**Appendix A**

**– Supplementary Methods and Results –**

**Clustering of patients and machine learning analyses.** Clustering analyses were performed by the R statistical software tool version 4.0.5 ([www.r-project.org](http://www.r-project.org)) using a set of 15 baseline characteristics reported in **Table A.1**. The pairwise distance between participants was estimated using the following approaches: 1) Euclidean distance from Factor Analysis for Mixed Data (FAMD) components explaining 80% of the variability (11 components); 2) Gower distance using untransformed variables and 3) Euclidean distance from the first 2 Multidimensional Scaling (MDS) dimensions using the dissimilarity between subjects () generated by the unsupervised Random Forest (RF) method as input. After distance matrices were estimated, Partition Around Medoids (PAM) and agglomerative hierarchical clustering (Ward method) were applied to identify clusters of subjects. An increasing number of clusters (from 2 to 10 in steps of 1) were tested and the Silhouette coefficient was computed. The optimal number of clusters by a combination of distance matrices and clustering methods was identified as the one reaching the highest Silhouette coefficient. PAMclustering applied to the Euclidean distance matrix derived from MDS dimensions on dissimilarities between subjects by unsupervised RF reached the largest Silhouette coefficient, thus it was applied for clusters identification (**Table A.2**)**.** Supervised RF using cluster labels as class while the set of variables used for clustering as attributes were used to assess the importance of each variable in discriminating among the identified clusters.

**Identification of decisional rules to predict patients’ cluster.** Supervised machine learning algorithms for multinomial classification have been trained and tested to assess and compare their classification accuracy in discriminating patients belonging to the three clusters using baseline values used for clustering as explanatory variables. Analyses were performed by the R statistical software tool version 4.2.1 ([www.r-project.org](http://www.r-project.org)). In details, two partitioning methods able to generate explainable decisional rules (classification trees and conditional inference trees – imposing a maximum tree depth = 2 and a minimum number of observations by terminal node = 50) have been trained and tested according to a 10-fold cross validation strategy. According to this schema, data were split randomly into 10 folds, used in turn as validation set to assess the discriminative performances of the models trained on data from the remaining folds (training set).Thus, a total number of 10 models by algorithm have been trained (10 classification trees and 10 conditional inference trees, each characterized by potentially different variables selected and tree structure depending on the training sets characteristics) and tested on the corresponding validation set not used for models learning. The mean classification accuracy (CA) of the predictions provided by the two evaluated algorithms in correctly assigning patients to clusters over the 10 validation sets are reported in **Table A.4**. Conditional inference trees showed the best performances, reaching mean CA = 82.49% ± standard deviation [SD] = 4.03%, with a mean improvement of +8.09% compared to the majority classifier (CA = 74.40% ± 4.09%) and of +2.55% compared to classification trees (CA = 79.94% ± 4.41%). **Table A.5** reports the mean performances reached by conditional inference trees by 10-fold cross validation in discriminating one cluster against the other two in turn: the mean balanced classification accuracy (BCA) was 84.73%, 73.21% and 69.25% in discriminating C1-severe, C2-intermediate and C3-mild from the other two clusters respectively (the mean BCA reached by the majority classifier was 50%). The final conditional inference tree model has been then learned on the whole dataset (n = 1159 patients): the tree structure and deriving decisional rules are reported in **Figure 3** (main text) and **Table A.6** respectively. Of note, the mean discriminative performances deriving from the cross-validation procedure do not represent the discriminative performances of the final conditional inference tree model presented.

**Appendix A**

**– Supplementary Results –**

**Table A.1 – Baseline demographic, anthropometric and clinical variables and corresponding values used for patients clustering.**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Type** | **Values** |
| Provenience | Nominal | Home, hospital |
| Age, years | Numeric discrete | - |
| Gender | Nominal | Females, males |
| CRF | Nominal | No, yes |
| Drugs Inhaler | Nominal | Triple, others |
| SpO2, % | Numeric discrete | - |
| BMI, Kg/m2 | Numeric continuous | - |
| Motor Barthel, points | Numeric discrete | - |
| BiD, points | Numeric discrete | - |
| CIRS, points | Numeric discrete | - |
| FEV1/FVC, % | Numeric discrete | - |
| GOLD quadrant stages | Ordinal | A, B, C, D |
| Baseline 6MWT, metres | Numeric discrete | - |
| Baseline CAT, points | Numeric discrete | - |
| Baseline MRC, points | Numeric discrete | - |

