

## ORIGINAL ARTICLE

## Stability of ARDS subphenotypes over time in two randomised controlled trials

Kevin Delucchi,<sup>1</sup> Katie R Famous,<sup>2</sup> Lorraine B Ware,<sup>3</sup> Polly E Parsons,<sup>4</sup>  
B Taylor Thompson,<sup>5</sup> Carolyn S Calfee,<sup>6</sup> for the ARDS Network

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2017-211090>).

<sup>1</sup>Department of Psychiatry, University of California, San Francisco, California, USA

<sup>2</sup>Critical Care Medicine, Kaiser Permanente Oakland Medical Center, Oakland, California, USA

<sup>3</sup>Departments of Medicine and Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, Tennessee, USA

<sup>4</sup>Department of Medicine, University of Vermont School of Medicine, Burlington, Vermont, USA

<sup>5</sup>Department of Medicine, Division of Pulmonary and Critical Care, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>6</sup>Departments of Medicine and Anesthesia, Division of Pulmonary and Critical Care Medicine, University of California, San Francisco, California, USA

## Correspondence to

Dr Carolyn S Calfee, Departments of Medicine and Anesthesia, Division of Pulmonary and Critical Care Medicine, University of California San Francisco, CA 94143-0111, USA; [Carolyn.Calfee@ucsf.edu](mailto:Carolyn.Calfee@ucsf.edu)

Received 26 September 2017

Revised 18 January 2018

Accepted 5 February 2018

Published Online First

24 February 2018



**To cite:** Delucchi K, Famous KR, Ware LB, et al. *Thorax* 2018;**73**:439–445.

## ABSTRACT

**Rationale** Two distinct acute respiratory distress syndrome (ARDS) subphenotypes have been identified using data obtained at time of enrolment in clinical trials; it remains unknown if these subphenotypes are durable over time.

**Objective** To determine the stability of ARDS subphenotypes over time.

**Methods** Secondary analysis of data from two randomised controlled trials in ARDS, the ARMA trial of lung protective ventilation (n=473; patients randomised to low tidal volumes only) and the ALVEOLI trial of low versus high positive end-expiratory pressure (n=549). Latent class analysis (LCA) and latent transition analysis (LTA) were applied to data from day 0 and day 3, independent of clinical outcomes.

**Measurements and main results** In ALVEOLI, LCA indicated strong evidence of two ARDS latent classes at days 0 and 3; in ARMA, evidence of two classes was stronger at day 0 than at day 3. The clinical and biological features of these two classes were similar to those in our prior work and were largely stable over time, though class 2 demonstrated evidence of progressive organ failures by day 3, compared with class 1. In both LCA and LTA models, the majority of patients (>94%) stayed in the same class from day 0 to day 3. Clinical outcomes were statistically significantly worse in class 2 than class 1 and were more strongly associated with day 3 class assignment.

**Conclusions** ARDS subphenotypes are largely stable over the first 3 days of enrolment in two ARDS Network trials, suggesting that subphenotype identification may be feasible in the context of clinical trials.

## INTRODUCTION

As the critical care community continues to try to understand why so many clinical trials for sepsis and the acute respiratory distress syndrome (ARDS) fail to show benefit, there has been increased attention to the clinical and biological heterogeneity of these syndromes as a potential culprit.<sup>1</sup> We recently reported that latent class analysis (LCA) identifies two clinically and biologically distinct subphenotypes in independent analyses of three large ARDS cohorts, all from the National Heart Lung and Blood Institutes's (NHLBI) ARDS Network: the ARMA trial of low tidal volume ventilation, the ALVEOLI trial of low versus high positive end-expiratory pressure (PEEP) and the Fluid and Catheter Treatment Trial of liberal versus conservative fluid management.<sup>2,3</sup> Importantly, the two

## Key messages

## What is the key question?

► Two distinct acute respiratory distress syndrome (ARDS) subphenotypes have been identified using data obtained at time of enrolment in clinical trials, with differential response to randomly assigned positive end-expiratory pressure and fluid management strategy; however, it remains unknown if these subphenotypes are durable over time.

## What is the bottom line?

► ARDS subphenotypes are largely stable over the first 3 days of enrolment in two ARDS Network trials.

## Why read on?

► This finding suggests that subphenotype identification may be feasible in the context of clinical trials and also supports the hypothesis that there are fundamental biological and clinical differences between the subphenotypes that are not dictated by timing of measurement.

subphenotypes had clear differences in biomarker profiles and differential responses to randomly assigned PEEP and fluid management strategies, suggesting that these subphenotypes may in fact be endotypes of ARDS.

