

Novel approach to classifying patients with pulmonary arterial hypertension using cluster analysis

Kishan S. Parikh^{1,2}, Youlan Rao³, Tariq Ahmad⁴, Kai Shen³, G. Michael Felker^{1,2} and Sudarshan Rajagopal²

¹Duke Clinical Research Institute, Durham, NC, USA; ²Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC, USA; ³United Therapeutics Corporation, Research Triangle Park, NC, USA; ⁴Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

Abstract

Pulmonary arterial hypertension (PAH) patients have distinct disease courses and responses to treatment, but current diagnostic and treatment schemes provide limited insight. We aimed to see if cluster analysis could distinguish clinical phenotypes in PAH. An unbiased cluster analysis was performed on 17 baseline clinical variables of PAH patients from the FREEDOM-M, FREEDOM-C, and FREEDOM-C2 randomized trials of oral treprostinil versus placebo. Participants were either treatment-naïve (FREEDOM-M) or on background therapy (FREEDOM-C, FREEDOM-C2). We tested for association of clusters with outcomes and interaction with respect to treatment. Primary outcome was 6-minute walking distance (6MWD) change. We included 966 participants with 12-week (FREEDOM-M) or 16-week (FREEDOM-C and FREEDOM-C2) follow-up. Four patient clusters were identified. Compared with Clusters 1 (n = 131) and 2 (n = 496), Clusters 3 (n = 246) and 4 (n = 93) patients were older, heavier, had worse baseline functional class, 6MWD, Borg Dyspnea Index, and fewer years since PAH diagnosis. Clusters also differed by PAH etiology and background therapies, but not gender or race. Mean treatment effect of oral treprostinil differed across Clusters 1–4 increased in a monotonic fashion (Cluster 1: 10.9 m; Cluster 2: 13.0 m; Cluster 3: 25.0 m; Cluster 4: 50.9 m; interaction *P* value = 0.048). We identified four distinct clusters of PAH patients based on common patient characteristics. Patients who were older, diagnosed with PAH for a shorter period, and had worse baseline symptoms and exercise capacity had the greatest response to oral treprostinil treatment.

Keywords

cluster analysis, pulmonary arterial hypertension, treprostinil

Date received: 18 January 2017; accepted: 30 March 2017

Pulmonary Circulation 2017; 7(2) 486–493

DOI: 10.1177/2045893217705891

Pulmonary arterial hypertension (PAH) is a heterogeneous disease affecting a range of patient populations,¹ but clinical characterization of the disease has lagged behind development of therapies. The widely used World Health Organization (WHO) classification scheme is based on disease etiology and does not capture dynamic changes including disease progression and response to treatment. In addition, gender, cardiovascular risk factors and age have recently been identified as potential predictors of therapeutic response.^{2–5}

Clustering, also referred to as phenomapping, is a statistical technique that utilizes unbiased machine learning and has been recently applied as a proof-of-concept strategy for classifying patients with heart failure.^{6–8} We aimed to use cluster analysis on PAH participants included in clinical

Corresponding author:

Kishan S. Parikh, Duke Clinical Research Institute, 2400 Pratt Street, Durham, NC 27705, USA.

Email: Kishan.pariikh@dm.duke.edu



Creative Commons Non Commercial CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>)

which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

© The Author(s) 2017.

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
journals.sagepub.com/home/pul



trials to identify distinct clinical phenotypes in PAH. We selected three recent trials assessing response to the same intervention, oral treprostinil. The trials had the same endpoint, change in 6-minute walk distance (6MWD), allowing us to assess whether change in 6MWD would differ across clinical phenotypes. The trials were FREEDOM-M (oral treprostinil monotherapy: median placebo-corrected 6MWD change, +23 m), FREEDOM-C, and FREEDOM-C2 (oral treprostinil in combination therapy, no significant placebo-corrected 6MWD change).^{9–11}

Methods

Study design

FREEDOM-M (NCT00325403), FREEDOM-C (NCT00325442), and FREEDOM-C2 (NCT00887978) were multicenter, randomized trials of oral treprostinil as mono- or part of combination therapy in patients with PAH. Participants from FREEDOM-M (n = 349) were randomized to oral treprostinil versus placebo as initial PAH treatment in a 2:1 fashion with 12-week follow-up. The FREEDOM-C (n = 350) and FREEDOM-C2 (n = 310) trials randomized participants to oral treprostinil versus placebo in a 1:1 ratio as part of combination therapy for total duration of 16 weeks.

