome: latent class analysis of data from two randomised controlled
368 trials. *The Lancet Respiratory Medicine* 2014;**2**:611–20. doi:10.1016/S2213-
369 2600(14)70097-9
- 370 7 Famous KR, Delucchi K, Ware LB, *et al.* Acute Respiratory Distress Syndrome
371 Subphenotypes Respond Differently to Randomized Fluid Management Strategy.
372 *Am J Respir Crit Care Med* 2017;**195**:331–8. doi:10.1164/rccm.201603-0645OC
- 373 8 Calfee CS, Delucchi KL, Sinha P, *et al.* Acute respiratory distress syndrome
374 subphenotypes and differential response to simvastatin: secondary analysis of a
375 randomised controlled trial. *The Lancet Respiratory Medicine* 2018;**6**:691–8.
376 doi:10.1016/S2213-2600(18)30177-2
- 377 9 for the NHLBI ARDS Network, Sinha P, Delucchi KL, *et al.* Latent class analysis of
378 ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs
379 from sepsis (SAILS) study. *Intensive Care Med* 2018;**44**:1859–69.
380 doi:10.1007/s00134-018-5378-3
- 381 10 Bos LD, Schouten LR, van Vught LA, *et al.* Identification and validation of distinct
382 biological phenotypes in patients with acute respiratory distress syndrome by cluster
383 analysis. *Thorax* 2017;**72**:876–83. doi:10.1136/thoraxjnl-2016-209719
- 384 11 Ferguson ND, Fan E, Camporota L, *et al.* The Berlin definition of ARDS: an
385 expanded rationale, justification, and supplementary material. *Intensive Care Med*
386 2012;**38**:1573–82. doi:10.1007/s00134-012-2682-1

- 387 12 Bellani G, Laffey JG, Pham T, *et al.* Epidemiology, Patterns of Care, and Mortality
388 for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50
389 Countries. *JAMA* 2016;**315**:788. doi:10.1001/jama.2016.0291
- 390 13 Gattinoni L, Vassalli F, Romitti F. Benefits and risks of the P/F approach. *Intensive*
391 *Care Med* 2018;**44**:2245–7. doi:10.1007/s00134-018-5413-4
- 392 14 Matthay MA, Arabi YM, Siegel ER, *et al.* Phenotypes and personalized medicine in
393 the acute respiratory distress syndrome. *Intensive Care Med* 2020;**46**:2136–52.
394 doi:10.1007/s00134-020-06296-9
- 395 15 Shankar-Hari M, Rubenfeld GD. Population enrichment for critical care trials:
396 phenotypes and differential outcomes. *Current Opinion in Critical Care*
397 2019;**25**:489–97. doi:10.1097/MCC.0000000000000641
- 398 16 Sinha P, Delucchi KL, McAuley DF, *et al.* Development and validation of
399 parsimonious algorithms to classify acute respiratory distress syndrome phenotypes:
400 a secondary analysis of randomised controlled trials. *The Lancet Respiratory*
401 *Medicine* 2020;**8**:247–57. doi:10.1016/S2213-2600(19)30369-8
- 402 17 Sinha P, Churpek MM, Calfee CS. Machine Learning Classifier Models Can Identify
403 Acute Respiratory Distress Syndrome Phenotypes Using Readily Available Clinical
404 Data. *Am J Respir Crit Care Med* 2020;**202**:996–1004. doi:10.1164/rccm.202002-
405 0347OC
- 406 18 Kitsios GD, Yang L, Manatakis DV, *et al.* Host-Response Subphenotypes Offer
407 Prognostic Enrichment in Patients With or at Risk for Acute Respiratory Distress
408 Syndrome*: *Critical Care Medicine* 2019;**47**:1724–34.
409 doi:10.1097/CCM.0000000000004018
- 410 19 The National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome
411 (ARDS) Clinical Trials Network. Ventilation with Lower Tidal Volumes as Compared
412 with Traditional Tidal Volumes for Acute Lung Injury and the Acute Respiratory
413 Distress Syndrome. *N Engl J Med* 2000;**342**:1301–8.
414 doi:10.1056/NEJM200005043421801
- 415 20 The National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome
416 (ARDS) Clinical Trials Network. Higher versus Lower Positive End-Expiratory
417 Pressures in Patients with the Acute Respiratory Distress Syndrome. *N Engl J Med*
418 2004;**351**:327–36. doi:10.1056/NEJMoa032193
- 419 21 The National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome
420 (ARDS) Clinical Trials Network. Pulmonary-Artery versus Central Venous Catheter
421 to Guide Treatment of Acute Lung Injury. *N Engl J Med* 2006;**354**:2213–24.
422 doi:10.1056/NEJMoa061895
- 423 22 The National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome
424 (ARDS) Clinical Trials Network, Rice TW, Wheeler AP, *et al.* Initial Trophic vs Full

