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

Figures and Tables

Innsbruck PAH registry

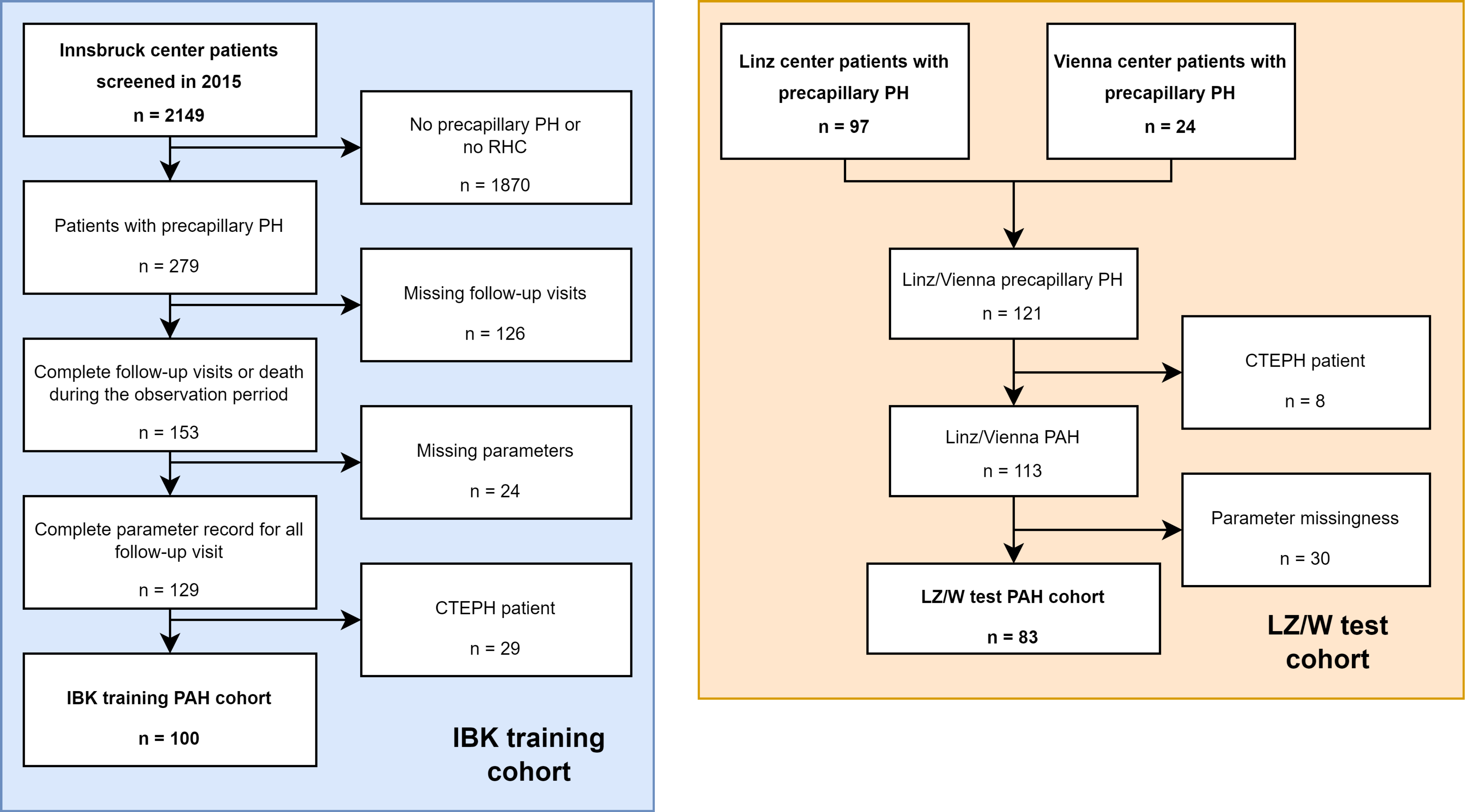
2023-03-27

# Tables

Table 1: Characteristic of the Innsbruck (IBK) and Linz/Vienna (LZ/W) study cohorts. Numeric variables are presented as medians with interquartile ranges (IQR) and ranges. Categorical variables are presented as percentages and counts within the complete observation set.