Patient population

Specific eligibility criteria have been previously published.^{9–11} As an overview, participants with idiopathic or hereditary PAH, drug-associated PAH, PAH associated with repaired congenital shunts, or PAH associated with collagen vascular disease or Human Immunodeficiency Virus (HIV) were recruited. Baseline 6MWD was required to be in the range of 100–450 m for FREEDOM-M and FREEDOM-C, and 150–425 m for FREEDOM-C2.

Additional inclusion criteria for FREEDOM-M were de novo PAH diagnosis and no concurrent background therapy. In contrast, FREEDOM-C and FREEDOM-C2 only included patients taking an approved phosphodiesterase type-5 (PDE-5) inhibitor or endothelin receptor antagonist (ERA) therapy, or combination of both, for at least 90 days and on a stable dose for at least 30 days before study entry. Participants with any evidence of significant left-sided heart disease or parenchymal lung disease were excluded from all studies. Of note, a significant proportion of participants were recruited from Chinese and Indian PAH centers, allowing for study of racial differences.

As is necessary for cluster analysis, patients with missing data for any of the cluster variables were excluded.

Cluster variables

Individual variables that were included for clustering of participants were ascertained at baseline and encompassed

clinical characteristics, laboratory data, background medications, and functional capacity. First, the selected variables were clustered to map out all interrelationships. Binary divisive clustering was performed using the PROC VARCLUS function in SAS software (SAS Institute, Cary, NC, USA) for continuous variables. All participants were initially included in one cluster at the beginning of our cluster analysis. In each divisive iteration, a principal component analysis (PCA) was performed on the set of variables within the cluster. The PCA, a procedure used to create predictive models, converts possibly correlated variables via orthogonal transformation into a set of values of linearly uncorrelated variables called principal components (PCs). The first two PCs, which had the highest variance, were identified, and the cluster was split into two corresponding sub-clusters according to the correlation between each variable and each PC. This divisive iteration process was continued until only one PC was found in the cluster (i.e., when the second eigenvalue ≤ 1). Each divisive iteration was also coupled to a test procedure in which each variable was tested to see if assigning it to a different cluster increased the amount of variance it explained. If it did, the variable was reassigned to that cluster and the components of the two clusters involved were recomputed before the next variable was tested. For categorical variables, PCA and PROC VARCLUS were not applicable. Categorical variables with k levels were therefore represented by k-1 dummy variables.

Our variable grouping with PCA resulted in five clusters for continuous variables (Cluster 1: age, weight, creatinine; Cluster 2: systolic blood pressure, diastolic blood pressure; Cluster 3: aspartate aminotransferase (AST), bilirubin; Cluster 4: years since PAH diagnosis, heart rate; Cluster 5: baseline 6MWD, sodium, baseline Borg Dyspnea Index) and two clusters for categorical variables (Cluster 1: ethnicity, functional class, background therapy; Cluster 2: gender, disease etiology).

Participant clustering

The goal of cluster analysis is to group participants who are more closely related to each other than to other participants. The participant clustering process begins with every individual as a separate cluster. We used PCA output of variable clusters to group participants based on standardized first PC score values derived from each of the seven variable clusters. We used the PROC CLUSTER function to accomplish agglomerative clustering by merging the two closest clusters based on weighted Euclidean distance. Clustering based on this distance is in accordance to Ward's minimum variance method.