- 425 Enteral Feeding in Patients With Acute Lung Injury: The EDEN Randomized Trial.
426 *JAMA: The Journal of the American Medical Association* 2012;**307**:795–803.
427 doi:10.1001/jama.2012.137
- 428 23 The National Heart, Lung, and Blood Institute ARDS Clinical Trials Network.
429 Rosuvastatin for Sepsis-Associated Acute Respiratory Distress Syndrome. *N Engl J*
430 *Med* 2014;**370**:2191–200. doi:10.1056/NEJMoa1401520
- 431 24 Writing Group for the Alveolar Recruitment for Acute Respiratory Distress Syndrome
432 Trial (ART) Investigators, Cavalcanti AB, Suzumura ÉA, *et al.* Effect of Lung
433 Recruitment and Titrated Positive End-Expiratory Pressure (PEEP) vs Low PEEP on
434 Mortality in Patients With Acute Respiratory Distress Syndrome: A Randomized
435 Clinical Trial. *JAMA* 2017;**318**:1335. doi:10.1001/jama.2017.14171
- 436 25 Yehya N, Harhay MO, Curley MAQ, *et al.* Reappraisal of Ventilator-Free Days in
437 Critical Care Research. *Am J Respir Crit Care Med* 2019;**200**:828–36.
438 doi:10.1164/rccm.201810-2050CP
- 439 26 Castela Forte J, Perner A, van der Horst ICC. The use of clustering algorithms in
440 critical care research to unravel patient heterogeneity. *Intensive Care Med*
441 2019;**45**:1025–8. doi:10.1007/s00134-019-05631-z
- 442 27 Ketchen DJ, Shook CL. The Application of Cluster Analysis in Strategic
443 Management Research: An Analysis and Critique. *Strategic Management Journal*
444 1996;**17**:441–58. doi:[https://doi.org/10.1002/\(SICI\)1097-](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G)
445 0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G
- 446 28 Caliński T, Harabasz J. A dendrite method for cluster analysis. *Communications in*
447 *Statistics* 1974;**3**:1–27. doi:10.1080/03610927408827101
- 448 29 Bulgarelli L, Deliberato RO, Johnson AEW. Prediction on critically ill patients: The
449 role of “big data.” *Journal of Critical Care* 2020;**60**:64–8.
450 doi:10.1016/j.jcrc.2020.07.017
- 451 30 Johnson AEW, Kramer AA, Clifford GD. A New Severity of Illness Scale Using a
452 Subset of Acute Physiology and Chronic Health Evaluation Data Elements Shows
453 Comparable Predictive Accuracy*: *Critical Care Medicine* 2013;**41**:1711–8.
454 doi:10.1097/CCM.0b013e31828a24fe
- 455 31 Deliberato RO, Escudero GG, Bulgarelli L, *et al.* SEVERITAS: An externally
456 validated mortality prediction for critically ill patients in low and middle-income
457 countries. *International Journal of Medical Informatics* 2019;**131**:103959.
458 doi:10.1016/j.ijmedinf.2019.103959
- 459 32 DeMerle KM, Angus DC, Baillie JK, *et al.* Sepsis Subclasses: A Framework for
460 Development and Interpretation. *Crit Care Med* Published Online First: 15 February
461 2021. doi:10.1097/CCM.0000000000004842

462 **FIGURES LEGENDS**

463 **Figure 1 - Differences of the Variables Included in the Cluster Algorithm Among**
464 **Clusters**

