Cluster analysis identifies novel real-world lung disease–pulmonary hypertension subphenotypes: implications for treatment response

Shelsey W. Johnson[1,2,3], Rui-Sheng Wang[4], Michael R. Winter[5], Kari R. Gillmeyer[1,2,6], Katarina Zeder[7,8], Elizabeth S. Klings[2], Ronald H. Goldstein[1], Renda Soylemez Wiener[1,2,6] and Bradley A. Maron[1,8,9]

[1]VA Boston Healthcare System, Boston, MA, USA. [2]The Pulmonary Center, Division of Pulmonary, Allergy, Sleep and Critical Care, and Boston University School of Medicine, Boston, MA, USA. [3]Department of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA. [4]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [5]Boston University School of Public Health, Biostatistics and Epidemiology Data Analytics Center, Boston, MA, USA. [6]Center for Healthcare Organization and Implementation Research, VA Bedford Healthcare System, Bedford, MA, USA. [7]Department of Pulmonology, Medical University of Graz and Ludwig Boltzmann Institute for Lung Vascular Research, Graz, Austria. [8]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [9]The University of Maryland-Institute for Health Computing, Bethesda, MD, USA.

Corresponding author: Shelsey W. Johnson (swjohnson@mgb.org)

Shareable abstract (@ERSpublications)
K-means cluster analysis identifies novel subphenotypes of real-world WHO Group 3 PH patients with distinct survival trajectories. In an exploratory within-cluster treatment analysis, cluster assignment determined treatment response. https://bit.ly/3wjK5Fw

Cite this article as: Johnson SW, Wang R-S, Winter MR, *et al*. Cluster analysis identifies novel real-world lung disease–pulmonary hypertension subphenotypes: implications for treatment response. *ERJ Open Res* 2024; 10: 00959-2023 [DOI: 10.1183/23120541.00959-2023].

Received: 1 Dec 2023
Accepted: 15 Feb 2024

Abstract

Background Clinical trials repurposing pulmonary arterial hypertension (PAH) therapies to patients with lung disease- or hypoxia-pulmonary hypertension (PH) (classified as World Health Organization Group 3 PH) have failed to show a consistent benefit. However, Group 3 PH clinical heterogeneity suggests robust phenotyping may inform detection of treatment-responsive subgroups. We hypothesised that cluster analysis would identify subphenotypes with differential responses to oral PAH therapy.

Methods Two k-means analyses were performed on a national cohort of US veterans with Group 3 PH; an inclusive model (I) of all treated patients (n=196) and a haemodynamic model (H) limited to patients with right heart catheterisations (n=112). The primary outcome was organ failure or all-cause mortality by cluster. An exploratory analysis evaluated within-cluster treatment effects.

Results Three distinct clusters of Group 3 PH patients were identified. In the inclusive model ($C1^I$ n=43, 21.9%; $C2^I$ n=102, 52.0%; $C3^I$ n=51, 26.0%), lung disease and spirometry drove cluster assignment. By contrast, in the haemodynamic model ($C1^H$ n=44, 39.3%; $C2^H$ n=43, 38.4%; $C3^H$ n=25, 22.3%), right heart catheterisation data surpassed the importance of lung disease and spirometry. In the haemodynamic model, compared to $C3^H$, $C1^H$ experienced the greatest hazard for respiratory failure or death (HR 6.1, 95% CI 3.2–11.8). In an exploratory analysis, cluster determined treatment response (p=0.006). Conclusions regarding within-cluster treatment responses were limited by significant differences between select variables in the treated and untreated groups.

Conclusions Cluster analysis identifies novel real-world subphenotypes of Group 3 PH patients with distinct clinical trajectories. Future studies may consider this methodological approach to identify subgroups of heterogeneous patients that may be responsive to existing pulmonary vasodilatory therapies.

Introduction

Pulmonary hypertension (PH) associated with chronic lung disease (CLD) and/or hypoxia (World Health Organization (WHO) Group 3 PH) is a highly morbid condition, with mortality risk exceeding that of pulmonary arterial hypertension (PAH) by 5-fold [1–3]. Currently, the cornerstone of treatment for Group 3 PH is optimising the underlying lung disease or lung transplantation for eligible patients [4, 5].

Clinical trials repurposing PAH therapies to Group 3 PH patients have failed to demonstrate consistent benefit [6–10]. However, these data are limited by small sample sizes, variable inclusion criteria and divergent trial end-points [6, 7]. Cumulative analyses from prospective trials [7, 11, 12] and clinical practice [13, 14] indicate signs of favourable outcomes in certain Group 3 PH subgroups, yet the precise clinical profile that informs treatment-responsiveness is unknown [15].

Importantly, the pathophysiology and clinical profile of Group 3 PH share features across CLDs, but are simultaneously heterogenous. For example, shared mechanisms of pulmonary vascular injury are implicated in both COPD-PH and interstitial lung disease-PH with common aspects of clinical presentation [16, 17]. However, divergent disease mechanisms exist and, when present, Group 3 PH occurs across a wide haemodynamic spectrum that strongly influences mortality risk [2, 3]. Additionally, overlapping risk factors for Group 2 PH are common in real-world patients who rarely present with the isolated pre-capillary disease represented in clinical trial populations [18]. Despite these observations, methods used in clinical studies to test PAH therapy response in Group 3 PH hinge on reductionist approaches that overemphasise a narrow subset of variables. Therefore, these methods may be insufficient to capture discrete subgroups and their associated therapeutic response [19]. Recently, agnostic and robust strategies have been used to phenotype and prognosticate PAH [20–23]. In particular, cluster analyses, which aim to identify subgroups based on shared but nonintuitive patterns within aggregated traits [24], have unmasked nuanced PAH subgroups that were otherwise missed using probabilistic models alone [25]. Thus, deploying novel analytical methods to subphenotype Group 3 PH patients may inform identification of within-group differences including treatment response patterns to repurposed PAH therapies.

The Veteran's Health Administration is the largest integrated national health system in the USA and includes a high prevalence of CLD and Group 3 PH [26, 27]. Within this cohort, off-label prescribing of PAH therapies to Group 3 PH patients is common and systematically reported [28–30]. We hypothesised that analysing a large cohort of veterans with Group 3 PH using k-means cluster analysis would identify real-world subgroups defined by discrete clinical and physiological profiles with differential responses to oral PAH therapies.

Methods

Study subjects

Veterans with Group 3 PH diagnosed between 1 January 2006 and 31 December 2017 were identified using a previously validated claims-based algorithm [29–31]. Within this group, patients exposed to oral PAH therapy between 1 January 2006 and 31 December 2016 were then selected to allow at least 1 year of follow-up (n=334). A pulmonologist (SWJ) reviewed the medical records of all exposed patients to confirm the Group 3 PH diagnosis and initiation of PAH therapy (n=196). For patients in whom right heart catheterisation (RHC) data were available, a pre-capillary PH diagnosis was made based on a mean pulmonary artery pressure (mPAP) >20 mmHg, pulmonary artery wedge pressure (PAWP) ⩽15 mmHg and pulmonary vascular resistance (PVR) >2 WU; a combined pre- and post-capillary diagnosis was made based on a PAWP >15 mmHg and PVR >2 WU [4, 5]. In the absence of available RHC data, as is common in real-world populations diagnosed with lung disease PH [14], a Group 3 PH diagnosis was made based on composite clinical data with emphasis on echocardiographic PH features [32]. Those with PH identified by echocardiogram alone were intentionally included in the analysis to understand the importance of clinical variables typically obtained in the evaluation of real-world lung disease PH populations and to compare their relevance to "gold standard" haemodynamic data.

