Investigating phenotypes of pulmonary COVID-19 recovery – a longitudinal observational prospective multicenter trial

Supplementary Material

CovILD study team

2022-01-05

Contents

Supplementary Tables

Table S1: **Study variables.**
Variable: variable name in the analysis pipeline, Reference time point: study visit, the variable was recorded at, Label: variable label in figures and tables.

Variable	Reference time point	Label	Variable type	Stratification cutoff
sex_male_V0	V0: acute COVID-19	Male sex	explanatory	NA
obesity_rec_V0	V0: acute COVID-19	Obesity	explanatory	BMI > 30 kg/m2
current_smoker_V0	V0: acute COVID-19	Current smoker	explanatory	NA
smoking_ex_V0	V0: acute COVID-19	Ex-smoker	explanatory	NA
CVDis_rec_V0	V0: acute COVID-19	CVD	explanatory	NA
hypertension_rec_V0	V0: acute COVID-19	Hypertension	explanatory	NA
PDis_rec_V0	V0: acute COVID-19	PD	explanatory	NA
COPD_rec_V0	V0: acute COVID-19	COPD	explanatory	NA
asthma_rec_V0	V0: acute COVID-19	Asthma	explanatory	NA
endocrine_metabolic_rec_V0	V0: acute COVID-19	Metabolic disorders	explanatory	NA
hypercholesterolemia_rec_V0	V0: acute COVID-19	Hypercholesterolemia	explanatory	NA
diabetes_rec_V0	V0: acute COVID-19	Diabetes	explanatory	NA
CKDis_rec_V0	V0: acute COVID-19	CKD	explanatory	NA
GITDis_rec_V0	V0: acute COVID-19	GITD	explanatory	NA
malignancy_rec_V0	V0: acute COVID-19	Malignancy	explanatory	NA
immune_deficiency_rec_V0	V0: acute COVID-19	Immune deficiency	explanatory	NA
weight_change_rec_V0	V0: acute COVID-19	Weight loss @V0	explanatory	at least 1 kg
dyspnoe_rec_V0	V0: acute COVID-19	Dyspnea @V0	explanatory	NA
cough_rec_V0	V0: acute COVID-19	Cough @V0	explanatory	NA
fever_rec_V0	V0: acute COVID-19	Fever @V0	explanatory	NA
night_sweat_rec_V0	V0: acute COVID-19	Night sweat @V0	explanatory	NA
pain_rec_V0	V0: acute COVID-19	Pain @V0	explanatory	NA
GI_sympt_rec_V0	V0: acute COVID-19	GI symptoms @V0	explanatory	NA
anosmia_rec_V0	V0: acute COVID-19	Anosmia @V0	explanatory	NA
ECOG_imp_rec_V0	V0: acute COVID-19	Impaired performance @V0	explanatory	ECOG at least 1
sleep_disorder_rec_V0	V0: acute COVID-19	Sleep disorders @V0	explanatory	NA
treat_antiinfec_rec_V0	V0: acute COVID-19	Anti-infectives @V0	explanatory	NA
treat_antiplat_rec_V0	V0: acute COVID-19	Anti-platelet @V0	explanatory	NA
treat_anticoag_rec_V0	V0: acute COVID-19	Anti-coagulatives @V0	explanatory	NA
treat_immunosuppr_rec_V0	V0: acute COVID-19	Immunosuppression @V0	explanatory	NA
anemia_rec_V1	V1: 60-day follow-up	Anemia @V1	explanatory	Male: Hb < 14 g/dL, Female: Hb < 12 g/dL
ferr_elv_rec_V1	V1: 60-day follow-up	Elevated ferritin @V1	explanatory	Male: > 300 ng/mL, Female: > 150 ng/ml
NTelv_rec_V1	V1: 60-day follow-up	Elevated NTproBNP @V1	explanatory	> 125 pg/mL
Ddimerelv_rec_V1	V1: 60-day follow-up	Elevated D-dimer @V1	explanatory	> 500 pg/mL FEU
CRP_elv_rec_V1	V1: 60-day follow-up	Elevated CRP @V1	explanatory	> 0.5 mg/dL
IL6_elv_rec_V1	V1: 60-day follow-up	Elevated IL-6 @V1	explanatory	> 7 pg/mL
iron_deficiency_30_rec_V1	V1: 60-day follow-up	Iron deficiency @V1	explanatory	TF-Saturation < 15%
age_65_V0	V0: acute COVID-19	Age over 65	explanatory	> 65 years
hosp_7d_V0	V0: acute COVID-19	Hospitalized >7 days @V0	explanatory	> 7 days
comorb_present_V0	V0: acute COVID-19	Any comorbidity	explanatory	> 0 comorbidities

Table S1: **Study variables.**
Variable: variable name in the analysis pipeline, Reference time point: study visit, the variable was recorded at, Label: variable label in figures and tables. *(continued)*