Variable = analysed variable; Type = type of variable; Values = values that nominal and ordinal variables assume. **Abbreviations** – CRF: Chronic respiratory failure; SpO2: pulsed oxygen saturation; BMI: Body-Mass index; BiD: Barthel index dyspnea; CAT: COPD assessment test; MRC: Medical research council; 6MWT: Six-minute walking distance test; FEV1: Forced expiratory volume at one second; FVC: Forced vital capacity; GOLD: Global strategy for prevention, diagnosis and management of COPD; CIRS: Comorbidity Index of Cumulative Illness Rating Scale.

**Table A.2 – Results from the clustering methods evaluated**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Distance between subjects** | **Clustering** | **Silhouette** | **# Clusters** |
| Euclidean (from 11 FAMD PCs explaining ~ 80%) | PAM | 0.15 | 5 |
|  | Agglomerative HC | 0.18 | 4 |
| Gower | PAM | 0.24 | 2 |
|  | Agglomerative HC | 0.20 | 6 |
| Euclidean (from 2 MDS dimensions on unsupervised RF dissimilarity) | PAM | 0.67 | 3 |
|  | Agglomerative HC | 0.62 | 3 |

Distance between subjects = distance metrics; Clustering = clustering method; Silhouette = Silhouette coefficient deriving from the analysis; # Clusters = optimal number of clusters based on the clustering methods. **Abbreviations** – FAMD: Factor Analysis for Mixed Data; MDS: Multidimensional Scaling; PAM: Partition Around Medoids; RF: Random Forest; PC = principal component; HC = Hierarchical Clustering.