While these data provide strong support for the presence of two subphenotypes of ARDS early in its course,<sup>4</sup> the stability of these subphenotypes over time is an important question that has not yet been studied. Some have queried whether the two subphenotypes may represent different temporal stages in ARDS evolution and therefore would not be expected to be stable over time. After time in the hospital and treatment of critical illness, subphenotypes may no longer exist; alternatively, more or fewer subphenotypes may be found, or patients may transition between subphenotypes over time. The durability of subphenotypes over time is critically important for understanding the underlying pathogenesis of each subgroup and for determining whether these subphenotypes can realistically be targeted in clinical trials, since the subphenotypes must be at least reasonably stable over several days if they are to be used to guide trial enrolment.<sup>5</sup> Therefore, we set out to study the durability of ARDS subphenotypes over time in the ARMA and ALVEOLI cohorts by addressing four key questions.

**Table 1** Latent class analysis of subjects enrolled in ARMA at day 3

N Classes	BIC	Entropy	Number assigned to each subphenotype				P value*
			Class 1, n (%)	Class 2, n (%)	Class 3, n (%)	Class 4, n (%)	
1	29865		458				
2	29303	0.83	321 (70)	137 (30)			0.14
3	28557	0.90	319 (70)	133 (29)	6 (1)		0.24
4	28452	0.92	301 (66)	84 (18)	67 (15)	6 (1)	0.80

\*P value represents the Vuong-Lo-Mendell-Rubin (VLMR) test, which tests if  $k$  classes confer an improved model fit over  $k-1$  classes.

Entropy, an index ranging between 0 and 1, which indicates separation of classes.

BIC, Bayesian information criteria, a measure of model fit.

First, is there evidence of the existence of distinct ARDS subphenotypes 3 days after study entry? Second, if subphenotypes are found 3 days after study entry, are they similar to the subphenotypes observed at day 0? Third, do subjects move between subphenotypes from day 0 to day 3? Finally, is stability of subphenotype membership related to clinical characteristics of the patients and/or clinical outcomes? We hypothesised that we would identify two subphenotypes at day 3 as in our prior studies and that these subphenotypes would be relatively stable over the 3-day period studied.

## METHODS

### Subjects

We used previously measured clinical and biomarker data from subjects enrolled in two randomised controlled trials of patients with ARDS. The full results of these studies have been previously published.<sup>6,7</sup> Briefly, ARMA enrolled 902 patients from 1996 to 1999. One arm of the study found that a lower tidal volume ventilatory strategy resulted in lower mortality.<sup>6</sup> Therefore, as in our previous study, subjects randomised to the higher tidal volume strategy in ARMA were excluded from the current analysis.<sup>2</sup> The ALVEOLI trial enrolled 549 subjects from 1999 to 2002 and found no mortality difference between a low versus high PEEP ventilatory strategy.<sup>7</sup> In both ARMA and ALVEOLI, subjects were included if they met ARDS diagnostic criteria within 36 hours prior to enrolment.

For the current analyses, we used clinical and biomarker data obtained on days 0 and 3 of study enrolment. Day 3 was selected because biospecimens were obtained at this time point in both clinical trials with minimal attrition. We included all subjects with clinical data who were alive at each time point. Study participants decreased by 3% (ARMA) and 4% (ALVEOLI) from day 0 to day 3, which was largely dropout due to death (online supplementary table S1). During the ARMA trial, some patients were randomised into one of two substudies.<sup>8,9</sup> In one substudy, some patients received ketoconazole or placebo; in a second substudy, some patients received lisofylline or placebo.

### Variable selection

We used as many of the same variables for these analyses as possible compared with our prior published analysis.<sup>2,3</sup> For the current analyses, however, we only included variables that were available on both study days 0 and 3. We did not include PEEP, because it was the randomised treatment in ALVEOLI. A full list of the variables used is available in the online supplement. Biomarkers for these analyses were previously measured for other studies.<sup>10–15</sup> Skewed variables were log-transformed to achieve a distribution closer to normal. As scales of measurement varied widely between variables, we standardised non-categorical variables to a mean of 0 and a SD of 1, for each study and day, as in our prior work.

### Latent class analysis

Latent class models, a subset of finite mixture modelling, were separately fit to the data from each trial. Modelling was conducted separately at day 0 and day 3. As the effects of ketoconazole and lisofylline on the subphenotypes are unknown, we also re-estimated the day 3 LCA models for ARMA without those participants who received active study drugs (either ketoconazole or lisofylline; total  $n=115$  at day 3). Models with from 1 to 4 classes were fit. For each model, each subject is assigned a probability of belonging to each class. Ideally, the probability will be near 1.0 for a single class and close to zero for the others. In order to determine the optimal number of latent classes at each day, we considered multiple factors, including (1) the Bayesian information criterion (BIC), in which a decreasing number indicates an improved model fit, (2) the number of subjects assigned to each class, in which a small number would be unlikely to represent a clinically significant subgroup and (3) the Vuong-Lo-Mendell-Rubin (VLMR) test, which tests if  $k$  classes fits the data better than  $k-1$  classes.<sup>16</sup> We also examined entropy, in which values  $\geq 0.80$  represent good class separation.<sup>17</sup> As in our prior studies, clinical outcomes were not considered in the latent class modelling. All latent class and latent transition modelling was conducted in Mplus (V7.4).