The clustering process is continued until only a single cluster remains that includes all participants. To choose an optimal cluster number, we looked for the stage in the clustering process that provided greatest consensus among three statistics: the local peaks of the cubic clustering criterion and pseudo F statistic, combined with a small value of

the pseudo t^2 statistic next to a larger pseudo t^2 value for the next cluster fusion.

Clinical endpoints

The primary endpoint for all three trials was change in 6MWD from baseline to end of study follow-up (FREEDOM-M: 12 weeks; FREEDOM-C and FREEDOM-C2: 16 weeks). Secondary endpoints of interest included clinical worsening, Borg Dyspnea Index and WHO functional class (FC).

Statistical analysis

Association between identified participant clusters and clinical outcomes was tested using one-way ANOVA for change

in 6MWD, and logistic regression for change in functional class and change in Borg Dyspnea Index. Changes were the differences between baseline and end-of-study values. We also assessed for interaction between treatment (treprostinil versus placebo) and outcomes with respect to cluster membership. Interactions for 6MWD change were assessed with two-way ANOVA, and for Borg Dyspnea Index and WHO-FC with logistic regression. To understand trends across clusters, linear trend test was used for continuous variables and the Cochran-Armitage test was used for categorical variables. Continuous variables were expressed as mean \pm standard deviation, and categorical variables were expressed as n (%). Normality of the 6MWD data was assessed by the histograms of residuals and quantile-quantile plots in all participants and treprostinil- and placebo-treated subgroups (Supplemental Figure). 6MWD

Table 1. Baseline characteristics of combined participants from FREEDOM-M, FREEDOM-C, and FREEDOM-C2 trials.*

Variable	All (n = 966)	Treprostinil (n = 541)	Placebo (n = 425)	P value
Age (years)	47.3 \pm 14.6	46.8 \pm 14.8	47.9 \pm 14.2	0.31
Weight (kg)	73.5 \pm 20.7	72.3 \pm 20.0	75.1 \pm 21.4	0.07
Female	77.8	77.3	78.6	0.62
Ethnicity				0.33
White	64.2	62.5	66.4	
Asian	26.6	28.5	24.2	
Other	9.2	9.1	9.4	
Years since PAH [†] diagnosis	2.6 \pm 3.9	2.3 \pm 3.5	3.1 \pm 4.2	<0.001
Etiology of PAH				0.88
Idiopathic or hereditary	68.8	68.9	68.7	
Collagen vascular disease	25.2	24.8	25.6	
Others	6.0	6.3	5.6	
Background PAH therapy				<0.001
None	34.5	41.2	25.9	
ERA only	15.6	14.0	17.6	
PDE-5 inhibitor only	22.0	19.8	24.9	
ERA and PDE-5 inhibitor	27.8	25.0	31.5	
Heart rate (bpm)	79.7 \pm 14.4	79.6 \pm 14.8	79.7 \pm 13.8	0.66
Systolic blood pressure (mmHg)	114 \pm 15.2	113 \pm 15.2	115 \pm 15.2	0.09
Diastolic blood pressure (mmHg)	71.2 \pm 9.4	71.3 \pm 9.6	71.1 \pm 9.2	0.64
Laboratory values				
Sodium (mmol/L)	140 \pm 3.4	140 \pm 3.3	140 \pm 3.4	0.80
Creatinine (mg/dL)	1.0 \pm 0.29	1.0 \pm 0.29	1.0 \pm 0.30	0.63
AST (IU/L)	27.8 \pm 15.8	28.3 \pm 17.7	27.2 \pm 13.1	0.10
Total bilirubin (mg/dL)	0.8 \pm 0.6	0.8 \pm 0.6	0.7 \pm 0.6	0.33
6MWD (m)	338 \pm 71.3	337 \pm 70.7	338 \pm 72.1	0.71
Borg Dyspnea Index	3.9 \pm 2.3	3.8 \pm 2.3	4.0 \pm 2.3	0.16
WHO-FC III or IV	70.7	67.8	74.4	0.03