Variable	Reference time point	Label	Variable type	Stratification cutoff
comorb_3_V0	V0: acute COVID-19	>3 comorbidities	explanatory	> 3 comorbidities
overweight_V0	V0: acute COVID-19	Overweight or obesity	explanatory	BMI > 25 kg/m2
sympt_6_V0	V0: acute COVID-19	>6 symptoms @V0	explanatory	> 6 symptoms
sympt_present_V1	V1: 60-day follow-up	Persistent symptoms @V1	explanatory	> 0 symptoms @V3
ab_0_V1	V1: 60-day follow-up	Anti-S1/S2 IgG Q1 @V1	explanatory	(0, 312] BAU/ml
ab_25_V1	V1: 60-day follow-up	Anti-S1/S2 IgG Q2 @V1	explanatory	(312, 644] BAU/ml
ab_50_V1	V1: 60-day follow-up	Anti-S1/S2 IgG Q3 @V1	explanatory	(644, 975] BAU/ml
ab_75_V1	V1: 60-day follow-up	Anti-S1/S2 IgG Q4 @V1	explanatory	> 975 BAU/ml
pat_group_G1_V0	V0: acute COVID-19	Ambulatory @V0	explanatory	NA
pat_group_G2_V0	V0: acute COVID-19	Hospitalized @V0	explanatory	NA
pat_group_G3_V0	V0: acute COVID-19	Oxygen therapy @V0	explanatory	NA
pat_group_G4_V0	V0: acute COVID-19	ICU @V0	explanatory	NA
CT_findings_V3	180-day follow-up	CT abnormalities @V3	outcome	NA
CT_sev_low_V3	180-day follow-up	CT Severity Score @V3 1 – 5	outcome	NA
CTsevabove5_V3	180-day follow-up	CT Severity Score @V3 > 5	outcome	NA
sympt_present_V3	180-day follow-up	Symptoms present @V3	outcome	NA
lung_function_impaired_V3	180-day follow-up	Lung function impairment @V3	outcome	NA

Table S2: **Results of univariate risk modeling.**
Outcome: outcome variable at the 180-days follow-up visit (V3), Co-variate: explanatory variable, Baseline: reference level of the explanatory variable, OR: odds ratios with 95% confidence intervals, pFDR: significance p value corrected for multiple testing with the Benjamini-Hochberg method (FDR: false discovery rate).
The table is available online.

—

—

—

Table S3: **Feature cluster assignment scheme.**

Cluster	Variable
Cluster #1	Male sex, CVD, Hypertension, Metabolic disorders, Anti-infectives @V0, Elevated NTproBNP @V1, Elevated D-dimer @V1, Hospitalized >7 days @V0, >3 comorbidities, Overweight or obesity
Cluster #2	Obesity, Current smoker, Ex-smoker, PD, COPD, Asthma, Hypercholesterolemia, Diabetes, CKD, GITD, Malignancy, Immune deficiency, GI symptoms @V0, Anosmia @V0, Sleep disorders @V0, Anti-platelet @V0, Anti-coagulatives @V0, Immunosuppression @V0, Anemia @V1, Elevated ferritin @V1, Elevated CRP @V1, Elevated IL-6 @V1, Iron deficiency @V1, Age over 65, >6 symptoms @V0, Anti-S1/S2 IgG Q1 @V1, Anti-S1/S2 IgG Q2 @V1, Anti-S1/S2 IgG Q3 @V1, Anti-S1/S2 IgG Q4 @V1, Ambulatory @V0, Hospitalized @V0, Oxygen therapy @V0, ICU @V0, CT Severity Score @V3 1 – 5, CT Severity Score @V3 > 5, Lung function imp. @V3
Cluster #3	Weight loss @V0, Dyspnea @V0, Cough @V0, Fever @V0, Night sweat @V0, Pain @V0, Impaired performance @V0, Any comorbidity, Persistent symptoms @V1, Symptoms present @V3

Table S4: **Development of machine learning models.**
Outcome: outcome variable at the 180-days follow-up visit (V3).

Outcome	Classifier type	Caret method	Description	Package	Optimal arguments
CT abnormalities @V3	model	C5.0	C5.0	C50	trials = 10, model = tree, winnow = FALSE
		rf	Random Forest	randomForest	mtry = 27
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0105, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.000431
	ensemble	glmnet	glmnet	glmnet	alpha = 1, lambda = 0.0523
CT Severity Score @V3 > 5	model	C5.0	C5.0	C50	trials = 1, model = rules, winnow = TRUE
		rf	Random Forest	randomForest	mtry = 52
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.00979, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.0419
	ensemble	glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.000379
Symptoms present @V3	model	C5.0	C5.0	C50	trials = 1, model = tree, winnow = FALSE
		rf	Random Forest	randomForest	mtry = 27
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0109, C = 1
		nnet	Neural Network	nnet	size = 3, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.000247
	ensemble	glmnet	glmnet	glmnet	alpha = 0.1, lambda = 0.0167
Lung function impairment @V3	model	C5.0	C5.0	C50	trials = 1, model = rules, winnow = FALSE
		rf	Random Forest	randomForest	mtry = 52
		svmRadial	Support Vector Machines with Radial Basis Function Kernel	kernlab	sigma = 0.0108, C = 0.5
		nnet	Neural Network	nnet	size = 1, decay = 0.1
		glmnet	glmnet	glmnet	alpha = 0.55, lambda = 0.0341
	ensemble	glmnet	glmnet	glmnet	alpha = 0.55, lambda = 0.0387

Table S5: **Performance of machine learning classifiers.**
Outcome: outcome variable at the 180-days follow-up visit (V3).
The table is available online.