**Table 2** Latent class analysis of subjects enrolled in ARMA at day 3 and not assigned to active drug in the ketoconazole or lisofylline substudies

N Classes	BIC	Entropy	Number assigned to each subphenotype				P value*
			Class 1, n (%)	Class 2, n (%)	Class 3, n (%)	Class 4, n (%)	
1	22106		343				
2	21552	0.87	247 (72)	96 (28)			<0.0001
3	21516	0.91	240 (70)	62 (18)	41 (12)		0.33
4	21523	0.84	162 (47)	92 (27)	49 (14)	40 (12)	0.23

\*P value represents the Vuong-Lo-Mendell-Rubin (VLMR) test, which tests if  $k$  classes confer an improved model fit over  $k-1$  classes.

Entropy, an index ranging between 0 and 1, which indicates separation of classes.

BIC, Bayesian information criteria, a measure of model fit.

**Table 3** Latent class analysis of subjects enrolled in ALVEOLI at day 3

N Classes	BIC	Entropy	Number assigned to each subphenotype				P value*
			Class 1, n (%)	Class 2, n (%)	Class 3, n (%)	Class 4, n (%)	
1	34 956		525				
2	34 380	0.85	394 (75)	131 (25)			<0.0001
3	34 234	0.79	299 (57)	113 (21.5)	113 (21.5)		0.39
4	34 140	0.84	266 (51)	102 (19)	79 (15)	78 (15)	0.38

\*P value represents the Vuong-Lo-Mendell-Rubin (VLMR) test, which tests if *k* classes confer an improved model fit over *k*-1 classes.

Entropy, an index ranging between 0 and 1, which indicates separation of classes.

BIC, Bayesian information criteria, a measure of model fit.

### Latent transition model

To determine if subphenotype assignment was stable over the first 3 days of study enrolment, we estimated and tested a two-class latent transition model.<sup>18</sup> This model simultaneously estimates the latent class model at each of the two time points and the relationship of the latent classes between these time points. This model also provides an estimate of latent class membership on each day and the probability of changing class. For this analysis, we incorporated all variables that were included in the initial latent class models.

### Comparisons of clinical outcomes

Based on the latent transition analysis results, we compared subjects who stayed in the same latent class over time to those for whom class assignment changed over time. We used Pearson's  $\chi^2$  test to compare the proportion of patients alive at day 90. The Kruskal-Wallis non-parametric test was used to compare ventilator free days and organ-failure free days. These statistical analyses were conducted in SAS V.9.4.

Some of these data have previously been reported in the form of an abstract.<sup>19</sup>

## RESULTS

### Day 0 subphenotypes

We began the analysis by repeating our prior latent class modelling at day 0 in both cohorts but using only the variables that were available at both day 0 and day 3. In both ARMA and ALVEOLI at day 0, as the number of classes increased, the BIC decreased, indicating improved model fit (online supplementary tables S2 and S3). Entropy was >0.80, indicating good separation between the classes in both cohorts. In ARMA, the VLMR test demonstrated that a two-class model provided an improved model fit over a one-class model ( $P=0.01$ ). A three-class model did not provide a statistically significant improvement in model fit ( $P=0.61$ ). In ALVEOLI, the VLMR test demonstrated an improved fit with a two-class model over a one-class model ( $P=0.04$ ); a three-class model did not result in improved model fit over a two-class model ( $P=0.33$ ). These results were similar to our previous analysis, which was expected given the similarities in the variables considered.<sup>2</sup> We retained a two-class model for both ARMA and ALVEOLI based on these results.

### Day 3 subphenotypes in ARMA

We next carried out an LCA in the ARMA cohort at day 3 ( $n=458$ ). As the number of classes was increased, the BIC decreased, indicating improved model fit with additional classes (tables 1 and 2). Entropy values >0.8 indicated good separation between the classes. The VLMR test did not indicate a statistically significant improvement in model fit with a two-class model over a one-class model ( $P=0.14$ ). The two-class model,

however, had a similar proportion of subjects assigned to each class as in our prior work. The three-class model had only six patients in the third class, which seemed unlikely to represent a clinically significant subgroup.

In the models re-estimated without subjects who received an active drug in the ARMA trial, we found that a two-class model best fit the data, based on the entropy, the BIC and VLMR P value ( $P<0.0001$ ; table 2). Additional analysis revealed that ketoconazole and lisofylline had differential effects on key class-defining biomarkers at day 3, including interleukin (IL)-6 for ketoconazole and protein C and soluble tumor necrosis factor receptor-1 (sTNFr-1) for lisofylline (online supplementary tables S4 and S5), suggesting that active drug administration may interfere with class identification at this time point. In light of these results both with and without the ketoconazole and lisofylline patients, we proceeded with a two-class model for subsequent analyses of ARMA at day 3.<sup>20 21</sup>

### Day 3 subphenotypes in ALVEOLI

In latent class models of the ALVEOLI cohort on day 3 ( $n=525$ ), the BIC decreased as the number of classes increased, and entropy was greater than 0.80 (table 3). Based on the VLMR test, a two-class model was a statistically significant improvement over a one-class model ( $P<0.0001$ ). The addition of a third class did not provide an improvement over a two-class model ( $P=0.39$ ). Based on these results, we also used a two-class model to describe the ALVEOLI cohort on day 3 of study enrolment.