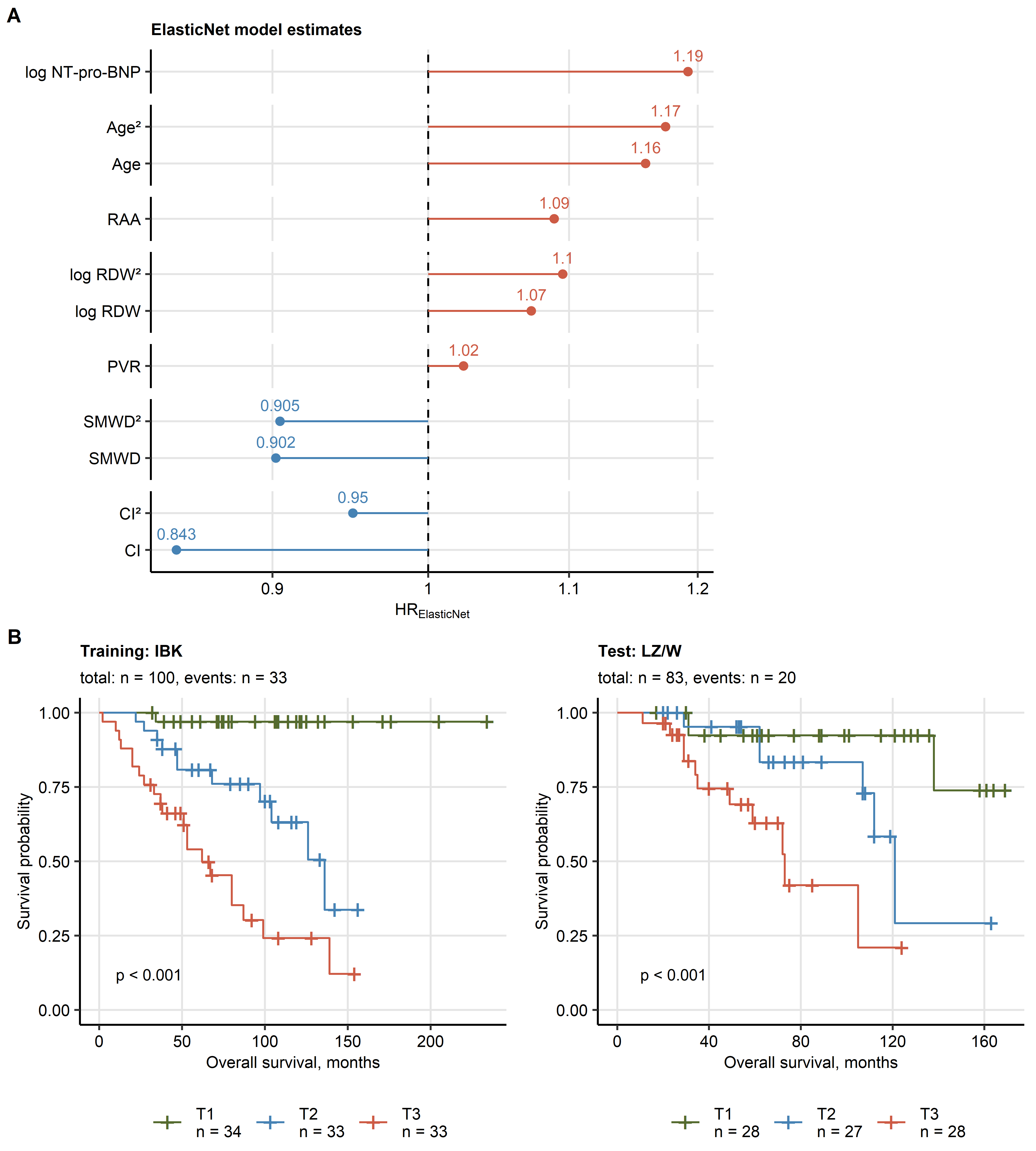
| **Variablea** | **IBK** | **LZ/W** | **Significanceb** | **Effect sizeb** |
| --- | --- | --- | --- | --- |
| Participants, n | 100 | 83 |  |  |
| Age, years | 66 [IQR: 53 - 71] 19 - 84 | 70 [IQR: 54 - 74] 23 - 82 | ns (p = 0.36) | r = 0.093 |
| Sex | female: 64% (64) male: 36% (36) | female: 66% (55) male: 34% (28) | ns (p = 0.94) | V = 0.024 |
| Anemia | 19% (19) | 17% (14) | ns (p = 0.94) | V = 0.028 |
| Renal insufficiency | 35% (35) | 18% (15) | p = 0.043 | V = 0.19 |
| Percardial effusion | 16% (16) | 3.6% (3) | p = 0.04 | V = 0.2 |
| WHO class | I/II: 39% (39) III/IV: 61% (61) | I/II: 53% (44) III/IV: 47% (39) | ns (p = 0.17) | V = 0.14 |
| SMWD, m | 320 [IQR: 200 - 400] 50 - 610 | 350 [IQR: 270 - 440] 50 - 620 | ns (p = 0.11) | r = 0.15 |
| mPAP, mmHg | 40 [IQR: 30 - 50] 26 - 120 | 39 [IQR: 31 - 49] 18 - 67 | ns (p = 0.91) | r = 0.024 |
| PVR, Wood | 10 [IQR: 6.7 - 17] 3.3 - 43 | 5 [IQR: 3.5 - 7.8] 1.4 - 20 | p < 0.001 | r = 0.54 |
| 5-year mortality | 21% (21) | 13% (11) | ns (p = 0.38) | V = 0.1 |
| OS, months | 70 [IQR: 46 - 110] 2 - 230 | 63 [IQR: 32 - 110] 11 - 170 | ns (p = 0.51) |  |
| aWHO class: WHO functional class; SMWD: six-minute walking distance; mPAP: mean pulmonary arterial pressure; PVR: pulmonary vascular resistance; OS: overall survival. | | | | |
| bNumeric variables: Mann-Whitney U test with r effect size statistic; categorical variables: χ² test with Cramer V effect size statistic; survival: log-rank test. | | | | |