To establish the cohort, patients exposed to PAH therapy were matched 1:1 to unexposed Group 3 PH patients. Matching criteria included 1) age within 5 years of exposed patient at time of PH diagnosis and 2) claim-based PH diagnosis in the same calendar year. Patients were not matched based on underlying CLD for several reasons. Several pathophysiological mechanisms implicated in the development of CLD-PH are shared across lung diseases, including the impact of exposure to chronic hypoxia and cigarette smoke and the influences of oxidative stress and cellular senescence on the pulmonary vasculature [16, 17]. Therefore, the "vascular phenotype" of CLD may reflect a common pathway of divergent upstream diagnoses, more likely to be captured and characterised with our inclusive approach. This approach aligns with current recommendations for consideration of pulmonary vasodilator therapy in patients with CLD, regardless of their underlying lung disease [33].

SWJ then reviewed records of matched untreated patients to confirm Group 3 PH and lack of exposure to PAH therapy (n=196). Overall cohort selection (n=392) and exclusion criteria are outlined in the supplemental material and supplemental figure S1. This research was approved by the VA Boston Healthcare System R&D committee (project #1590467).

Cluster variables

Variables for inclusion in the cluster analysis were extracted from the medical record and entered into a de-identified REDCap research database [34]. For patients treated with PAH therapies, data were extracted from the outpatient visit at which PAH treatment was prescribed. To standardise data extraction for the untreated patients, data were similarly extracted closest to the script date of the matched treated patient. Extracted variables included standard-of-care testing for a suspected diagnosis of Group 3 PH including pulmonary function testing, echocardiography and RHC data (if available) [4, 5]. For treated patients, only RHC data collected prior to start of therapy were collected. Because isolated Group 3 PH is rare in real-world populations within the age range of our study cohort [18], variables suggesting concomitant left heart disease including left ventricular hypertrophy or left atrial enlargement were also included. Finally, outpatient vital signs including supplemental oxygen requirements and blood pressure were extracted for their relevance to risk of adverse effects of PAH therapy. Radiographic data were not extracted because individual images could not be reviewed and validated independently [35]. From this dataset of extracted variables (supplemental table S1), those with ⩽20% missing values were eligible for inclusion in cluster analyses (n=25 variables; supplemental table S2). Mean substitution and higher frequency substitution were used to impute data for missing continuous and binary variables, respectively [36].

Primary and secondary outcomes

The primary outcome was a composite time to death by any cause or time to acute hospitalisation or emergency department visit for acute respiratory, right heart or renal failure, by treated patient clusters. The secondary outcome was a composite of mortality or a respiratory failure event. These were selected based on literature indicating that PAH therapy may alter physiological ventilation/perfusion (V/Q) matching, resulting in hypoxia and increased risk of respiratory hospitalisations when used by patients with Group 3 PH [10, 37]. In an exploratory analysis of treatment effect within cluster, outcomes were identified following the PAH therapy script start date for both treated and untreated patients. To mitigate immortal time bias, untreated patients were followed from the script date of their matched treated counterpart to ensure untreated patients were at a time point beyond PH diagnosis [31].

Statistical analyses

Two k-means cluster analyses were performed in MATLAB to identify subgroups defined by distinct clinical profiles [38]. Accepting the limited uptake of RHC to confirm Group 3 PH in real-world practice [14, 39], the first k-means analysis (termed the inclusive model, I) was performed with all Group 3 PH patients treated with PAH therapy (n^{I}=196), including those diagnosed by echocardiography alone [38]. Owing to the importance of RHC data to diagnose PH and define Group 3 PH severity, a second cluster analysis (termed the haemodynamic model, H) included patients with available RHC data (n^{H}=112) (supplemental table S2). Following selection of k, clustering was run 1000 times to obtain consensus clustering (supplemental figure S2). Random forest-based variable importance evaluation was performed in the R package "caret" (https://rbasics.org/packages/caret-package-in-r/) to rank the importance of individual variables in cluster assignment. Cluster matching was then used to assign each untreated patient to a treated patient cluster according to the Euclidean distance of the scaled clinical profiles to the treated cluster centres [36]. Additional methodological details are outlined in the supplemental material. Descriptive outputs from cluster analyses are presented as mean±SD or median (interquartile range (IQR)) for normally and non-normally distributed continuous variables, respectively, and n (%) for categorical variables. Continuous variables were compared across clusters using the one-way ANOVA test for means and Kruskal–Wallis test for medians; categorical variables were compared using chi-square testing.

Survival analysis with Kaplan–Meier curves and log-rank tests were used to evaluate time-to-event across clusters and within clusters by treatment. Univariate Cox regression analysis was performed to assess the effect of cluster, a defined composite of relevant covariates, on outcomes. In our exploratory analysis, we additionally used univariate Cox regression to assess the effect of treatment within cluster and to look for evidence of effect modification by cluster using a cluster-by-treatment interaction [40]. The proportional hazards assumption was confirmed for all analyses. A sensitivity analysis was performed with Fine–Gray subdistribution hazard models for respiratory failure, with death as a competing risk, and reported with cumulative incidence functions [41]. Survival analysis was performed using SAS statistical software, version 9.4 (SAS Institute, Cary, NC, USA). All p-values are two-sided with a significance level of 0.05.

Results

From a total of 392 veterans (97.7% male) with Group 3 PH, treated and untreated patients (n=196 each) were characterised by similar age and rates of COPD (67.4% *versus* 73.0%). However, oxygen dependence (79.1% *versus* 43.4%) and idiopathic pulmonary fibrosis (IPF) (24.0% *versus* 6.1%) were more common in the treated group (supplemental table S3). In the overall cohort, RHC was performed in 34.9% of patients

(n=137): 57.1% of treated patients (n=112) and 12.8% of untreated patients (n=25). The haemodynamic profile of treated patients was more severe than that of the untreated patients (mean PVR 6.4 *versus* 2.9 WU). Phosphodiesterase-5 inhibitors (PDE5i) were the most common PAH medication prescribed (n=184, 93.9%), followed by endothelin receptor antagonists (ERA) (n=25, 12.8%) and combination PDE5i+ERA therapy (n=14, 7.1%).

Group 3 PH clusters

In the inclusive model, k-means identified three distinct patient clusters: Cluster 1^{I} ($C1^{I}$; n=43), Cluster 2^{I} ($C2^{I}$; n=102) and Cluster 3^{I} ($C3^{I}$; n=51), which accounted for 21.9%, 52.0% and 26.0% of the cohort, respectively. In the haemodynamic model, restricted to treated patients with RHC data (n=112), three subphenotypes were identified: Cluster 1^{H} ($C1^{H}$; n=44), Cluster 2^{H} ($C2^{H}$; n=43) and Cluster 3^{H} ($C3^{H}$; n=25), which accounted for 39.3%, 38.4% and 22.3% of patients with RHC data, respectively. The overall clinical characteristics of each cluster are presented in tables 1 and 2 and the importance of specific variables that were used to drive the assignment of patients to particular clusters is presented in rank order in supplemental figures S3 and S4.