### Defining characteristics of days 0 and 3 subphenotypes

In ARMA, comparison of the continuous variables in the two classes at day 0 and day 3 indicated that many of the same variables contributed to distinguishing the two classes at both time points (figure 1; online supplementary figure S1A). Specifically, of the top 15 measures with the greatest absolute value difference (using standardised variables) between the classes at day 0, 13 were also in the top 15 at day 3. Likewise, in the ALVEOLI cohort, 13 of the top 15 measures with the greatest absolute value difference between the classes at day 0 were the same at day 3 (figure 2; online supplementary figure S1B). As in our prior work, plasma biomarkers of inflammation tended to contribute heavily to class identification at both time points and suggested that one class is relatively 'hyper-inflammatory' at both time points compared with the other class (figures 1 and 2).

We next compared the results between the ARMA and ALVEOLI cohorts. On day 3, 9 of the 10 continuous measures with the greatest absolute value difference between the classes were similar in ARMA and ALVEOLI: IL-8, IL-6, bilirubin, intercellular adhesion molecule-1, tumour necrosis factor receptor-1, mean airway pressure, protein C, creatinine and respiratory rate. (figures 1 and 2) Likewise, the variables that did not exhibit

differences between the classes were similar. The raw values for selected clinical and biological variables at day 0 and day 3 in each class and each cohort are shown in online supplementary tables S6 and S7.

In comparing the categorical variables, there was no statistically significant difference in gender between the two classes on day 3 in either cohort (table 4). There were more white patients in the 'hypo-inflammatory' class 1 at day 3 for both studies, but the difference was statistically significant only in ALVEOLI ( $P=0.03$ ). There were significantly more patients on vasopressors at day 3 in the 'hyper-inflammatory' class 2 in both cohorts ( $P<0.0001$  for both). There was a statistically significant difference in primary ARDS risk factor between the two classes in both cohorts as well, with a higher proportion of patients with sepsis in the 'hyper-inflammatory' class 2 compared with class 1 ( $P<0.01$  for both).

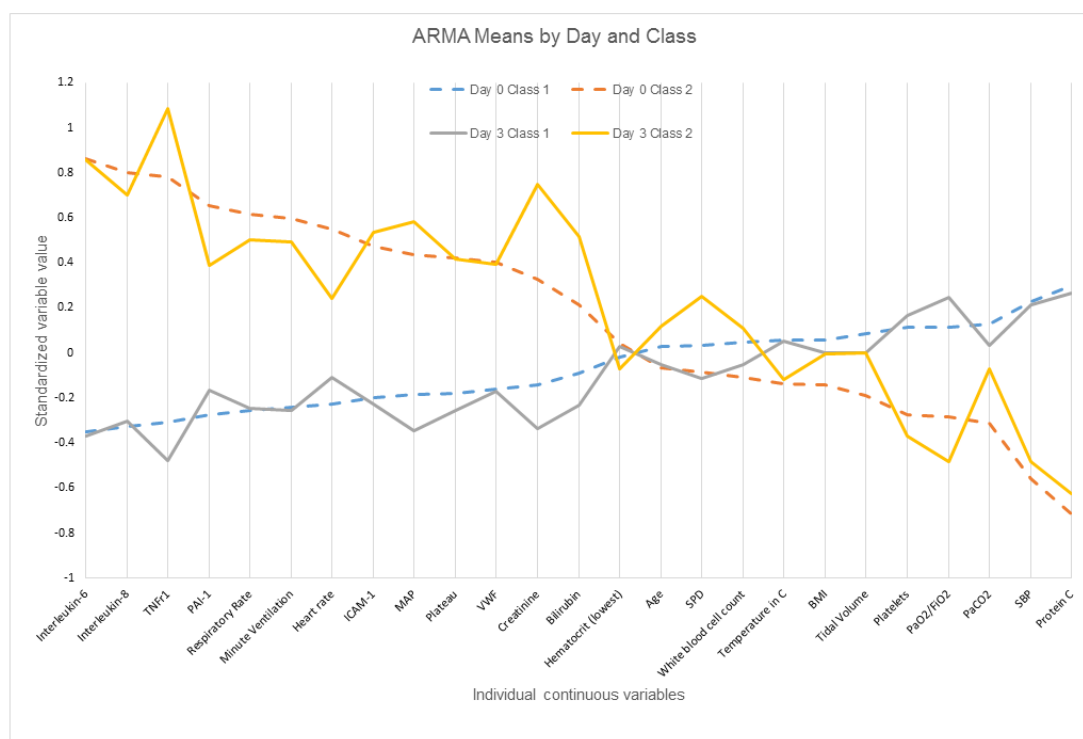
### Stability of ARDS subphenotypes from day 0 to day 3

In the latent transition model, estimates of class membership at each time point were very similar to those obtained from the latent class models. In ARMA, the sizes for classes 1 and 2 in the latent transition model were 330 and 128, respectively, on Day 0, and 321 and 137, respectively, on Day 3. In ALVEOLI, the sizes were 376 and 149, respectively, on Day 0, and 369 and 156, respectively, on day 3. Entropy was 0.89 in both cohorts.