# Figures



**Figure 1. Flow diagram of the study analysis inclusion process.**

*PH: pulmonary hypertension; RHC: right heart catheterization; CTEPH: chronic thromboembolic pulmonary hypertension.*

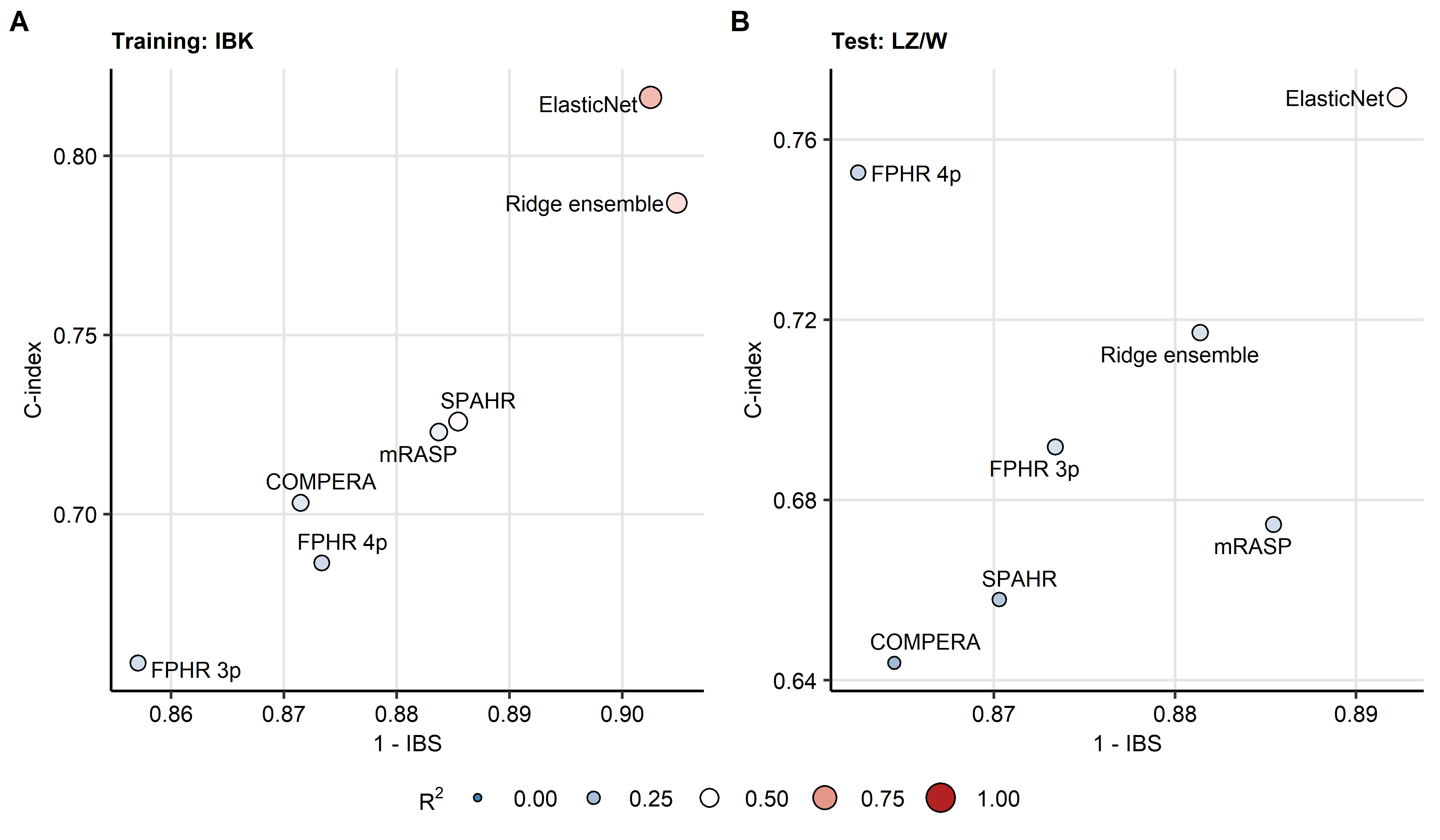


**Figure 2. Multi-parameter modeling of PAH survival with Elastic Net Cox regression.**

*The Elastic Net multi-parameter Cox regression model with the set of 19 (Supplementary Table S1) independent variables