Clinical profiles of patients by cluster

In the inclusive model, there were no clinically meaningful differences across C1–C3 in age at PH diagnosis ($C1^{I}$: 69 years, IQR 66–76 years; $C2^{I}$: 72 years, IQR 68–78 years; $C3^{I}$: 68 years, IQR 66–76 years; p=0.02). Patients in $C1^{I}$ had a significantly greater body mass index (BMI) (32 $kg \cdot m^{-2}$, IQR

TABLE 1 Clinical profiles of treated World Health Organization Group 3 pulmonary hypertension patients by cluster among all patients treated with pulmonary arterial hypertension therapy (n=196) included in the inclusive model

	Cluster 1	Cluster 2	Cluster 3	p-value
Patients	43 (21.9)	102 (52.0)	51 (26.0)	
Age at PH diagnosis (years)	69 (66–76)	72 (68–78)	68 (66–76)	0.02
Body mass index ($kg \cdot m^{-2}$)	32 (29.5–38.3)	26.1 (23.1–29.4)	29.5 (24.9–31.9)	<0.001
Systolic blood pressure (mmHg)	132 (12– 146)	124 (113–135)	120 (110–128)	<0.001
Creatinine ($mg \cdot dL^{-1}$)	1.1 (1.0–1.3)	1.1 (0.9–1.2)	1.03 (0.9–1.2)	0.35
Comorbidities				
Hypertension	35 (81.4)	76 (74.5)	32 (62.7)	0.11
Diabetes	24 (55.8)	21 (20.6)	16 (31.4)	<0.001
Chronic lung disease				
COPD	32 (74.4)	90 (88.2)	10 (19.6)	<0.001
IPF	1 (2.3)	2 (2.0)	44 (86.3)[¶]	<0.001
Non-IPF interstitial lung disease	7 (16.3)	30 (29.4)	10 (19.6)	0.17
Obstructive sleep apnoea	32 (74.4)	13 (12.7)	7 (13.7)	<0.001
Oxygen dependence	17 (39.5)	91 (89.2)	47 (92.2)	<0.001
Pulmonary function test data				
FEV_1 (% predicted)	67.0±19.3	54.8±23.8	66.0±22.4	
FVC (% predicted)	75.1±18.3	74.2±21.9	63.0±22.7	
FEV_1/FVC (%)	67 (61–74)	57.6 (41.5–64.5)[¶]	80 (74–86.5)	
D_{LCO} (% predicted)	50 (42–62.5)[¶]	34.7 (24.3–45.2)	31.7 (23.1–43.8)	
Echocardiogram data[#]				
Inferior vena cava dilation	13 (30.2)	18 (17.6)	3 (5.9)	0.008
Right atrial dilation	28 (65.1)	78 (76.4)	34 (66.7)	0.26
Right ventricular dilation	27 (62.8)	77 (75.4)	36 (70.6)	0.30
RVSP (mmHg)	60 (50–77)	60 (57–80)	60 (55–73)	0.66
Left ventricle ejection fraction (%)	60 (55–60)	58 (55–63)	58 (55–60)	0.78
Left atrial dilation	28 (65.1)	27 (26.4)	12 (23.5)	<0.001
Left ventricular hypertrophy	20 (46.5)	22 (21.6)	14 (27.5)	0.01

Data for continuous variables are presented as mean±SD or median (interquartile range) for normally and non-normally distributed variables, respectively, and categorical variables are presented as n (%). p-values are calculated using Kruskal–Wallis test for medians, one-way ANOVA for means and chi-square for percentages. PH: pulmonary hypertension; IPF: idiopathic pulmonary fibrosis; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lungs for carbon monoxide; RVSP: right ventricular systolic pressure estimate. [#]: n (%) refers to the number of patients with the described echocardiographic feature; [¶]: denotes the variable most important in cluster determination (also see supplemental figure S3).

TABLE 2 Clinical profiles of treated World Health Organization Group 3 pulmonary hypertension patients by cluster among those with right heart catheterisation (n=112) included in the haemodynamic model

	Cluster 1	Cluster 2	Cluster 3	p-value
Patients	44 (39.3)	43 (38.4)	25 (22.3)	
Age at PH diagnosis (years)	70 (66–77)	69 (66–74)	68 (66–73)	0.42
Body mass index (kg·m^{-2})	27.0 (24.8–29.6)	27.4 (24.1–30.0)	33.8 (31.5–41.7)	<0.001
Systolic blood pressure (mmHg)	120 (110–134)	125 (113–136)	130 (123–142)	0.14
Creatinine (mg·dL^{-1})	1.1 (0.9–1.3)	1.0 (0.8–1.2)	1.1 (1.0–1.3)	0.36
Comorbidities				
Hypertension	37 (84.1)	24 (55.8)	19 (76.0)	0.01
Diabetes	12 (27.3)	15 (34.9)	12 (48.0)	0.22
Chronic lung disease				
Chronic obstructive pulmonary disease	35 (79.5)	30 (69.8)	16 (64.0)	0.34
IPF	7 (15.9)	8 (18.6)	1 (4.0)	0.23
Non-IPF interstitial lung disease	12 (27.3)	14 (32.6)	1 (4.0)	0.02
Obstructive sleep apnoea	8 (18.2)	7 (16.3)	22 (88.0)	<0.001
Oxygen dependence	40 (90.9)	39 (90.7)	7 (28.0)	<0.001
Pulmonary function test data				
FEV_1 (% predicted)	72.5±20.4	51.1±21.8¶	63.8±17.2	<0.001
FVC (% predicted)	88.0±18.8	65.5±19.7	70.2±17.5	<0.001
FEV_1/FVC (%)	65.3 (56.5–72.5)	61 (46.5–73.5)	68 (61.5–75)	0.0966
D_{LCO} (% predicted)	29.5 (23–45.2)	32 (23.9–37.5)	51.3 (40–63)¶	<0.001
Echocardiogram data#				
Inferior vena cava dilation	14 (31.8)	2 (4.7)	4 (16.0)	0.004
Right atrial dilation	38 (86.4)	27 (62.8)	17 (68.0)	0.04
Right ventricular dilation	42 (95.5)	26 (60.5)	17 (68.0)	<0.001
RVSP (mmHg)	80 (64–95)	59 (48–60)	60 (50–75)	<0.001
Left ventricle ejection fraction (%)	60 (55–65)	58 (55–60)	60 (55–60)	0.11
Left atrial dilation	24 (54.5)	5 (11.6)	12 (48.0)	<0.001
Left ventricular hypertrophy	18 (40.9)	5 (11.6)	11 (44.0)	0.003
Right heart catheterisation data				
Mean pulmonary artery pressure (mmHg)	46.9±9.6	34.8±7.9	40.1±11.4	<0.001
Pulmonary artery wedge pressure (mmHg)	12 (8.8–14)	13 (9–15)	16 (14–18)	<0.001
Pulmonary vascular resistance (WU)	8.3 (5.7–10.9)¶	4.6 (3.1–5.7)	4.6 (3.4–5.7)	<0.001

Data for continuous variables are presented as mean±SD or median (interquartile range) for normally and non-normally distributed variables, respectively, and categorical variables are presented as n (%). p-values are calculated using Kruskal–Wallis test for medians, one-way ANOVA for means and chi-square for percentages. PH: pulmonary hypertension; IPF: idiopathic pulmonary fibrosis; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lungs for carbon monoxide; RVSP: right ventricular systolic pressure estimate. #: n (%) refers to the number of patients with the described echocardiographic feature; ¶: denotes the variable most important in cluster determination (also see supplemental figure S4).

29.5–38.3 kg·m^{-2} *versus* C2^{I}: 26.1 kg·m^{-2}, IQR 23.1±29.4 kg·m^{-2} and C3^{I}: 29.5 kg·m^{-2}, IQR 24.9–31.9 kg·m^{-2}; p<0.001) and systolic blood pressure (132 mmHg, IQR 128–146 mmHg *versus* C2^{I}: 124 mmHg, IQR 113–135 mmHg and C3^{I}: 120 mmHg, IQR 110–128 mmHg; p<0.001). Most patients in C1^{I} (74.4%) had moderate COPD (forced expiratory volume in 1 s (FEV_1) 67.0±19.3% predicted) or COPD–obstructive sleep apnoea (OSA) overlap syndrome (74.4%). In fact, a concomitant diagnosis of OSA was strongly important in determining cluster assignment (supplemental figure S3). C1^{I} patients often had systemic hypertension (81.4%) and diabetes mellitus (55.8%), which are established risk factors for pathogenic remodelling of the left heart [42] that may predispose to post-capillary PH. Indeed, the echocardiographic profile of C1^{I} patients showed that 65% of this subphenotype had left atrial dilation and 47% had left ventricular hypertrophy, which were 2.8-fold and 2.2-fold greater, respectively, than in the clusters with the lowest prevalence.