As seen in table 5 and online supplementary figure S2, in both cohorts, most patients assigned to a class at day 0 remained assigned to the same class at day 3 ( $>94\%$ ). While the probabilities of switching classes were low (all at or under 11%), there was a slightly higher probability of moving from the hyperinflammatory class (class 2) to the hypoinflammatory class (class 1) in both cohorts. Similar results were found when patients were assigned to their most likely class using the original latent class models (data not shown). The probability of class assignment generated in the original latent class models was not associated with the likelihood of moving between classes over time (online supplementary table S8).

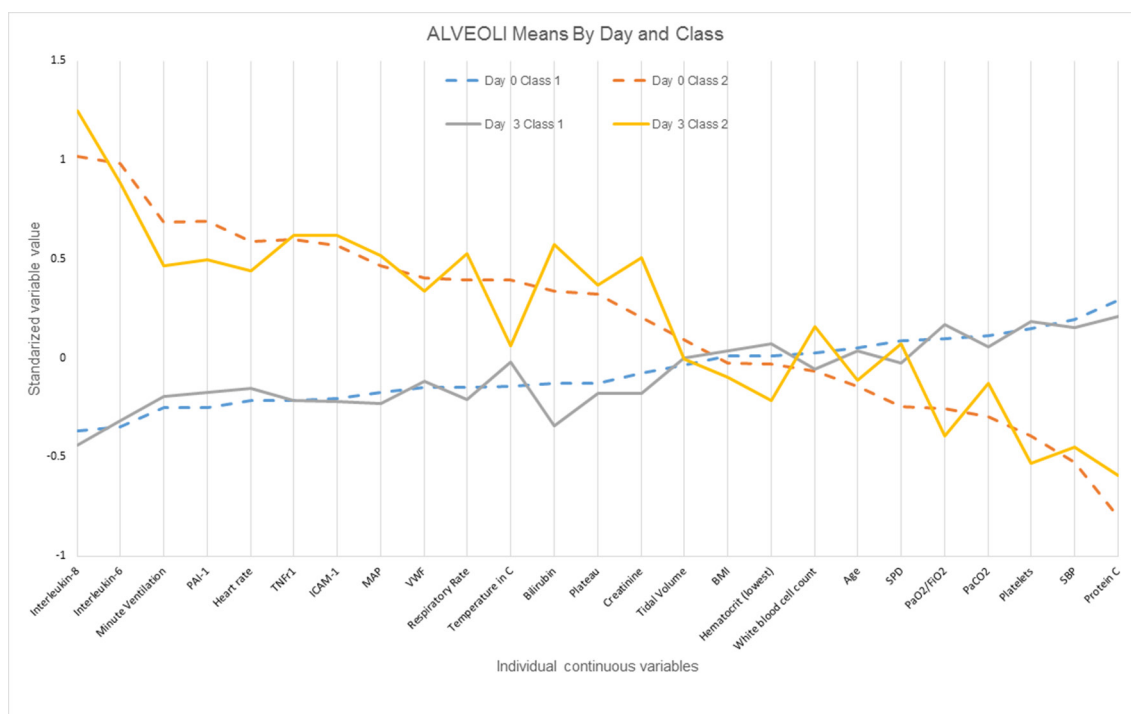
### Transition status is related to clinical outcomes

In the latent transition model for ARMA, 25 patients changed latent classes between day 0 and day 3, and 29 changed in ALVEOLI. We compared the outcomes among four groups: (1) those who changed from class 1 to class 2; (2) those who changed from class 2 to class 1; (3) those who stayed in class 1; and (4) those who stayed in class 2. In both cohorts, 90-day mortality was lowest among those who were in the hypoinflammatory class 1 at both day 0 and day 3 (18% ARMA, 16% ALVEOLI) (table 6). For those remaining in the hyperinflammatory class 2 on both days, the mortality rates were 48% in ARMA and 44% in ALVEOLI. Subjects who moved from the hypoinflammatory class 1 to the hyperinflammatory class 2 in



**Figure 1** Differences in standardised values of each continuous variable by class in the ARMA cohort by day of assessment and latent class. The variables are sorted on the basis of the degree of separation between the classes at day 0, from maximum positive separation at day 0 on the left (ie, class 2 higher than class 1) to maximum negative separation at day 0 on the right (ie, class 2 lower than class 1). The y-axis represents standardised variable values, in which all means are scaled to 0 and SDs to 1. A value of +1 for the standardised variable signifies that the mean value for a given phenotype was one SD higher than the mean value in the cohort as a whole. Mean values are joined by lines to facilitate displaying class profiles. BMI, body mass index; ICAM-1, intercellular adhesion molecule-1; MAP, mean airway pressure; PaCO<sub>2</sub>, pressure of carbon dioxide in arterial blood; PAI, plasminogen activator inhibitor-1; Plateau, plateau pressure; SBP, systolic blood pressure; SPD, surfactant protein D; TNF1, tumour necrosis factor receptor-1; VWF, vonWillebrand factor.