and overall survival as a response was developed in the training Innsbruck cohort. Numeric independent variables were median centered and their first and second order terms included in the model. Numbers of complete observations and mortality is indicated in B.*

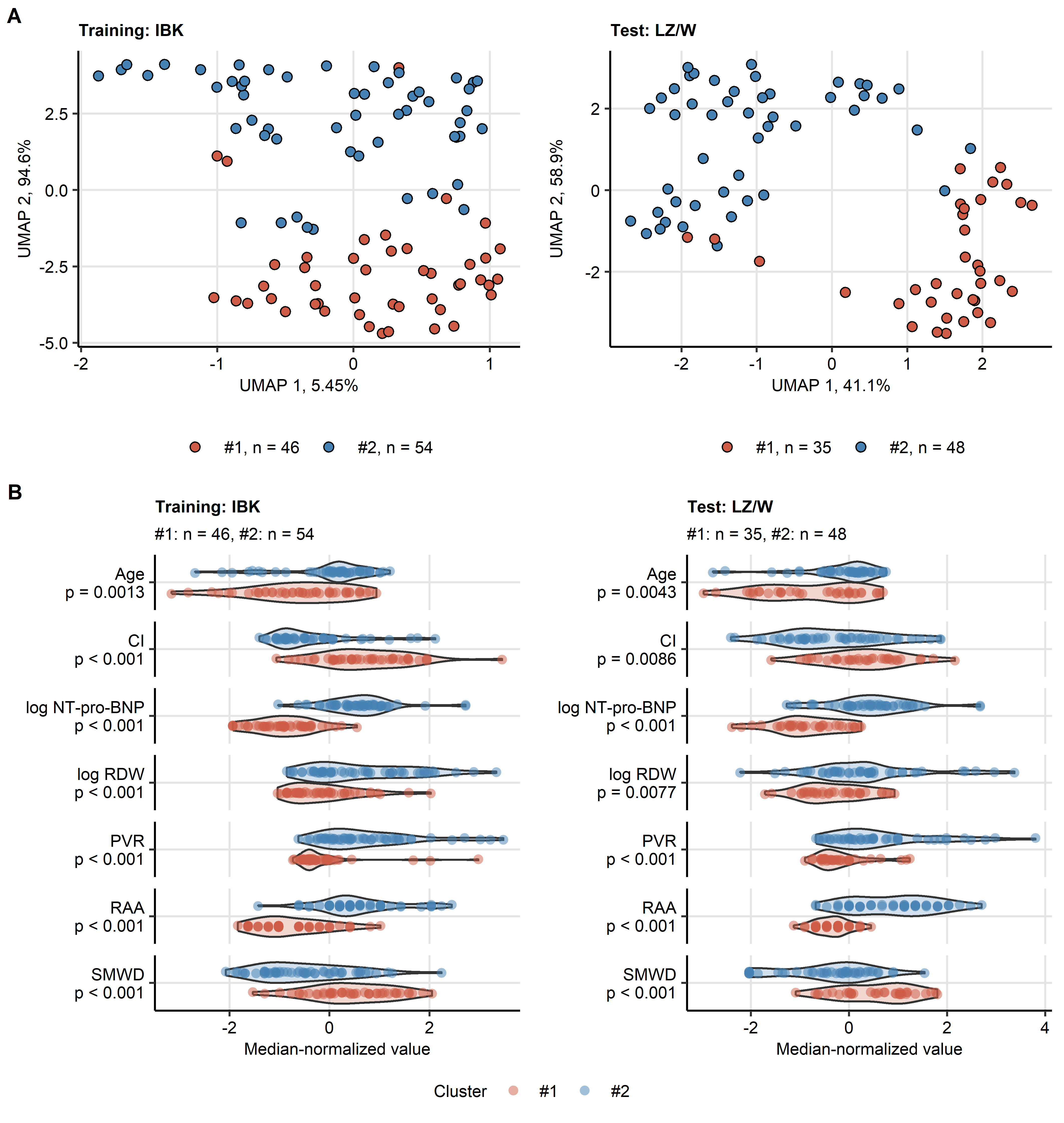
*(A) Non-zero Elastic Net model coefficients (Elastic Net signature) represented as hazard ratios (HR). Plot points are labeled with their HR values.*