The predominant lung pathophenotype in C2^{I} was also COPD (88.2%). Compared to C1^{I} and C3^{I}, these patients had the most severe airflow obstruction (FEV_1 for C2^{I}: 54.8±23.8% predicted *versus* C1^{I}: 67.0 ±19.3% predicted and C3^{I}: 66.0±22.4% predicted). Few patients had overlapping OSA (12.7%). In this cluster, one third of patients had evidence of overlapping interstitial lung disease (29.4%). In contrast to C1^{I} and C2^{I}, C3^{I} included mainly patients with a diagnosis of IPF (86.3%), which was internally consistent with our observation that this subphenotype also had the most severely impaired diffusing

capacity of the lung for carbon monoxide (D_{LCO}) ($C1^I$: 50%, IQR 42–62.5% predicted; $C2^I$: 34.7%, IQR 24.3–45.2% predicted; $C3^I$: 31.7%, IQR 23.1–43.8% predicted; p<0.001) and the highest number of patients dependent on supplemental oxygen ($C1^I$: 39.5%; $C2^I$: 89.2%; $C3^I$: 92.2%; p<0.001). In $C3^I$, preserved FEV_1 to forced vital capacity (FVC) ratio (80.8%, IQR 74–86.5%), reflective of restrictive lung disease, was more important for determining cluster assignment than D_{LCO} (supplemental figure S3). Echocardiographic features of PH, including right atrial and ventricular dilation and elevated right ventricular systolic pressure (RVSP) estimates, were present across $C1–3^I$, but not significantly different between clusters (table 1).

In the haemodynamic model, age did not significantly differ between groups; however, BMI was greatest in $C3^H$ (33.8 kg·m^{-2}, IQR 31.5–41.7 kg·m^{-2} *versus* $C1^H$: 27.0 kg·m^{-2}, IQR 24.8–29.6 kg·m^{-2} and $C2^H$: 27.4 kg·m^{-2}, IQR 31.5–41.7 kg·m^{-2}; p<0.001). In this model, the underlying lung disease distributed more evenly across clusters. For example, the prevalence of COPD and IPF were not significantly different across clusters. Echocardiographic parameters supportive of PH were most prevalent in $C1^H$ as compared to $C2^H$ and $C3^H$, including the percentage of patients with at least mild right atrial dilation (86.4% *versus* $C2^H$: 62.8% and $C3^H$: 68.0%; p=0.04), right ventricular dilation (95.5% *versus* $C2^H$: 60.5% and $C3^H$: 68.0%; p<0.001) and elevated RVSP (80 mmHg, IQR 64–95 mmHg *versus* $C2^H$: 59 mmHg, IQR 48–60 mmHg and $C3^H$: 60 mmHg, IQR 50–75 mmHg; p<0.001). Focusing on cardiopulmonary haemodynamic parameters, $C1^H$ included patients with the highest mPAP (46.9±9.6 mmHg) and PVR (8.3 WU, IQR 5.7–10.9 WU) but normal PAWP (12 mmHg, IQR 8.8–14 mmHg). $C2^H$ patients had less severe haemodynamic abnormalities (mPAP 34.8±7.9 mmHg and PVR 4.6 WU, IQR 3.1–5.7 WU) and $C3^H$ included patients with a profile that was, overall, most consistent with combined pre- and post-capillary PH (mPAP 40.1±11.4 mmHg, PAWP 16 mmHg, IQR 14–18 mmHg and PVR 4.6 WU, IQR 3.4–5.7 WU) (table 2).

PVR was most important in determining assignment to $C1^H$ (supplemental figure S4), which was largely characterised by COPD patients (79.5%) on supplemental oxygen (90.9%). $C2^H$ included patients with the most severe obstructive lung disease (FEV_1 51.1±21.8% predicted) and reduced D_{LCO} (32%, IQR 23.9–37.5% predicted). In this cluster, spirometric data and haemodynamic parameters were more important to cluster assignment than D_{LCO}. OSA was most common in $C3^H$ (88%), which included patients with the highest BMI (33.8 kg·m^{-2}, IQR 31.5–41.7 kg·m^{-2}) and the most preserved D_{LCO} (51.3%, IQR 40–63% predicted) (table 2).

Outcomes by Group 3 PH clusters

Patients were followed for 3.4±2.6 years after the start of PAH therapy. In the inclusive model, 29 patients (67.4%) in $C1^I$, 94 patients (92.2%) in $C2^I$ and 50 patients (98.0%) in $C3^I$ experienced the primary outcome. In the haemodynamic model, the primary outcome occurred in 43 patients (97.7%), 36 patients (83.7%) and 14 patients (56%) in $C1^H$, $C2^H$ and $C3^H$, respectively. For both the inclusive and haemodynamic models, the primary and secondary outcomes differed significantly by cluster (log rank p<0.001) (figure 1). In the inclusive model, compared to $C1^I$, patients in $C2^I$ and $C3^I$ had a significantly increased risk of both organ-specific failure or death ($C2^I$: HR 2.3, 95% CI 1.5–3.6; $C3^I$: HR 3.0, 95% CI 1.9–4.8) as well as respiratory failure or death ($C2^I$: HR 2.6, 95% CI 1.7–4.0; $C3^I$: HR 3.1, 95% CI 1.9–4.9). In a multivariate analysis adjusting for CLD, the differential outcomes by cluster remained significant. In the haemodynamic model, $C1^H$ experienced a 6-fold increased hazard of respiratory failure or death (HR 6.1, 95% CI 3.2–11.8) compared with $C3^H$ (table 3).

Treatment effect within Group 3 PH clusters

Within clusters, the clinical profiles of the treated compared to the untreated patients were largely similar but with important and significant differences for select variables (supplemental tables S4–S7). In the inclusive model, treated patients were more likely to be on oxygen, with more severe impairments in D_{LCO} and elevations in RVSP. In the haemodynamic model, treated patients had more significant elevations in mPAP and PVR across clusters, with similar PAWP. Acknowledging these differences, exploratory analyses of treatment effect within cluster demonstrated increased risk of adverse outcomes associated with therapy (log rank p<0.05; figure 2 and supplemental figure S5). For clusters defined by the haemodynamic model, there was evidence of effect modification by cluster in which clinical risk by treatment depended on cluster assignment. For example, treatment increased the hazard for death or respiratory failure by 3-fold for $C1^H$ (97.7% treated *versus* 67.2% untreated; HR 3.2, 95% CI 2.0–4.9; p<0.001) whereas no significant effect was observed for treatment for $C3^H$ patients (52% treated *versus* 41.9% untreated; HR 1.1, 95% CI 0.5–2.2; p=0.81) (table 4). In a competing risk sensitivity analysis, the hazard associated with treatment was equivocal across all clusters, both for the inclusive and haemodynamic models