**Figure 2** Differences in standardised values of each continuous variable by class in the ALVEOLI cohort by day of assessment and latent class. The variables are sorted on the basis of the degree of separation between the classes at day 0, from maximum positive separation at day 0 on the left (ie, class 2 higher than class 1) to maximum negative separation at day 0 on the right (ie, class 2 lower than class 1). The y-axis represents standardised variable values, in which all means are scaled to 0 and SDs to 1. A value of +1 for the standardised variable signifies that the mean value for a given phenotype was one SD higher than the mean value in the cohort as a whole. Mean values are joined by lines to facilitate displaying class profiles. BMI, body mass index; HCTL, hematocrit (low); ICAM-1, intercellular adhesion molecule-1; MAP, mean airway pressure; PaCO<sub>2</sub>, pressure of carbon dioxide in arterial blood; PAI, plasminogen activator inhibitor-1; Plateau, plateau pressure; SBP, systolic blood pressure; SPD, surfactant protein D; TNFr1, tumour necrosis factor receptor-1; VWF, vonWillebrand factor.

the ARMA cohort had a high mortality in ARMA ( $n=13/17$ , 77%), though this was not observed in ALVEOLI ( $n=7/18$ , 39%). For those moving from the hyperinflammatory class 2 into the hypoinflammatory class 1, mortality was 25% in ARMA and 9% in ALVEOLI. These differences were statistically significant in both studies (ARMA:  $P<0.0001$ ; ALVEOLI:  $P<0.0001$ .) For ventilator-free days, in both studies, the fewest days were observed for those who were in the hyperinflammatory class 2 on day 3, regardless of initial class assignment (table 6). A similar pattern was observed for organ-failure free days. No statistically significant differences between those who changed class and

those who did not were found for age, race, gender, ARDS risk and, for ALVEOLI patients, PEEP assignment.

## DISCUSSION

In this analysis of the stability of ARDS subphenotypes over the first several days of enrolment in two separate ARDS clinical trials, we found strong evidence supporting the durability of two distinct subphenotypes of ARDS over time. While the evidence in support of a two-class model on day 3 was stronger in the ALVEOLI cohort than in the ARMA cohort, the two classes at

**Table 4** Distribution of categorical variables by class for ARMA and ALVEOLI on day 3

Measure	ARMA			ALVEOLI		
	Class 1 (n=321)	Class 2 (n=137)	P value*	Class 1 (n=394)	Class 2 (n=131)	P value*
Male	197 (61)	77 (56)	0.30	216 (55)	70 (53)	0.78
White race	247 (77)	100 (73)	0.37	303 (77)	88 (67)	0.03
Vasopressors	33 (10)	70 (51)	<0.0001	13 (3)	33 (25)	<0.0001
ARDS risk factor			<0.01			<0.01
Trauma	52 (16)	7 (5)		39 (10)	5 (4)	
Sepsis	73 (23)	46 (34)		69 (18)	43 (35)	
Aspiration	49 (15)	21 (15)		65 (17)	15 (12)	
Pneumonia	91 (28)	48 (35)		167 (44)	47 (38)	
Other	55 (17)	15 (11)		36 (10)	13 (11)	

Values represent n (%). Sample size and therefore percentages may vary for each row due to missing data.

\*P value is from Pearson's  $\chi^2$  test of proportions.

ARDS, acute respiratory distress syndrome.

**Table 5** Transitions between classes from day 0 to day 3 in ARMA (panel A) and ALVEOLI (panel B) for latent transition model

(A) ARMA			
		Day 3 (n=458)	
		Class 1	Class 2
Day 0	Class 1	0.93 (313)	0.07 (17)
	Class 2	0.09 (8)	0.91 (120)
(B) ALVEOLI			
		Day 3 (n=525)	
		Class 1	Class 2
Day 0	Class 1	0.95 (358)	0.05 (18)
	Class 2	0.11 (11)	0.89 (138)

Tables represent transition probabilities for each class at day 3 compared with day 0; numbers in parentheses are numbers of patients in each group.

day 3 in both cohorts had distinct clinical and biological phenotypes, similar in nature and distribution to those we previously reported at day 0 and that appeared largely stable over time. In both cohorts, clinical outcomes seemed to be most strongly related to class assignment at day 3, though the number of subjects moving between the classes over time was small.