*(B) Association of overall survival with the Elastic Net model linear prediction score in the training IBK and test Linz/Vienna (LZ/W) cohort was assessed by Kaplan-Meier analysis. Significance of the survival differences in the study participants stratified by the linear predictor score tertiles (T1: 0 - 33, T2: 34 - 66, T3: 66 - 100 percentile) was determined by log-rank test adjusted for multiple testing with Benjamini-Hochberg method. P values are shown in the plots, numbers of complete observations and mortality are indicated in the plot captions.*



**Figure 3. Performance of PAH risk assessment tools.**

*The Elastic Net signature was developed as presented in Figure 2. The ensemble of the established 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 developed by Ridge Cox regression as presented in Supplementary Figure S3. Predictive performance of the Elastic Net signature, 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, point size and color codes for .*



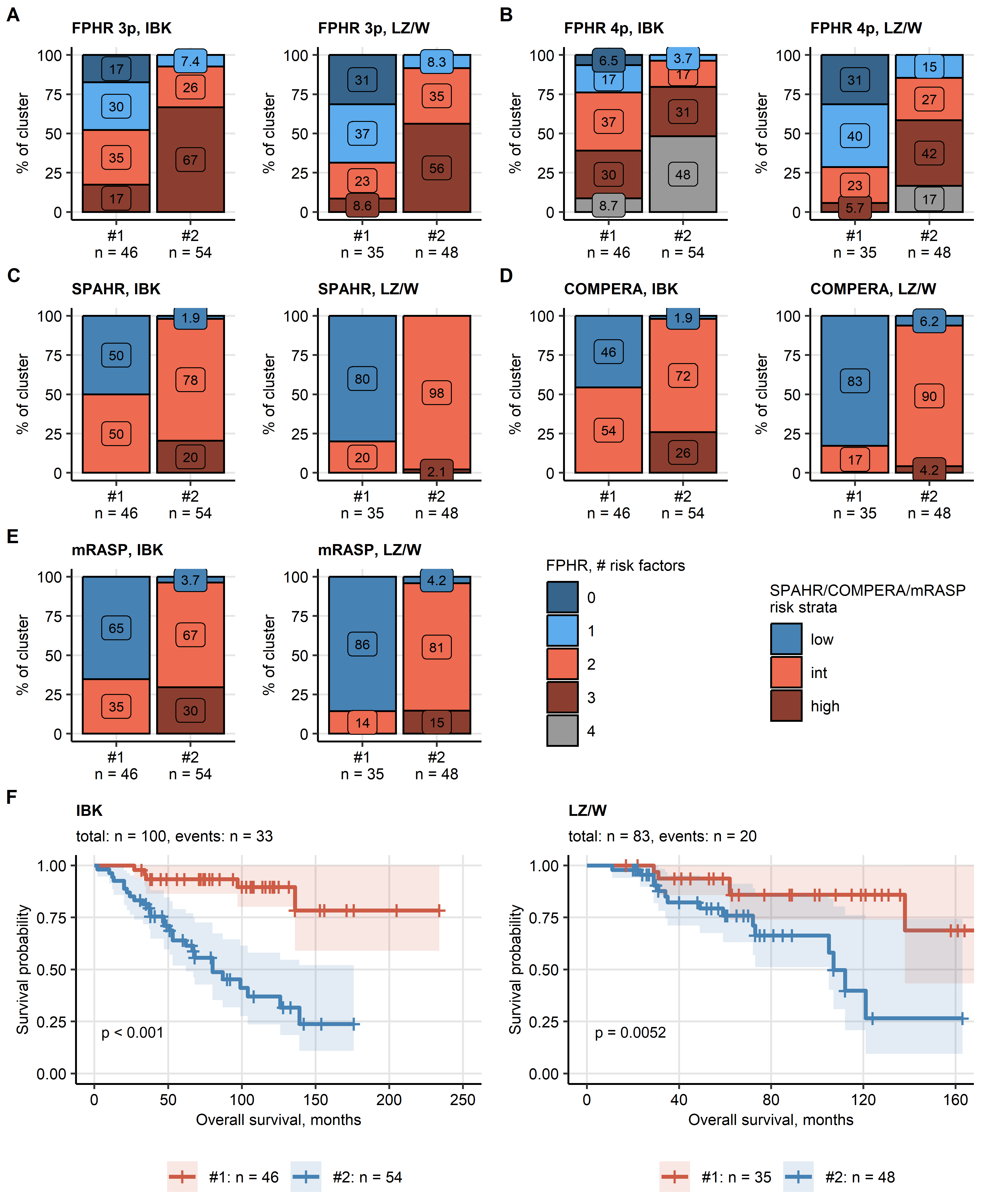
**Figure 4. Clustering of the study participants.**

*Clustering of the training Innsbruck (IBK) cohort participants in respect to the survival-associated factors identified by Elastic Net modeling (Figure 2) was investigated by PAM (partition around medoids) algorithm with cosine distance. Numeric clustering features were median centered prior to the clustering. Cluster assignment in the training Linz/Vienna cohort (LZ/W) was done by an inverse distance weighted k-nearest neighbor classifier. Numbers of individuals assigned to the PAH clusters are presented in the plot captions or legends.*

*(A) PAH cluster assignment overlaid on the 2-dimensional cosine-distance UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) layout plots. Percentages of variance associated with the components are indicated in the plot axes.*

*(B) Differences in the clustering features between the PAH clusters were assessed by Mann-Whitney test corrected for multiple testing with Benjamini-Hochberg method. Normalized, median-centered values of the clustering factors are shown in violin plots. Points represent single observations. P values are indicated in the Y axes.*

*CI: cardiac index; NT-pro-BNP: N terminal pro brain natriuretic peptide; RDW: red blood cell distribution width; PVR: pulmonary vascular resistance; RAA: right atrial area; SMWD: six minute walking distance.*