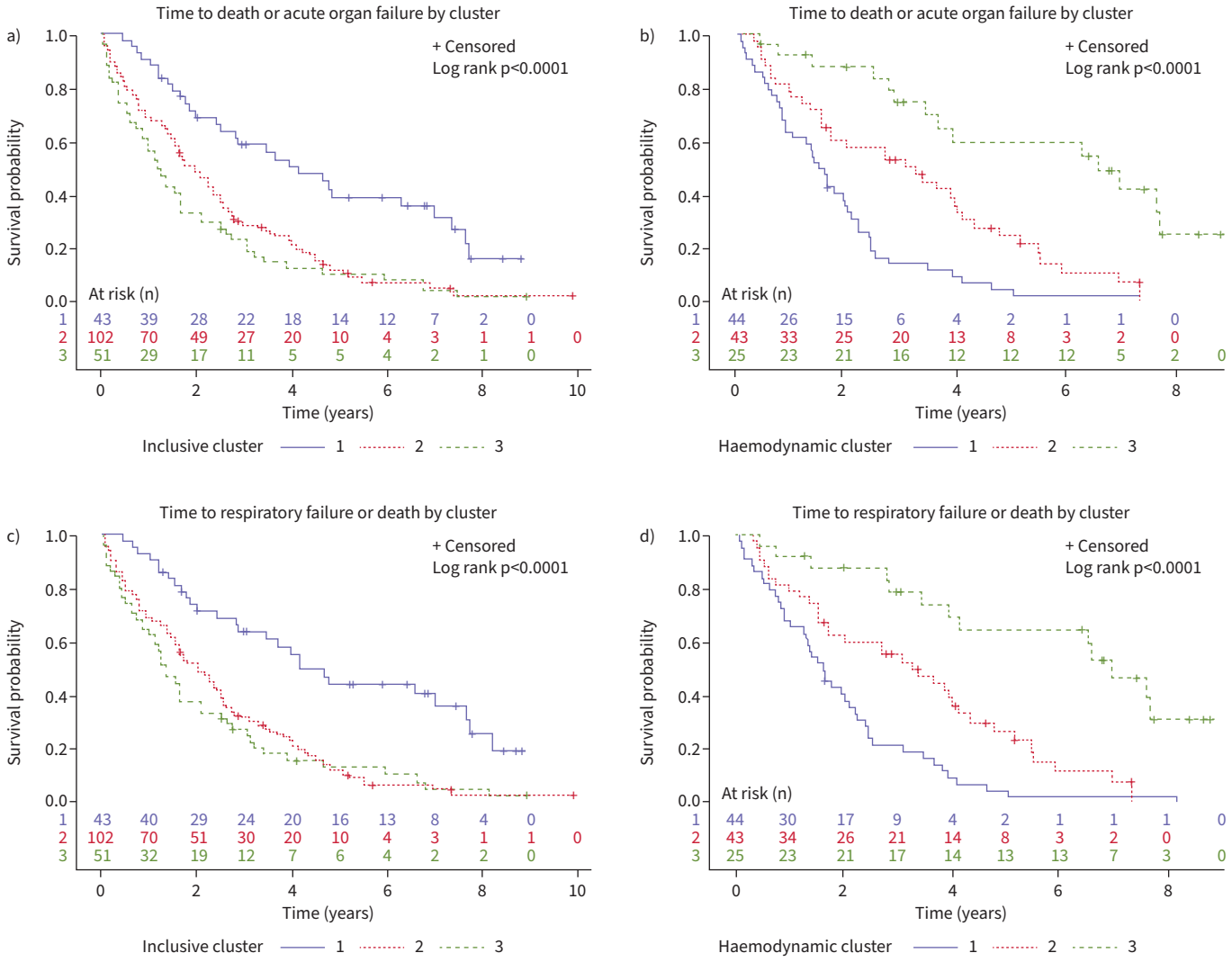

FIGURE 1 Survival curves by cluster analysis-defined Group 3 pulmonary hypertension subphenotypes. a, b) Primary outcome (composite time to death by any cause or time to acute hospitalisation or emergency department visit for acute respiratory, right heart or renal failure) by cluster. c, d) Secondary outcome of a composite of time to death or respiratory failure by cluster. Panels a and c represent the inclusive model and panels b and d represent the haemodynamic model.

(supplemental table S8). For example, 34.1% of treated and 38.8% of untreated patients from $C1^H$ of the haemodynamic model experienced respiratory failure (Gray's test p=0.58) (supplemental figure S6).

Discussion

To our knowledge, this is the first study using cluster analysis to profile novel real-world subphenotypes of Group 3 PH patients and to use clusters to explore heterogeneity of treatment response to pulmonary vasodilator therapy. In a large real-world cohort, our approach identified unique patient subgroups that were independent of underlying lung disease and distinct by virtue of unifying clinical profiles. In fact, CLD was not sufficient to differentiate groups, suggesting that the current characterisation of Group 3 patients by CLD alone may be limited. Subphenotypes in our analysis were associated with significantly disparate clinical trajectories, including death and respiratory failure, and in this way, were well positioned to elucidate differences in outcomes within groups treated with oral pulmonary vasodilators compared to untreated patients. In an exploratory analysis limited by the differences between treated and untreated patients inherent in a retrospective real-world cohort, we observed that treatment effect depended on cluster assignment and that treatment was associated with a greater clinical event rate compared to no treatment. Our findings identify opportunities to deconstruct clinical heterogeneity and nuance Group 3 PH phenotyping using agnostic analytical methods. Simultaneously, these results identify an opportunity to

TABLE 3 Primary and secondary end-points for cluster analysis-defined subgroups of Group 3 pulmonary hypertension patients treated with pulmonary arterial hypertension therapy

	Inclusive model		Haemodynamic model	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Death or organ-specific failure (respiratory, renal, right heart)				
Cluster 1			6.4 (3.3–12.6)	<0.001
Cluster 2	2.3 (1.5–3.6)	<0.001	3.1 (1.6–6.0)	0.001
Cluster 3	3.0 (1.9–4.8)	<0.001		
Death or respiratory failure				
Cluster 1			6.1 (3.2–11.8)	<0.001
Cluster 2	2.6 (1.7–4.0)	<0.001	3.1 (1.6–6.0)	0.001
Cluster 3	3.1 (1.9–4.9)	<0.001		

The inclusive model includes all treated and untreated patients (n=392); the haemodynamic model includes treated patients exposed to pulmonary arterial hypertension therapy with available right heart catheterisation data (n=112). For the inclusive model, cluster 1^{I} is the reference and for the haemodynamic model including treated patients who underwent right heart catheterisation, cluster 3^{H} is the reference. The reference was determined by the cluster with the most favourable outcome as per figure 1.