These findings have important implications for future clinical trials of targeted therapies in ARDS.<sup>22</sup> Specifically, if subphenotypes are to be used to help target clinical trials and/or future therapeutics,<sup>4,5</sup> it is critical to understand their stability over time. These findings also support the hypothesis that there are fundamental biological and clinical differences between the subphenotypes that are not dictated by timing of measurement.

While the evidence for two classes in ARMA at day 3 is less strong than in ALVEOLI, we proceeded with a two-class model in this cohort for several reasons. First, the BIC and entropy values indicated respectively that a two-class model was a better fit for the data than a one class model and that the classes in the two-class model were well separated. Second, when we removed patients who received ketoconazole or lisofylline from the analysis (both of which impacted the biological markers used to identify subphenotype), the evidence for a two-class model was just as strong as it was in ALVEOLI. Third, the size of the two classes at day 3 in ARMA was quite similar to our previous findings from day 0. Fourth, the evidence for a two-class model in ARMA at day 0 in both our previous study<sup>2</sup> and the current analyses is quite strong; thus, it is of considerable interest to determine how these two classes evolve over time. Fifth, the similarity in the clinical and biological phenotypes of the two

classes at day 0 and day 3 in ARMA, and also between day 3 in ARMA and day 3 in ALVEOLI, supports a two-class model in ARMA at day 3. Finally, the findings of the latent transition analysis showing stability of class assignment from day 0 to 3 strongly support the presence of two classes at day 3 in ARMA. However, this finding should be interpreted with more caution than our findings in ALVEOLI. A smaller sample size at day 3, the lack of several key variables that distinguished the two classes in our prior analyses (most notably bicarbonate) and additional variance in the data due to substudies of ketoconazole and lisofylline may have contributed to a higher VLMR P value at day 3 in ARMA.

We analysed subphenotype membership over time in two different and complementary ways: first, using latent transition analysis, and second, by examining most likely class assignment in independent analyses of day 0 and day 3. The findings with these two different methods were qualitatively highly concordant, both indicating strong stability of subphenotype membership over time, thus providing further support for this conclusion.

While most patients remained in the same subphenotype from day 0 to day 3, approximately 5.5% of patients in each cohort transitioned between subphenotypes over time. Interestingly, some patients moved from the hypoinflammatory class 1 to the hyperinflammatory class 2 over time, suggesting that some patients progressively developed more severe inflammation than was initially present at the time of ARDS development. Notably, for patients who changed subphenotype over time, clinical outcomes were more strongly associated with the later subphenotype, indicating that the transition from the hypoinflammatory class 1 to the hyperinflammatory class 2 may be reflective of an overall declining clinical trajectory. However, the number of patients who changed subphenotype over time was small, so additional replication is needed to confirm these findings.

These data provide an interesting opportunity to observe how the clinical and biological features of each subphenotype change over time (figures 1 and 2; online supplementary tables S6 and S7). While many of the variables distinguishing the two groups are the same at both time points, some variables contribute less to subphenotype identification over time, as reflected by more similar values between the two classes at day 3 compared with day 0: for instance, PaCO<sub>2</sub> and heart rate. In contrast, other variables clearly become more different between the two subphenotypes over time; this list includes indices of organ failure (bilirubin, PF ratio and creatinine). This finding is concordant with our previously published data that patients in class 2 go on to have fewer organ failure-free days compared with those in class 1.<sup>2,3</sup> It also suggests that

**Table 6** Clinical outcomes by transition status in ARMA (panel A) and ALVEOLI (panel B)

Outcome*	Stay in class 1 (n=313)	Class 1 to class 2 (n=17)	Class 2 to class 1 (n=8)	Stay in class 2 (n=120)
(A) ARMA				
Mortality at day 90, n (%)	57 (18)	13 (77)	2 (25)	58 (48)
Ventilator-free days, median (25th–75th percentile)	20 (6, 24)	0 (0, 0)	21 (9, 22)	0 (0, 14.5)
Organ-failure free days, median (25th–75th percentile)	23 (11, 27)	0 (0, 2)	19.5 (4.5, 26)	0 (0, 14.5)
Outcome*	Stay in class 1 (n=358)	Class 1 to class 2 (n=18)	Class 2 to class 1 (n=11)	Stay in class 2 (n=138)
(B) ALVEOLI				
Mortality at day 90, n (%)	56 (16)	7 (39)	1 (9)	61 (44)
Ventilator-free days, median (25th–75th percentile)	22 (12, 25)	0 (0, 16)	23 (19, 25)	3.5 (0, 18)
Organ-failure free days, median (25th–75th percentile)	23 (14, 27)	6 (0, 18)	23 (11, 25)	5.5 (0, 17.5)

\*All comparisons significant at P<0.0001.

ongoing biological differences between the classes remain relevant at day 3 (and therefore potentially targetable).