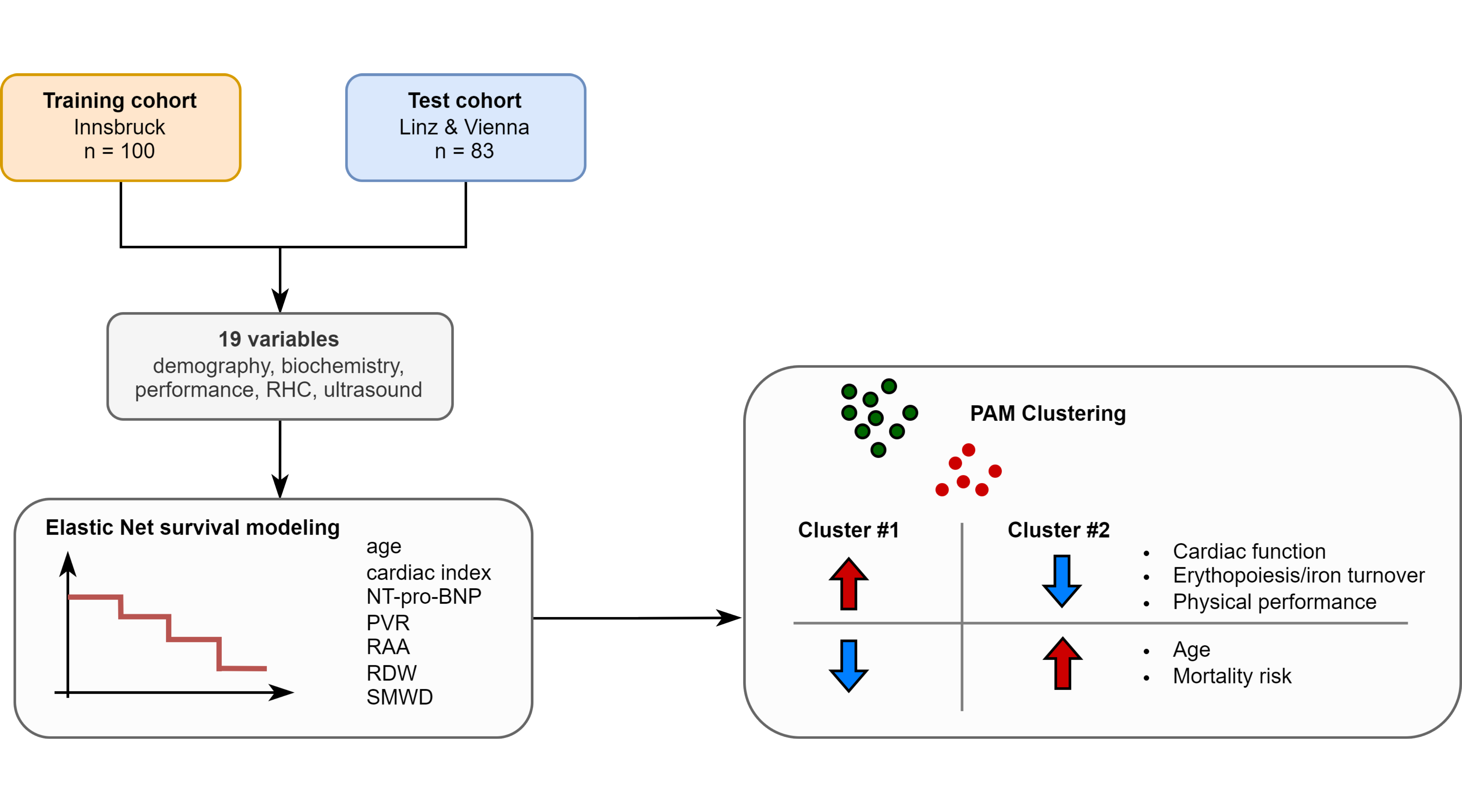
**Figure 5. Risk assessment and survival differences in the PAH clusters.**

*Risk assessment strata distribution and overall survival was compared between the study participant clusters with and log-rank test, respectively. P values were adjusted for multiple testing with Benjamini-Hochberg method. Numbers of individuals assigned to the clusters are presented in the Y axes or in the plot legends.*

*(A - E) Risk assessment strata frequencies in the PAH clusters presented in stack plots. All differences were significant with p < 0.001.*

*(F) Differences in overall survival in the PAH clusters visualized in Kaplan-Meier plots. P values are indicated in the plots. Numbers of complete observations and deaths are shown in the plot captions.*

*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; int.: intermediate.*



**Figure 6. Summary of the analysis results.**

*RHC: right heart catheterization; CI: cardiac index; NT-pro-BNP: N terminal pro brain natriuretic peptide; RDW: red blood cell distribution width; PVR: pulmonary vascular resistance; RAA: right atrial area; SMWD: six minute walking distance.*