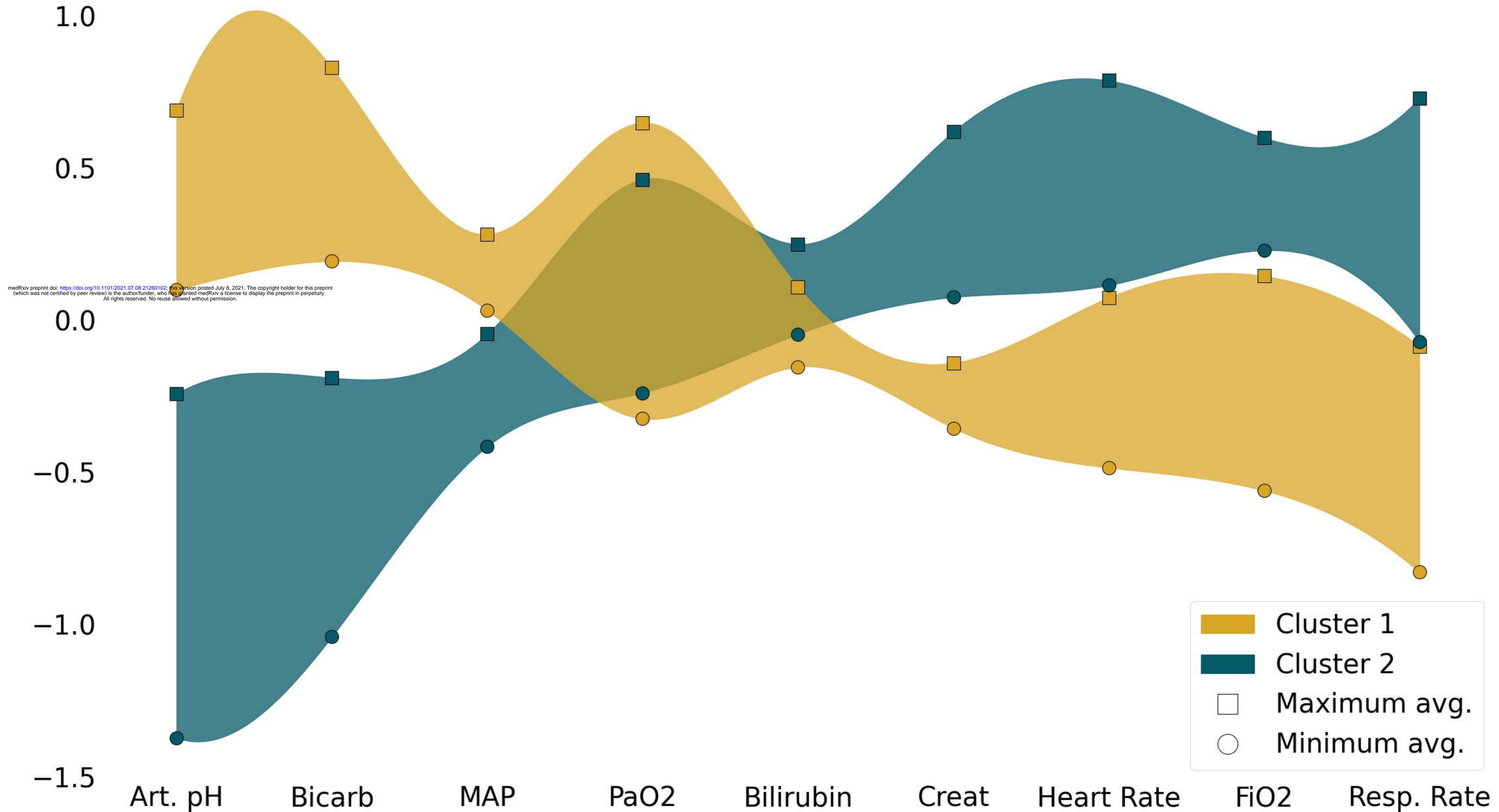
465 Square symbols represent the study with the highest mean z score for each phenotype;
466 Circles represent the study with the lowest mean z score for each phenotype. The
467 colored bands are exclusively to help visualize the opposite trends of the variables on
468 the different clusters; Art.pH: arterial pH; Bicarb: bicarbonate; MAP: mean arterial
469 pressure; Creat: creatinine; Resp.Rate: respiratory rate

470 **Figure 2 - Heat Map of the Biomarkers Available for the ARMA and ALVEOLI Trials**

471 For better visualization and due to difference in scales, the values were log-normalized
472 and z-scored. Subphenotypes A and B are shown separately to highlight their
473 differences.

Range of variable averages between studies

Z-scored average



Biomarkers