This study has several important strengths, including the consistency of findings in two distinct independent cohorts, a diverse patient sample drawn from a wide variety of centres in highly protocolised clinical trials and the measurement of eight biomarkers reflective of specific aspects of ARDS pathogenesis. This study also has some limitations. First, not all data points used in our initial identification of ARDS subphenotypes were available at day 3, likely contributing to a modest loss of power. Second, some patients included at day 0 were not included at day 3, almost entirely due to death before that time point. Since these patients likely represented extreme phenotypes, this attrition may have diminished the heterogeneity in the cohort at day 3 and therefore our ability to distinguish subtypes.

In summary, we found strong evidence in support of the stability of two distinct ARDS subphenotypes over the first 3 days of enrolment in two clinical trials. These data provide important evidence that ARDS subphenotypes are durable, providing strong additional support for the potential value of subphenotype-targeted therapies in ARDS. Future studies should focus on methods for rapidly classifying patients by subphenotype in real time including the development of point-of-care assays, on how these subphenotypes may respond differently to ARDS therapies and on deeper study of the biology of each subphenotype.<sup>4</sup>

**Contributors** KD, KRF and CSC drafted the manuscript. KD and KRF performed the statistical analyses. LBW and PEP contributed biomarker measurements. CSC supervised the study design, data interpretation and manuscript preparation. All authors contributed significantly to study design and/or interpretation. All authors made significant intellectual contributions to the final manuscript and approve its submission.

**Funding** NIH grants T32 HL1785-39 and F32 HL129680-01 (KRF); HL103836 and HL112656 (LBW); HL131621 and HL133390 (CSC, KD). NIH/NHLBI ARDS Network contracts: N01-HR 46054-46064.

**Competing interests** CSC has ongoing grant funding from Bayer, prior grant funding from GlaxoSmithKline and consulting for GlaxoSmithKline, Bayer and Boehringer Ingelheim. Other authors have no competing interests to declare.

**Ethics approval** Ethics approval obtained for original studies at each participating institution.

**Provenance and peer review** Not commissioned; externally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

- Prescott HC, Calfee CS, Thompson BT, et al. Toward smarter lumping and smarter splitting: Rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. *Am J Respir Crit Care Med* 2016;194:147–55.
- Calfee CS, Delucchi K, Parsons PE, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014;2:611–20.
- Famous KR, Delucchi K, Ware LB, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med* 2017;195:331–8.
- Thompson BT, Chambers RC, Liu KD. Acute respiratory distress syndrome. *N Engl J Med* 2017;377:562–72.
- Matthay MA, McAuley DF, Ware LB. Clinical trials in acute respiratory distress syndrome: challenges and opportunities. *Lancet Respir Med* 2017;5:524–34.
- Brower RG, Matthay MA, Morris A, 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;342:1301–8.
- Brower RG, Lanken PN, MacIntyre N, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med* 2004;351:327–36.
- The ARDS Network. Ketoconazole for early treatment of acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. The ARDS Network. *JAMA* 2000;283:1995–2002.
- Randomized, placebo-controlled trial of lisofylline for early treatment of acute lung injury and acute respiratory distress syndrome. *Crit Care Med* 2002;30:1–6.
- Parsons PE, Eisner MD, Thompson BT, et al. Lower tidal volume ventilation and plasma cytokine markers of inflammation in patients with acute lung injury. *Crit Care Med* 2005;33:1–6.
- Ware LB, Matthay MA, Parsons PE, et al. Pathogenetic and prognostic significance of altered coagulation and fibrinolysis in acute lung injury/acute respiratory distress syndrome. *Crit Care Med* 2007;35:1821–8.
- Calfee CS, Eisner MD, Parsons PE, et al. Soluble intercellular adhesion molecule-1 and clinical outcomes in patients with acute lung injury. *Intensive Care Med* 2009;35:248–57.
- Eisner MD, Parsons P, Matthay MA, et al. Plasma surfactant protein levels and clinical outcomes in patients with acute lung injury. *Thorax* 2003;58:983–8.
- Ware LB, Eisner MD, Thompson BT, et al. Significance of von Willebrand factor in septic and nonseptic patients with acute lung injury. *Am J Respir Crit Care Med* 2004;170:766–72.
- Ware LB, Koyama T, Billheimer DD, et al. Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. *Chest* 2010;137:288–96.
- Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 1989;57:307–33.
- Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* 1996;13:195–212.
- Collins LM, Flaherty BP. Latent class models for longitudinal data. In: Hagennars JA, McCutcheon AL, eds. *Applied latent class analysis*. New York: Cambridge University Press, 2002:287–305.
- Famous KR, Delucchi K, Ware LB, et al. Latent class analysis demonstrates the stability of ARDS endotypes over time. *Am J Respir Crit Care Med* 2016;193:A7791.
- Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal* 2007;14:535–69.
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978;6:461–4.
- Shankar-Hari M, McAuley DF. Acute respiratory distress syndrome phenotypes and identifying treatable traits. The dawn of personalized medicine for ARDS. *Am J Respir Crit Care Med* 2017;195:280–1.