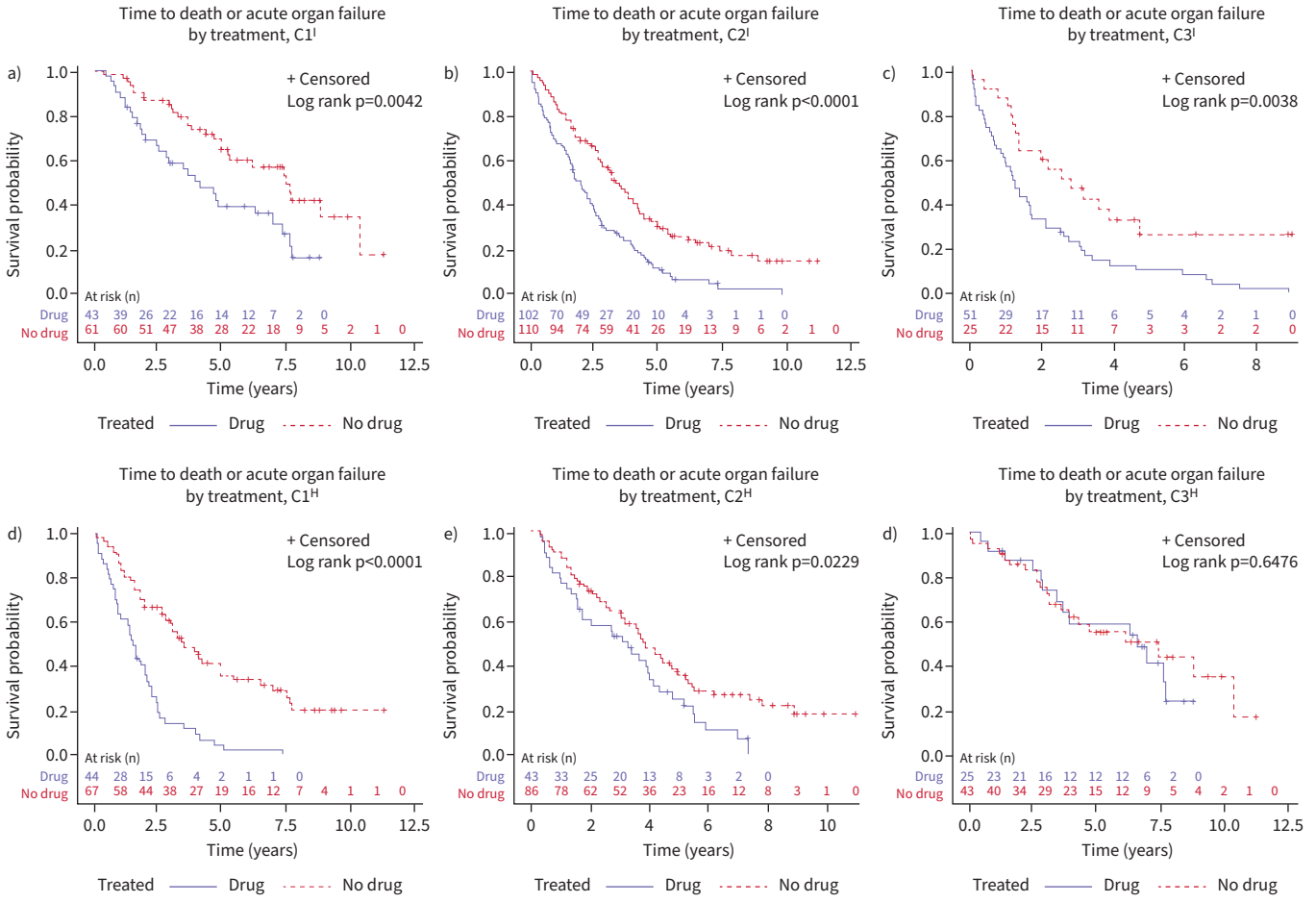

FIGURE 2 Survival curves by treatment within Group 3 pulmonary hypertension subphenotypes. a–c) The effect of treatment on the primary outcome of death or organ failure within inclusive model clusters ($C1^{I}$–$C3^{I}$). d–f) The effect of treatment on the same primary outcome within haemodynamic model clusters ($C1^{H}$–$C3^{H}$).

TABLE 4 Effect of pulmonary arterial hypertension therapy exposure on outcomes within cluster analysis-defined subgroups of Group 3 pulmonary hypertension patients

	Inclusive model		Haemodynamic model	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Death or organ-specific failure (respiratory–renal–right heart)[#]				
Cluster 1	2.1 (1.3–3.6)	0.005	3.2 (2.1–5.0)	<0.001
Cluster 2	2.0 (1.5–2.7)	<0.001	1.6 (1.1–2.5)	0.02
Cluster 3	2.2 (1.3–3.9)	0.005	1.2 (0.6–2.3)	0.65
Death or respiratory failure[¶]				
Cluster 1	1.9 (1.1–3.2)	0.025	3.2 (2.0–4.9)	<0.001
Cluster 2	2.1 (1.6–2.9)	<0.001	1.6 (1.1–2.5)	0.02
Cluster 3	2.1 (1.2–3.8)	0.009	1.1 (0.5–2.2)	0.81

The inclusive model includes all treated and untreated patients (n=392); the haemodynamic model includes treated patients exposed to pulmonary arterial hypertension therapy with available right heart catheterisation data (n=112). Treatment effects represent the effect of pulmonary vasodilator therapy within the treated clusters as compared to all untreated patients (n=196) *via* cluster matching. In a model with a cluster-by-treatment interaction, the interaction term was significant for the primary[#] and secondary[¶] outcomes (p=0.006 and p=0.01, respectively).

employ novel methods to identify subphenotypes of Group 3 PH that are more likely to benefit from, or experience harm when treated with, pulmonary vasodilator therapies.

The inclusive and haemodynamic models in this study generated clusters that were distinct, with variable and clinically meaningful profiles. The inclusive model generated clusters that incorporated the overlapping physiology that is characteristic of real-world patients. For example, patients in $C1^I$ were characterised by a profile at risk for concomitant Group 2 PH physiology whereas $C2^I$ had a profile suggestive of a combined pulmonary fibrosis and emphysema subphenotype with characteristically preserved lung volumes (FVC) and a reduced FEV_1/FVC compared to $C3^I$, where a diagnosis of IPF predominated [43]. The fact that outcomes remained discrete across clusters even after adjusting for underlying CLD is an important finding from our study. In the current PH guidelines, Group 3 disease is separated by CLD; however, pulmonary diagnosis alone may be insufficient to differentiate subphenotypes of Group 3 patients. Specifically, our findings suggest that a broader range of patient data should be considered to deconstruct subgroup heterogeneity. Importantly, in the inclusive model, echocardiographic features of PH including RVSP and dilation of the right heart chambers did not differ across clusters. In fact, the median RVSP was the same across $C1^I$–$C3^I$, reinforcing limitations of echocardiography alone for subphenotyping CLD populations [44, 45].

Inclusive model outcomes were most favourable for $C1^I$, with $C2^I$ and $C3^I$ experiencing a 2–3-fold increased risk of death and organ-specific failures, respectively. These outcomes in part reflect the underlying lung disease and overlapping comorbidities because treatment of sleep-disordered breathing may normalise pulmonary artery pressures [46, 47], guideline-directed heart failure therapies including diuretics may partially address post-capillary PH, and the mortality risk associated with IPF-PH is greater than that of COPD-PH [48]. When considering the effect of PAH therapy within these clusters, all subphenotypes were associated with a similar 2-fold increase in the risk for both the primary and secondary outcomes, compared to untreated patients. It is noted that in our sensitivity analysis of organ failure, with death as the competing event, we were still unable to identify a cluster in whom treatment was efficacious although the effect estimates were less. However, conclusions surrounding within-cluster treatment responses and recommendations for use of these therapies by cluster are limited by the retrospective design of the study and differences in select variables between treated and untreated patients, despite baseline and cluster matching. These findings expand on existing clinical trial data focused on short-term outcomes including 6-min walk distance (6MWD) and haemodynamic end-points, and mirror studies that have identified clinical worsening associated with treatment [9, 10].

The results of our haemodynamic model, restricted to patients with available RHC data (57% of treated patients) [4, 5], highlight the importance of invasive haemodynamic assessment for patients with Group 3 PH, especially when considering initiation of pulmonary vasodilators [4, 5]. The low frequency of RHC across our cohort prior to initiation of PAH therapy is consistent with previously established PH practice patterns [14, 39]. Hesitation surrounding haemodynamic assessment in Group 3 patients warrants ongoing

education and may be partly responsible for the heterogeneity of treatment effect seen across Group 3 PH trials when enrolment is determined by echocardiographic PH features [6, 7]. In this model, we identified three subphenotypes defined by haemodynamic and spirometric parameters, which definitively surpassed underlying lung disease in terms of variable importance to cluster assignment. These results align with studies in both COPD [49, 50] and interstitial lung disease [15, 33, 51], proposing a vascular lung disease phenotype characterised by preserved spirometry, significantly reduced D_{LCO}, hypoxia, oxygen dependence and a more severe haemodynamic profile [4, 5]. In a cohort of COPD patients referred for lung volume reduction surgery or transplant (50.2% of whom had PH confirmed by RHC), k-means cluster analysis similarly identified a subphenotype of patients with severe PH, moderate obstruction and hypoxaemia [50]. Therefore, it may be the case that cluster analysis is well positioned to advance knowledge of the clinical profiles hidden within the spectrum across groups commonly encountered in real-world practice, currently defined as Group 1/3 PH overlap or Group 2/3 PH overlap.