*Variables shown as mean \pm standard deviation or %. P values are for comparison of oral treprostinil vs. placebo groups. 6MWD, 6-minute walk distance; AST, aspartate aminotransferase; ERA, endothelin receptor antagonist; IU, International units; PDE-5, phosphodiesterase type-5; PAH, pulmonary arterial hypertension; WHO-FC, World Health Organization functional class.

change was also presented as median (Q1, Q3) for comparison to individual study results, and we tested this data for non-parametric significance with the Kruskal–Wallis test. A P value < 0.05 was considered statistically significant for all analyses.

Results

Of 1009 patients who participated in the FREEDOM-M, FREEDOM-C, and FREEDOM-C2 trials, 966 (96%) had complete baseline data available for the pre-specified 17 clinical variables of interest and were included in our study. Within the combined group, participants receiving treprostinil compared with those receiving placebo had similar demographics, distribution of PAH etiology, baseline 6MWD, and Borg Dyspnea Index ($P > 0.06$ for all), but had better WHO-FC ($P = 0.03$) and fewer years since diagnosis of PAH ($P < 0.001$) (Table 1). The cluster analysis identified four groups of patients: Cluster 1 ($n = 131$); Cluster 2 ($n = 496$); Cluster 3 ($n = 246$); and Cluster 4 ($n = 93$). Cluster dendrogram showing standardized Euclidean distances confirmed four distinct cluster numbers, and is shown with corresponding criteria for optimal number of clusters in Fig. 1. Key characteristics compared across patient clusters are described below.

Comparison of baseline data among clusters

When compared with other clusters, patients in Cluster 1 had were younger, second lightest, more likely to be non-White/non-Asian, men, have idiopathic or hereditary PAH, and had longer time since PAH diagnosis (Table 2). Cluster 2 was the largest cluster, containing 51% of the overall cohort. This cluster was second youngest, lightest, most likely to be Asian, and be on background therapy. Cluster 3 was the second largest group. It was the second oldest and heaviest, had the highest proportion of women, most likely to have no background PAH therapy, and had the highest heart rate. Patients in Cluster 4 were the oldest, heaviest, had the largest proportion of Whites, and had the least amount of time since PAH diagnosis. Almost all patients in Cluster 4 (98.9%) had idiopathic, hereditary, or collagen vascular disease-associated PAH. This group also had heart rate higher than Clusters 1 and 2, and similar to Cluster 3.

Baseline functional capacity and dyspnea severity were assessed by 6MWD, Borg Dyspnea Index and proportion of patients identified as having WHO-FC III or IV symptoms. All three measures tracked together across clusters from best functional capacity (highest 6MWD, lowest Borg Dyspnea Index, and lowest proportion with WHO-FC III/IV symptoms) to worst functional capacity (lowest 6MWD, highest Borg Dyspnea Index, and highest proportion with WHO-FC III/IV symptoms) in the following order: Cluster 1 $>$ Cluster 2 $>$ Cluster 3 $>$ Cluster 4 ($P \leq 0.001$ for trend in all measures).

