The Combination of Supervised and Unsupervised Learning based Risk Stratification and Phenotyping in Pulmonary Arterial Hypertension - a Long-term Retrospective Multicenter Trial

Figures for Reviewers

Innsbruck PAH registry

2023-03-24

# Figures

![Figure 1: Clustering analysis with the pooled IBK and LZ/W data set.](data:application/pdf;base64,)

Figure 1: Clustering analysis with the pooled IBK and LZ/W data set.

**Figure 1. Clustering analysis with the pooled IBK and LZ/W data set.**

*Clustering of the pooled data set with participants of the Innsbruck (IBK) and Linz/Vienna cohorts in respect to the survival-associated factors identified by elastic-net modeling (Figure 2A) was investigated by several algorithms (PAM: partitioning around medoids, HCl: hierarchical clustering, k-MEANS and SOM/HCl: combined self-organizing map/hierarchical clustering). The optimal cluster number k was determined by analysis of the curve of within-cluster sum of squares and maximal silhouette statistic. Explanatory performance of the algorithms was assessed by the fraction of ‘explained’ clustering variance (ratio of between-cluster sum of squares to total sum of squares) and correct cluster assignment rate in 10-fold cross-validation (CV). Statistic values are presented in bar plots. Algorithm, distance metric and k number of clusters are indicated in the Y axis.*

![Figure 2: Development of a Random Forest ensemble of PAH risk assessment tools and its performance at predicting overall survival.](data:application/pdf;base64,)

Figure 2: Development of a Random Forest ensemble of PAH risk assessment tools and its performance at predicting overall survival.

**Figure 2. Development of a Random Forest ensemble of PAH risk assessment tools and its performance at predicting overall survival.**

*Elastic Net score was developed as presented in Figure 2. The ensemble models of the Elastic Net score and established PAH risk assessment tools (FPHR 3p: French Pulmonary Hypertension Registry 3 parameter score, FPHR 4p: French Pulmonary Hypertension Registry 4 parameter score, COMPERA: Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension score, mRASP: modified Risk Assessment Score of PAH.) was established with the LASSO Cox regression (Supplementary Figure S3) and survival Random Forest algorithms (number of trees: 1000, mtry = 5, splitting rule: logrankscore, minimal node size: 2).*

*(A) Permutation importance of explanatory variables of the RF ensemble model.*

*(B) Predictive performance of the Elastic Net score, LASSO ensemble, Random Forest (RF) ensemble and single PAH risk scores at predicting overall survival was assessed by concordance index (C-index) and integrated Brier score (IBS). C-indexes and IBS for the risk assessment tools in the Innsbruck (IBK) and Linz/Vienna (LZ/W) cohorts are displayed in scatter plots.*