**Table A.3 – Changes in clinical variables distribution by cluster in each centre.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Lumezzane (n = 414)** | | |  |  |
|  | **C1-severe** | **C2-intermediate** | **C3-mild** |  |  |
| **Variable** | **(n = 48, 11.59%)** | **(n = 325, 78.50%)** | **(n = 41, 9.90%)** | **p-value** |  |
| MRC, points |  |  |  |  |  |
| Change | -1 (-1, -1) | -1 (-2, -1) | -1 (-1, 0) | 0.1024 |  |
| Change reaching MCID | 42 (87.5%) | 256 (78.77%) | 30 (73.17%) | 0.2338 |  |
| CAT, points |  |  |  |  |  |
| Change | -5 (-8.25, -2) | -7 (-11, -4) | -7 (-10, -3) | 0.0445 | \* |
| Change reaching MCID | 39 (81.25%) | 298 (91.69%) | 37 (90.24%) | 0.0734 |  |
| 6MWT, metres |  |  |  |  |  |
| Change | 66.5 (26.5, 133.75) | 50 (20, 85) | 30 (7, 75) | 0.0362 | \* |
| Change reaching MCID | 35 (72.92%) | 227 (69.85%) | 25 (60.98%) | 0.4332 |  |
| *Change reaching MCID for all outcomes* | *24 (50%)* | *166 (51.08%)* | *17 (41.46%)* | *0.5436* |  |
|  |  |  |  |  |  |
|  | **Montescano (n = 138)** | | |  |  |
|  | **C1-severe** | **C2-intermediate** | **C3-mild** |  |  |
| **Variable** | **(n = 22, 15.94%)** | **(n = 116, 84.06%)** | **(n = 0, 0%)** | **p-value** |  |
| MRC, points |  |  |  |  |  |
| Change | -2 (-2, -2) | -1 (-1, -1) |  | <0.0001 | \* |
| Change reaching MCID | 21 (95.45%) | 102 (87.93%) |  | 0.4736 |  |
| CAT, points |  |  |  |  |  |
| Change | -10 (-12, -4.25) | -3 (-4, -1.75) |  | <0.0001 | \* |
| Change reaching MCID | 18 (81.82%) | 87 (75%) |  | 0.5928 |  |
| 6MWT, metres |  |  |  |  |  |
| Change | 105 (67, 167.5) | 40 (10, 74) |  | <0.0001 | \* |
| Change reaching MCID | 21 (95.45%) | 72 (62.07%) |  | 0.0017 | \* |
| *Change reaching MCID for all outcomes* | *17 (77.27%)* | *51 (43.97%)* |  | *0.0046* | *\** |
|  |  |  |  |  |  |
|  | **Pavia (n = 176)** | | |  |  |
|  | **C1-severe** | **C2-intermediate** | **C3-mild** |  |  |
| **Variable** | **(n = 48, 27.27%)** | **(n = 128, 72.73%)** | **(n = 0, 0%)** | **p-value** |  |
| MRC, points |  |  |  |  |  |
| Change | -2 (-2, -2) | -1 (-2, -1) |  | <0.0001 | \* |
| Change reaching MCID | 48 (100%) | 118 (92.19%) |  | 0.0663 |  |
| CAT, points |  |  |  |  |  |
| Change | -9 (-11, -7) | -3 (-5, -2) |  | <0.0001 | \* |
| Change reaching MCID | 45 (93.75%) | 98 (76.56%) |  | 0.0161 | \* |
| 6MWT, metres |  |  |  |  |  |
| Change | 81 (53.25, 133.5) | 36 (0, 72.75) |  | <0.0001 | \* |
| Change reaching MCID | 42 (87.5%) | 69 (53.91%) |  | <0.0001 | \* |
| *Change reaching MCID for all outcomes* | *39 (81.25%)* | *47 (36.72%)* |  | *<0.0001* | *\** |
|  |  |  |  |  |  |
|  | **Tradate (n = 369)** | | |  |  |
|  | **C1-severe** | **C2-intermediate** | **C3-mild** |  |  |
| **Variable** | **(n = 2, 0.54%)** | **(n = 239, 64.77%)** | **(n = 128, 34.69%)** | **p-value** |  |
| MRC, points |  |  |  |  |  |
| Change | -1.5 (-1.75, -1.25) | -1 (-1, -1) | -1 (-1, -1) | 0.0089 | \* |
| Change reaching MCID | 2 (100%) | 225 (94.14%) | 102 (79.69%) | 0.0025 | \* |
| CAT, points |  |  |  |  |  |
| Change | -17 (-18.5, -15.5) | -7 (-11, -3) | -4.5 (-7, -2) | <0.0001 | \* |
| Change reaching MCID | 2 (100%) | 215 (89.96%) | 107 (83.59%) | 0.177 |  |
| 6MWT, metres |  |  |  |  |  |
| Change | 71 (63, 79) | 42 (10, 75.5) | 29 (6.5, 53.5) | 0.0339 | \* |
| Change reaching MCID | 2 (100%) | 147 (61.51%) | 64 (50%) | 0.0317 | \* |
| *Change reaching MCID for all outcomes* | *2 (100%)* | *125 (52.3%)* | *47 (36.72%)* | *0.0028* | *\** |
|  |  |  |  |  |  |
|  | **Milano + Telese + Veruno (n = 62)** | | |  |  |
|  | **C1-severe** | **C2-intermediate** | **C3-mild** |  |  |
| **Variable** | **(n = 6, 9.68%)** | **(n = 54, 84.10%)** | **(n = 2, 3.23%)** | **p-value** |  |
| MRC, points |  |  |  |  |  |
| Change | -2 (-2, -2) | -1 (-2, -1) | 0 (0, 0) | 0.0315 | \* |
| Change reaching MCID | 5 (83.33%) | 49 (90.74%) | 0 (0%) | 0.0156 | \* |
| CAT, points |  |  |  |  |  |
| Change | -9.5 (-13.75, -4.5) | -3 (-7, 0) | -2.5 (-2.75, -2.25) | 0.3344 |  |
| Change reaching MCID | 5 (83.33%) | 32 (59.26%) | 2 (100%) | 0.374 |  |
| 6MWT, metres |  |  |  |  |  |
| Change | 114.5 (91.5, 139) | 46.5 (46, 74) | 105 (96.5, 113.5) | 0.0766 |  |
| Change reaching MCID | 6 (100%) | 48 (88.89%) | 2 (100%) | 0.6892 |  |
| *Change reaching MCID for all outcomes* | *5 (83.33%)* | *27 (50%)* | *0 (0%)* | *0.1144* |  |

Variable = analysed variable; Variables’ distribution by cluster is described as absolute frequency (relative frequency, %) or median (25th, 75th percentiles). \* p-value < 0.05.  **Abbreviations** – CAT: COPD assessment test; MRC: Medical research council; 6MWT: Six-minute walking distance test; MCID = minimal clinically important difference.