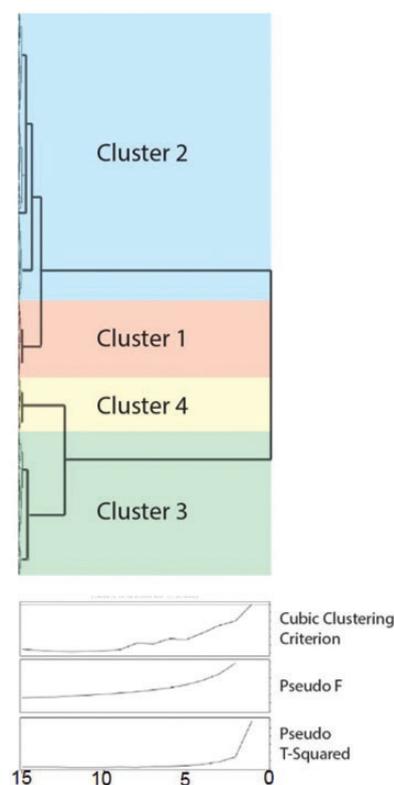


Fig. 1 Cluster dendrogram (top) and analysis to identify the optimal cluster number (bottom). The arrangement of the branches in the dendrogram represents relative degree of similarity between participants (i.e., Euclidean distance). The greater the height of the branch points, the greater the differences between the branches. The cubic clustering criterion (CCC), pseudo-F statistic, and pseudo T-squared statistic suggest four clusters can be created from the study population.

Follow-up

Follow-up 6MWD was available for 88.3% of patients (Cluster 1 = 90.1%; Cluster 2 = 91.9%; Cluster 3 = 82.5%; Cluster 4 = 81.7%). Among all patients, mean change in 6MWD did not vary by cluster ($P = 0.425$, Table 3), but there was a trend in association between cluster membership and mean 6MWD change in the subgroup of patients receiving oral treprostinil ($P = 0.075$). We next assessed treatment effect (i.e., difference in the 6MWD change between participants on oral treprostinil versus placebo) across clusters and observed a statistically significant interaction (mean treatment effect for Clusters 1–4, respectively: 10.9 m; 13.0 m; 25.0 m; 50.9 m; interaction P value = 0.048) (Fig. 2). Non-parametric analysis of all participants and treatment subgroups confirmed similar relationships between cluster grouping and 6MWD change (Supplemental Table). The percentage of participants who stopped study drug prematurely also varied by cluster (Cluster 1 = 13.7%; Cluster 2 = 12.5%; Cluster 3 = 22.4%; Cluster 4 = 25.8%; $P < 0.001$).

Table 2. Baseline patient characteristics according to patient clusters.*

Variable	Cluster 1 (n = 131)	Cluster 2 (n = 496)	Cluster 3 (n = 246)	Cluster 4 (n = 93)	P value
Age (years)	42.5 ± 14.1	45.9 ± 13.9	50.0 ± 14.8	54.8 ± 13.9	<0.001
Weight (kg)	72.7 ± 20.1	72.6 ± 19.8	73.6 ± 21.6	79.5 ± 22.1	0.027
Female	74.0	77.2	80.9	78.5	0.465
Ethnicity					0.2686
White	66.4	62.5	64.6	68.8	
Asian	19.8	28.8	27.2	22.6	
Other	13.7	8.7	8.1	8.6	
Years since PAH diagnosis	3.1 ± 4.2	2.9 ± 4.1	2.2 ± 3.6	1.5 ± 1.8	<0.001
Etiology of PAH					0.0019
Idiopathic or hereditary	74.8	68.8	66.3	67.7	
Collagen vascular disease	15.3	24.2	30.1	31.2	
Others	9.9	7.1	3.7	1.1	
Background PAH therapy					0.0421
None	36.6	29.4	42.7	36.6	
ERA only	17.6	17.3	11.4	15.1	
PDE-5 inhibitor only	16.8	24.8	20.3	19.4	
ERA and PDE-5 inhibitor	29.0	28.4	25.6	29.0	
Heart rate (bpm)	77.6 ± 13.3	77.7 ± 14.0	83.6 ± 14.9	82.2 ± 14.2	<0.001
Systolic blood pressure (mmHg)	114 ± 14.6	112 ± 14.6	116 ± 15.9	115 ± 16.9	0.020
Diastolic blood pressure (mmHg)	70.9 ± 9.5	70.5 ± 9.1	72.5 ± 9.4	71.6 ± 10.4	0.095
Laboratory values					
Sodium (mmol/L)	141 ± 3.1	140 ± 3.2	140 ± 3.4	139 ± 4.3	0.046
Creatinine (mg/dL)	0.9 ± 0.3	0.9 ± 0.3	1.0 ± 0.3	1.1 ± 0.4	<0.001
AST (IU/L)	25.4 ± 8.5	28.4 ± 17.7	27.7 ± 16.0	29.0 ± 12.0	0.123
Total bilirubin (mg/dL)	0.7 ± 0.5	0.7 ± 0.6	0.8 ± 0.8	0.8 ± 0.7	0.281
Exercise tolerance					
6MWD (m)	424 ± 15.8	370 ± 23.1	284 ± 29.4	187 ± 26.0	<0.001
Borg Dyspnea Index	3.1 ± 2.2	3.4 ± 2.0	4.6 ± 2.3	5.6 ± 2.4	<0.001
WHO-FC III or IV	55.0	67.1	79.6	88.2	<0.0001