Recent studies have investigated whether haemodynamic parameters may predict treatment response in Group 3 PH. Concordant with these retrospective data, in our haemodynamic model we did not identify a cluster in which treatment was associated with a mortality benefit or reduced risk of organ failure. However, we did find that treatment effect depended on cluster assignment. This may be due to fundamental differences between our treated and untreated patients and the retrospective nature of our study in that treated patients often had more severe PH and/or lung disease. Additionally, inclusion of measures of exercise tolerance (*e.g.* 6MWD) may have strengthened the validity of our findings, but were not available for inclusion in the cluster models. In a retrospective study of COPD-PH patients, Vizza *et al.* [12] identified baseline WHO functional class and 6MWD as predictors of treatment response whereas haemodynamic profiles did not differ between responders and nonresponders. Similarly, in a recent single-centre cohort study of patients with severe Group 3 PH, PVR did not predict treatment response, defined by a 30-m improvement in 6MWD [52]. While recent guidelines allow for consideration of PAH therapy in patients with lung disease and PVR >5 WU, cumulatively, our results and those previously published highlight the importance of prospective studies, particularly within subgroups enriched for treatment response, possibly with multicomponent end-points [53, 54]. Finally, our study highlights the importance of end-points that capture clinical deterioration, including morbidity and mortality, and is the first to examine long-term treatment effects within Group 3 PH subgroups [31, 54].

Strengths and limitations

To our knowledge, this is the first study to assess the impact of treatment in Group 3 PH using k-means cluster analysis, which is a methodology that has proven effective for identifying novel patient subgroups in PAH populations [20, 22, 23]. In this study, patients were followed for several years, expanding our appreciation of treatment consequences beyond the narrower timeframe typical of rigorously controlled clinical trials. However, key variables that are used to define treatment response in clinical practice (*e.g.* 6MWD and WHO functional class) were not available in this analysis, and missingness of variables in the analytical models may have confounded our results. Future studies incorporating these important variables will expand our understanding of the utility of k-means cluster analysis for Group 3 PH phenotyping. In fact, findings from our k-means analysis delineating Group 3 PH patient subgroups effectively show proof of concept for the use of this method when considering approaches that aim to refine enrolment criteria in prospective clinical trials; expansion to validation cohorts, including those within specific CLD populations, will be important prior to pursing this application. As with all retrospective methods, confounding by indication is a possible source of bias. Specifically, sicker patients may have been more likely to receive treatment, which would limit our ability to detect within-cluster treatment effects. Therefore, application of similar methodologies to randomised clinical trial populations with lack of bias regarding treatment decisions may be better positioned to expose succinct clinical profiles that inform treatment response.

In conclusion, cluster analysis is an effective method to identify Group 3 PH patients with similar but previously unrecognised traits, clinical trajectories and, possibly, treatment responses. Despite limitations associated with the retrospective study design and data missingness, our results demonstrate the potential for our methodology to deconstruct clinical heterogeneity and enhance subphenotyping among real-world patients, often with overlapping World Symposium on Pulmonary Hypertension Group features. Furthermore, when applied to well-matched cohorts, we believe our work demonstrates proof of concept that clustering may be a tool to aid investigations to identify Group 3 PH patients who stand to benefit from, or experience harm when treated with, pulmonary vasodilator therapies.

Provenance: Submitted article, peer reviewed.

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the US Government.

Conflict of interest: S.W. Johnson, R-S. Wang, M.R. Winter, K.R. Gillmeyer, R.H. Goldstein and R.S. Wiener have nothing to disclose; K. Zeder has received a Max Kade Fellowship from the Max Kade Foundation NY, and support from MSD, Ferrer and Janssen (outside the scope of the current work); E.S. Klings has received support from CSL Behring, Pfizer, Novartis, Novo Nordisk and United Therapeutics (all outside of the scope of the current work); B.A. Maron reports support from Actelion Pharmaceuticals, Deerfield Company and Tenax Therapeutics (all outside the scope of the current work).

Support statement: This work was supported by the National Institutes of Health (1F32HL164090-01 to S.W. Johnson, and R01HL139613-01, R01HL163960, R01HL153502 and R01HL155096-01 to B.A. Maron) and by resources from the VA Boston Healthcare System. Support for VA/CMS data was provided by the Department of Veterans Affairs, VA Health Services Research and Development Service, VA Information Resource Center (Project Numbers SDR 02-237 and 98-004). Funding information for this article has been deposited with the Crossref Funder Registry.

Ethics statement: This research was approved by the Boston VA Healthcare System Institutional Review Board of the VA Boston Healthcare System R&D committee (project number 1590467-7).

References

1 Hemnes AR, Beck GJ, Newman JH, *et al.* PVDOMICS: a multi-center study to improve understanding of pulmonary vascular disease through phenomics. *Circ Res* 2017; 121: 1136–1139.
2 Zeder K, Avian A, Bachmaier G, *et al.* Elevated pulmonary vascular resistance predicts mortality in COPD patients. *Eur Respir J* 2021; 58: 2100944.
3 Olsson KM, Hoeper MM, Pausch C, *et al.* Pulmonary vascular resistance predicts mortality in patients with pulmonary hypertension associated with interstitial lung disease: results from the COMPERA registry. *Eur Respir J* 2021; 58: 2101483.
4 Humbert M, Kovacs G, Hoeper MM, *et al.* 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J* 2022; 43: 3618–3731.
5 Humbert M, Kovacs G, Hoeper MM, *et al.* 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J* 2023; 61: 2200879.
6 Blanco I, Santos S, Gea J, *et al.* Sildenafil to improve respiratory rehabilitation outcomes in COPD: a controlled trial. *Eur Respir J* 2013; 42: 982–992.
7 Vitulo P, Stanziola A, Confalonieri M, *et al.* Sildenafil in severe pulmonary hypertension associated with chronic obstructive pulmonary disease: a randomized controlled multicenter clinical trial. *J Heart Lung Transplant* 2017; 36: 166–174.
8 Maron BA, Choudhary G, Goldstein RL, *et al.* Tadalafil for veterans with chronic obstructive pulmonary disease-pulmonary hypertension: a multicenter, placebo-controlled randomized trial. *Pulm Circ* 2022; 12: e12043.
9 Nathan SD, Behr J, Collard HR, *et al.* Riociguat for idiopathic interstitial pneumonia-associated pulmonary hypertension (RISE-IIP): a randomised, placebo-controlled phase 2b study. *Lancet Respir Med* 2019; 7: 780–790.
10 Raghu G, Behr J, Brown KK, *et al.* Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. *Ann Intern Med* 2013; 158: 641–649.
11 Prins KW, Duval S, Markowitz J, *et al.* Chronic use of PAH-specific therapy in World Health Organization Group III pulmonary hypertension: a systematic review and meta-analysis. *Pulm Circ* 2017; 7: 145–155.
12 Vizza CD, Hoeper MM, Huscher D, *et al.* Pulmonary hypertension in patients with COPD: results from the Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA). *Chest* 2021; 160: 678–689.
13 Trammell AW, Pugh ME, Newman JH, *et al.* Use of pulmonary arterial hypertension-approved therapy in the treatment of non-Group 1 pulmonary hypertension at US referral centers. *Pulm Circ* 2015; 5: 356–363.
14 Minai OA, Nathan SD, Hill NS, *et al.* Pulmonary hypertension in lung diseases: survey of beliefs and practice patterns. *Respir Med* 2010; 104: 741–748.
15 Nathan SD. Progress in the treatment of pulmonary hypertension associated with interstitial lung disease. *Am J Respir Crit Care Med* 2023; 208: 238–246.
16 Borek I, Birnhuber A, Voelkel NF, *et al.* The vascular perspective on acute and chronic lung disease. *J Clin Invest* 2023; 133: e170502.
17 Singh N, Dorfmuller P, Shlobin OA, *et al.* Group 3 pulmonary hypertension: from bench to bedside. *Circ Res* 2022; 130: 1404–1422.