**Table A.4 – Overall accuracy of the considered methods in assigning patients to the correct cluster.**

|  |  |
| --- | --- |
| **Method** | **CA (%)** |
| Majority classifier | 74.40 ± 4.09 |
| Classification trees | 79.94 ± 4.41 |
| Conditional inference trees | 82.49 ± 4.03 |

Method = Machine learning method tested; CA (%) = Mean classification accuracy ± standard deviation. Majority classifier corresponds to the mean value and standard deviation of the discriminative performances obtained by assigning the most frequent class of the training set to all examples in the validation set, to be used as comparison term for classification trees and conditional inference trees classification accuracy assessment.

**Table A.5 – Conditional inference trees performances in discriminating between clusters**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Conditional Inference Trees** | | | | |
| **Comparison** | **Sens. (%)** | **Spec. (%)** | **PPV (%)** | **NPV (%)** | **BCA (%)** |
| **C1-severe** vs. C2-intermediate + C3-mild | 72.97 | 96.50 | 72.27 | 96.51 | 84.73 |
| **C2-intermediate** vs. C1-severe + C3-mild | 92.37 | 54.05 | 85.33 | 70.89 | 73.21 |
| **C3-mild** vs. C1-severe + C2-intermediate | 41.50 | 97.00 | 68.12 | 90.51 | 69.25 |
| Mean estimates | 68.94 | 82.52 | 75.24 | 85.97 | 75.73 |

Comparison = pairwise comparison (the reference cluster is highlighted in bold). The mean performances in discriminating the reference cluster from the other two clusters by cross-validation expressed in terms of sensitivity [Sens], specificity [Spec], Positive Predictive Value [PPV], Negative Predictive Value [NPV] and Balanced Classification Accuracy [BCA] are reported. As an example, a mean sensitivity of 72.97% for the comparison C1-severe vs. C2-intermediate + C3-mild indicate that on average 72.97% of all C1-severe patients were predicted as C1-severe.

**Table A.6 – Decisional rules derived from the conditional inference tree structure.**

|  |  |
| --- | --- |
| **Rule** | **Predicted cluster** |
| Baseline 6MWT ≤ 159 m AND BiD ≤ 28 | C2-intermediate |
| Baseline 6MWT ≤ 159 m AND BiD > 28 | C1-severe |
| Baseline 6MWT > 159 m AND GOLD quadrant stages = “A” | C3-mild |
| Baseline 6MWT > 159 m AND GOLD quadrant stages = “B”, “C” or “D” | C2-intermediate |

**Figure A.1 – Frequency distribution of assessed outcome measures.**

****

Histograms describe the frequency distribution of outcome measures at admission (leftmost), discharge (middle), and corresponding change after the program (rightmost). The black vertical dashed lines indicate the MCID (Minimal Clinically Important Difference). CAT: COPD Assessment test; MRC: Medical Research Council; 6MWT: Six-minute walking distance test.

**Figure A.2 – Multivariate variables importance in discriminating among clusters.**

****

Variables are ordered as function of their decreasing level of importance in discriminating among clusters from supervised RF using clusters’ label as dependent variable (class variable) and variables measured at admission used for clustering as independent variables (attributes). **Abbreviations** – BID: Barthel Index Dyspnea; BMI: Body-Mass Index; CAT: COPD Assessment test; CIRS: Cumulative Illness Rating Scale; GOLD: Global Initiative for Obstructive Lung Disease; CRF: Chronic respiratory failure; FEV1/FVC: Forced Expiratory Volume at one second/Forced Vital Capacity ratio; MRC: Medical Research Council; 6MWT: Six-minute walking distance test; SpO2: Pulsed Oxygen Saturation.