*Variables shown as mean ± standard deviation or % unless stated otherwise. P values are for comparisons of variables across clusters.

6MWD, 6-minute walk distance; AST, aspartate aminotransferase; ERA, endothelin receptor antagonist; IU, International units; PDE-5, phosphodiesterase type-5; PAH, pulmonary arterial hypertension; WHO-FC, World Health Organization functional class.

Table 3. Mean change (SD) in 6MWD (m) from baseline by cluster.

Group	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P value
Overall (n = 853)	18.7 ± 53.3 (9.1, 28.3)	21.6 ± 55.6 (16.5, 26.7)	28.7 ± 63.5 (20.0, 37.4)	24.7 ± 75.7 (7.7, 41.7)	0.425
Placebo (n = 390)	13.4 ± 48.4 (1.2, 25.6)	14.4 ± 47.9 (7.8, 21.0)	14.9 ± 63.0 (2.0, 27.8)	-1.4 ± 49.9 (-17.5, 14.7)	0.378
Treprostinil (n = 463)	24.3 ± 57.9 (9.4, 39.2)	27.4 ± 60.5 (20.0, 34.8)	39.9 ± 61.2 (28.6, 51.2)	49.5 ± 87.5 (22.0, 77.0)	0.075

Discussion

We used an unbiased approach to identify four phenotypically different clusters from a cohort of 966 PAH patients

enrolled in clinical trials. We then looked for change in 6MWD in each cluster. The main findings of our study are the following: (1) cluster analysis can identify distinct profiles within a heterogeneous PAH cohort; (2) patients

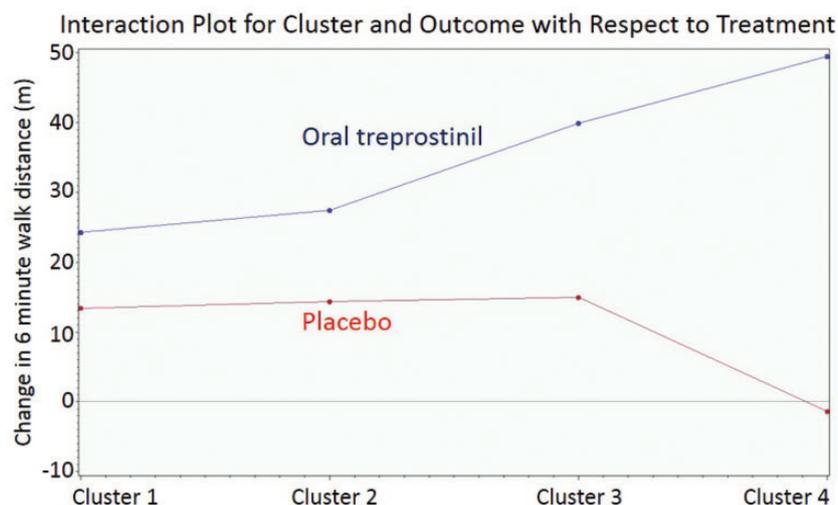


Fig. 2 Interaction plot for cluster and mean change in 6MWD with respect to treatment. Treatment effect increased monotonically with cluster number, from Clusters 1 to 4.

in the clusters have different functional capacity and symptoms; and (3) predicted change in 6MWD may differ based on the participant's clinical phenotype. Our study highlights clinical variation among participants enrolled in PAH trials, a consideration that has implications for future clinical trial design and understanding what constitutes optimal therapy in patients with PAH.