18 Gillmeyer KR, Rinne ST, Walkey AJ, *et al.* How closely do clinical trial participants resemble "real-world" patients with Groups 2 and 3 pulmonary hypertension? A structured review. *Ann Am Thorac Soc* 2020; 17: 779–783.
19 Semler MW, Bernard GR, Aaron SD, *et al.* Identifying clinical research priorities in adult pulmonary and critical care. NHLBI Working Group Report. *Am J Respir Crit Care Med* 2020; 202: 511–523.
20 Parikh KS, Rao Y, Ahmad T, *et al.* Novel approach to classifying patients with pulmonary arterial hypertension using cluster analysis. *Pulm Circ* 2017; 7: 486–493.
21 Badagliacca R, Rischard F, Papa S, *et al.* Clinical implications of idiopathic pulmonary arterial hypertension phenotypes defined by cluster analysis. *J Heart Lung Transplant* 2020; 39: 310–320.
22 Hoeper MM, Pausch C, Grunig E, *et al.* Idiopathic pulmonary arterial hypertension phenotypes determined by cluster analysis from the COMPERA registry. *J Heart Lung Transplant* 2020; 39: 1435–1444.
23 Launay D, Montani D, Hassoun PM, *et al.* Clinical phenotypes and survival of pre-capillary pulmonary hypertension in systemic sclerosis. *PLoS One* 2018; 13: e0197112.
24 James G, Witten D, Hastie T, *et al.* An introduction to statistical learning: with applications in R. New York, Springer, 2013.
25 Hoeper MM, Dwivedi K, Pausch C, *et al.* Phenotyping of idiopathic pulmonary arterial hypertension: a registry analysis. *Lancet Respir Med* 2022; 10: 937–948.
26 Maron BA, Hess E, Maddox TM, *et al.* Association of borderline pulmonary hypertension with mortality and hospitalization in a large patient cohort: insights from the Veterans Affairs clinical assessment, reporting, and tracking program. *Circulation* 2016; 133: 1240–1248.
27 Maron BA, Brittain EL, Hess E, *et al.* Pulmonary vascular resistance and clinical outcomes in patients with pulmonary hypertension: a retrospective cohort study. *Lancet Respir Med* 2020; 8: 873–884.
28 Gillmeyer KR, Johnson SW, Bolton RE, *et al.* Organization of pulmonary hypertension care in non-expert care settings: lessons learned from a multi-site study. *Health Serv Res* 2023; 58: 663–673.
29 Gillmeyer KR, Rinne ST, Glickman ME, *et al.* Factors associated with potentially inappropriate phosphodiesterase-5 inhibitor use for pulmonary hypertension in the United States, 2006 to 2015. *Circ Cardiovasc Qual Outcomes* 2020; 13: e005993.
30 Kim D, Lee KM, Freiman MR, *et al.* Phosphodiesterase-5 inhibitor therapy for pulmonary hypertension in the United States. Actual versus recommended use. *Ann Am Thorac Soc* 2018; 15: 693–701.
31 Gillmeyer KR, Miller DR, Glickman ME, *et al.* Outcomes of pulmonary vasodilator use in veterans with pulmonary hypertension associated with left heart disease and lung disease. *Pulm Circ* 2021; 11: 20458940211001714.
32 Johnson SW, Gillmeyer KR, Bolton RE, *et al.* Variable monitoring of veterans with Group 3 pulmonary hypertension treated with off-label pulmonary vasodilator therapy. *Ann Am Thorac Soc* 2022; 19: 1236–1239.
33 Nathan SD, Barbera JA, Gaine SP, *et al.* Pulmonary hypertension in chronic lung disease and hypoxia. *Eur Respir J* 2019; 53: 1801914.
34 Harris PA, Taylor R, Minor BL, *et al.* The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019; 95: 103208.
35 Walsh SL, Calandriello L, Sverzellati N, *et al.* Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016; 71: 45–51.
36 Oldham WM, Oliveira RKF, Wang RS, *et al.* Network analysis to risk stratify patients with exercise intolerance. *Circ Res* 2018; 122: 864–876.
37 Seeger W, Adir Y, Barbera JA, *et al.* Pulmonary hypertension in chronic lung diseases. *J Am Coll Cardiol* 2013; 62: Suppl. 25, D109–D116.
38 Hartigan JA, Wong MA. A k-means clustering algorithm. *J R Stat Soc Series C* 1979; 28: 100–108.
39 Maron BA, Choudhary G, Khan UA, *et al.* Clinical profile and underdiagnosis of pulmonary hypertension in US veteran patients. *Circ Heart Fail* 2013; 6: 906–912.
40 Corraini P, Olsen M, Pedersen L, *et al.* Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. *Clin Epidemiol* 2017; 9: 331–338.
41 Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation* 2016; 133: 601–609.
42 Lawson CA, Zaccardi F, Squire I, *et al.* Risk factors for heart failure: 20-year population-based trends by sex, socioeconomic status, and ethnicity. *Circ Heart Fail* 2020; 13: e006472.
43 Cottin V, Selman M, Inoue Y, *et al.* Syndrome of combined pulmonary fibrosis and emphysema: an official ATS/ERS/JRS/ALAT research statement. *Am J Respir Crit Care Med* 2022; 206: e7–e41.
44 Fisher MR, Criner GJ, Fishman AP, *et al.* Estimating pulmonary artery pressures by echocardiography in patients with emphysema. *Eur Respir J* 2007; 30: 914–921.
45 Nathan SD, Shlobin OA, Barnett SD, *et al.* Right ventricular systolic pressure by echocardiography as a predictor of pulmonary hypertension in idiopathic pulmonary fibrosis. *Respir Med* 2008; 102: 1305–1310.
46 Sajkov D, Wang T, Saunders NA, *et al.* Continuous positive airway pressure treatment improves pulmonary hemodynamics in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 2002; 165: 152–158.

47 Ip MS, Lam B, Chan LY, *et al.* Circulating nitric oxide is suppressed in obstructive sleep apnea and is reversed by nasal continuous positive airway pressure. *Am J Respir Crit Care Med* 2000; 162: 2166–2171.

48 Gall H, Felix JF, Schneck FK, *et al.* The Giessen Pulmonary Hypertension Registry: survival in pulmonary hypertension subgroups. *J Heart Lung Transplant* 2017; 36: 957–967.

49 Kovacs G, Agusti A, Barbera JA, *et al.* Pulmonary vascular involvement in chronic obstructive pulmonary disease. Is there a pulmonary vascular phenotype? *Am J Respir Crit Care Med* 2018; 198: 1000–1011.

50 Thabut G, Dauriat G, Stern JB, *et al.* Pulmonary hemodynamics in advanced COPD candidates for lung volume reduction surgery or lung transplantation. *Chest* 2005; 127: 1531–1536.

51 Lettieri CJ, Nathan SD, Barnett SD, *et al.* Prevalence and outcomes of pulmonary arterial hypertension in advanced idiopathic pulmonary fibrosis. *Chest* 2006; 129: 746–752.

52 Naud R, Bermudez J, Resseguier N, *et al.* Impact of targeted pulmonary arterial hypertension therapies in severe pulmonary hypertension in chronic lung diseases. *ERJ Open Res* 2023; 9: 00027-2023.

53 Nathan SD, Fernandes P, Psotka M, *et al.* Pulmonary hypertension in interstitial lung disease: clinical trial design and endpoints: a consensus statement from the Pulmonary Vascular Research Institute's Innovative Drug Development Initiative–Group 3 Pulmonary Hypertension. *Pulm Circ* 2022; 12: e12178.

54 Weatherald J, Boucly A, Peters A, *et al.* The evolving landscape of pulmonary arterial hypertension clinical trials. *Lancet* 2022; 400: 1884–1898.