To our knowledge this is the first published utilization of machine learning to phenotype patients with PAH. Cluster analysis has also been successfully applied to both acute and chronic HF patients in recent years.^{6–8} Like HF, a heterogeneous syndrome, PAH can vary by etiology, co-morbidities, and endotypes, resulting in a spectrum of limitations on physical activity and dyspnea. However, this range of patient-specific characteristics is not captured by hemodynamic definitions or the etiology-based WHO classification system alone. By combining data from three studies with study sites in the USA, Europe, and Asia, the composition of our cohort was diverse.

Phenotypes in PAH

Comparison of phenotypes between clusters reveals several points of interest. Broadly, patients at time of their study enrollment tend to be either younger and lighter, have longer-standing PAH diagnosis, and lower resting heart rate (Clusters 1 and 2, approximately two-thirds of the cohort) or older, heavier, with shorter time since their PAH diagnosis, and higher resting heart rate (Clusters 3 and 4, the remaining one-third of the cohort). Patients fitting into the younger two-thirds of the cohort were more likely to be assigned specifically to Cluster 1 if they were non-Asian, had idiopathic or hereditary PAH, and no background PAH therapy. These patients had the best

functional capacity and symptoms across all clusters. Among the older third of the cohort (Clusters 3 and 4), patients in Cluster 4 were the heaviest, more likely to be White and on background therapy. Cluster 4 patients had the lowest 6MWD and almost 90% of these patients had WHO-FC III or IV symptoms. Interestingly, although time since PAH diagnosis decreased across clusters in concordance with worsening of functional capacity and symptoms, there was no corresponding trend in: (1) PAH etiology; (2) proportion of patients on no background therapy; or (3) frequency of combination regimens among patients on background therapy. Time since PAH diagnosis may represent a source of length time bias, where “less sick” patients were diagnosed earlier in their disease course than others due to more significant symptomatology or other work-up.

Cluster association with clinical outcome

We assessed response to therapy with oral treprostiniil as a case study utilizing cluster analysis of PAH trial data. The Food and Drug Administration approved oral treprostiniil in December 2013 for PAH patients to improve exercise capacity. The minimal 6MWD change yielding significant difference in quality of life approximates 33 m when standardized to quality of life assessments in PAH patients.¹² Although the mean improvement in 6MWD for all participants on study drug was less than 33 m, we observed that the treatment effect (i.e., the difference in 6MWD over time based on study intervention) varied with cluster membership. The PAH patients in Cluster 4 had the worst baseline functional capacity and obtained clinically meaningful improvement in 6MWD with treatment. Several potential reasons for this relationship may exist. A “ceiling effect” for additive treatment may exist in certain patients whose

pulmonary vasodilation is already maximized by their current treatment. Also, the therapeutic pathway yielding the most profound response may be variable between patients (and clusters). Although the heavier weight of Cluster 3 and 4 participants may have suggested a potentially greater tolerability of twice daily dosing of oral treprostinil, these participants had nearly twice as high rates of premature discontinuation of study drug as participants in Clusters 1 and 2, suggesting a “high risk/high reward” phenotype. As a next step, the cluster analysis should be validated in another PAH population and with other therapies.

Our findings have important implications for PAH classification and trial design. Within the context of precision medicine, our cluster analysis is a pragmatic approach to phenotypic description that could guide therapeutic decisions.^{13–16} In an era with several classes of PAH drugs, with novel therapies in the pipeline, this is a necessary next step for management of PAH. The overall PAH population is evolving over time, becoming older and heavier, and having more co-morbidities, and has an increasing number of therapeutic regimen options.^{17,18} In-depth phenotyping of this population is needed to shift towards more efficient and effective future clinical trial design.

Limitations

Although we included a large sample size from three multicenter randomized controlled trials, a limitation of our study is the lack of validation of our clustering with an external cohort. Also, our predefined cluster variables did not include hemodynamic or echocardiographic data due to limitation of the data collected in the trials, which would have likely added depth to the phenotypic description of our clusters. Finally, we did not have long-term clinical outcomes data available for our analysis, which would have also aided in understanding similarities and differences between the clusters.

Conclusions

This is the first study to our knowledge that performed cluster analysis for PAH patients and showed distinct clinical phenotypes among participants enrolled in PAH clinical trials. Although this technique needs to be validated in other PAH populations, cluster analysis may be effective in developing more efficient clinical trials and informing the timing and selection of therapies in patients with PAH.

Conflict of interest

The author(s) declare that there is no conflict of interest.

Funding

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

References

- Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J* 2016; 37: 67–119.
- Charalampopoulos A, Howard LS, Tzoulaki I, et al. Response to pulmonary arterial hypertension drug therapies in patients with pulmonary arterial hypertension and cardiovascular risk factors. *Pulm Circ* 2014; 4: 669–678.
- Mathai SC, Hassoun PM, Puhon MA, et al. Sex differences in response to tadalafil in pulmonary arterial hypertension. *Chest* 2015; 147: 188–197.
- Rose JA, Cleveland JM, Rao Y, et al. Effect of age on phenotype and outcomes in pulmonary arterial hypertension trials. *Chest* 2016; 149(5): 1234–1244.
- Rusiecki J, Rao Y, Cleveland J, et al. Sex and menopause differences in response to tadalafil: 6-minute walk distance and time to clinical worsening. *Pulm Circ* 2015; 5: 701–706.
- Ahmad T, Desai N, Wilson F, et al. Clinical implications of cluster analysis-based classification of acute decompensated heart failure and correlation with bedside hemodynamic profiles. *PLoS One* 2016; 11: e0145881.
- Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol* 2014; 64: 1765–1774.
- Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015; 131: 269–279.
- Jing ZC, Parikh K, Pulido T, et al. Efficacy and safety of oral treprostinil monotherapy for the treatment of pulmonary arterial hypertension: a randomized, controlled trial. *Circulation* 2013; 127: 624–633.
- Tapson VF, Jing ZC, Xu KF, et al. Oral treprostinil for the treatment of pulmonary arterial hypertension in patients receiving background endothelin receptor antagonist and phosphodiesterase type 5 inhibitor therapy (the FREEDOM-C2 study): a randomized controlled trial. *Chest* 2013; 144: 952–958.
- Tapson VF, Torres F, Kermeen F, et al. Oral treprostinil for the treatment of pulmonary arterial hypertension in patients on background endothelin receptor antagonist and/or phosphodiesterase type 5 inhibitor therapy (the FREEDOM-C study): a randomized controlled trial. *Chest* 2012; 142: 1383–1390.
- Mathai SC, Puhon MA, Lam D, et al. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2012; 186: 428–433.
- Council NR. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New*

- Taxonomy of Disease*. Washington DC: The National Academies Press, 2011, pp.1–128.
14. Collins FS and Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; 372: 793–795.
 15. Austin ED and Loyd JE. Toward precision medicine in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2015; 192: 1272–1274.
 16. Ashley EA. The precision medicine initiative: a new national effort. *JAMA* 2015; 313: 2119–2120.
 17. Frost AE, Badesch DB, Barst RJ, et al. The changing picture of patients with pulmonary arterial hypertension in the United States: how REVEAL differs from historic and non-US Contemporary Registries. *Chest* 2011; 139: 128–137.
 18. Poms AD, Turner M, Farber HW, et al. Comorbid conditions and outcomes in patients with pulmonary arterial hypertension: a REVEAL registry analysis. *Chest* 2013; 144: 169–176.