—

—

—

Table S6: **Performance of machine learning classifiers.**
Outcome: outcome variable at the 180-days follow-up visit (V3), Cohort subset: cohort acute COVID-19 severity strata (mild-moderate: outpatient or hospitalized without oxygen, severe-critical: oxygen therapy or ICU).
The table is available online.

—

—

—

Supplementary Figures

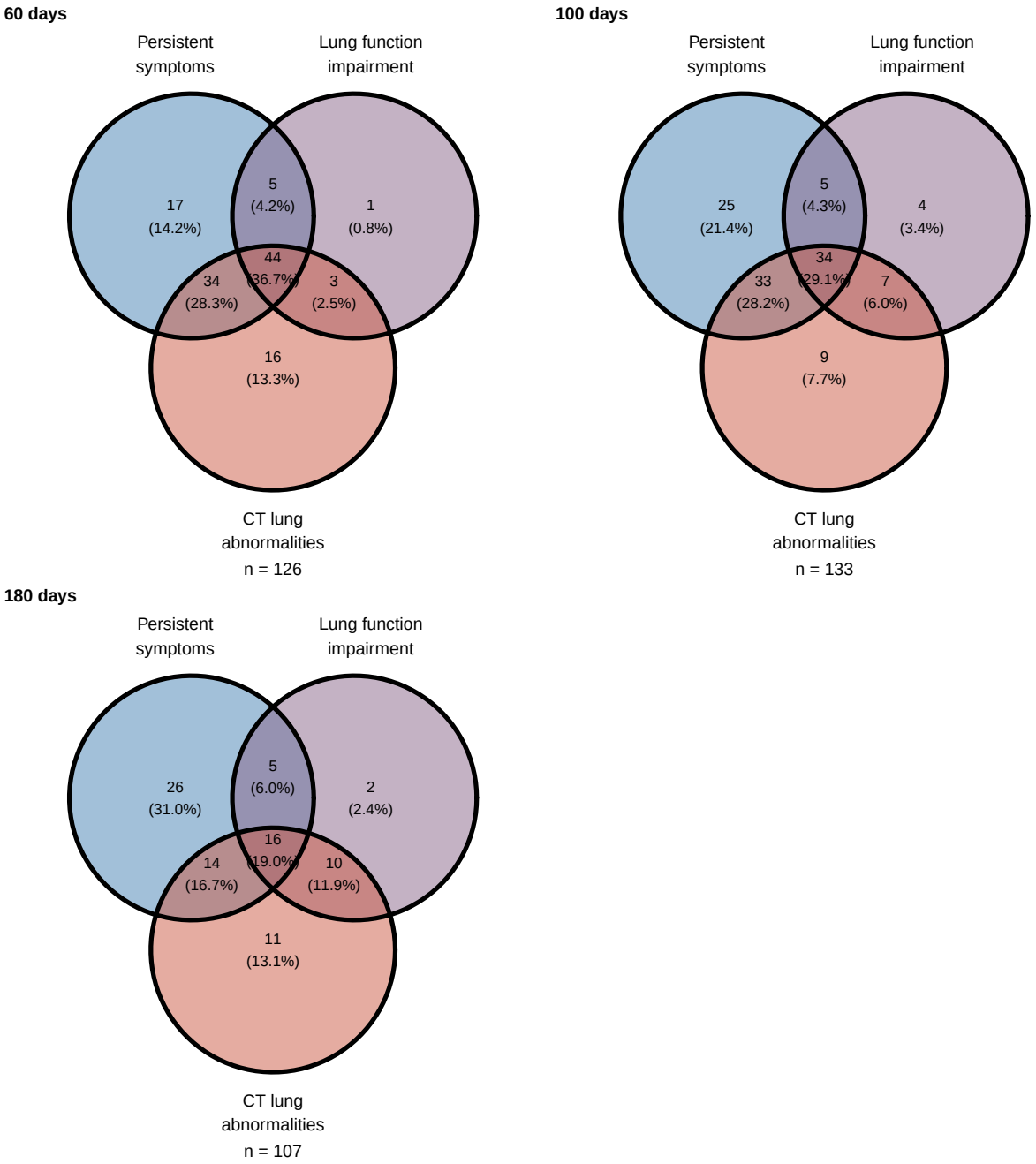

Figure S1: Co-occurrence of any lung CT abnormalities, functional lung impairment and persistent symptoms.

Supplementary Figure S1. Co-occurrence of any lung CT abnormalities, functional lung impairment and persistent symptoms.

Numbers and percentages of the study participants with persistent symptoms, functional lung impairment or any lung CT abnormalities at the consecutive follow-up visits presented in Venn diagrams. Numbers of complete observations are shown under the plots.

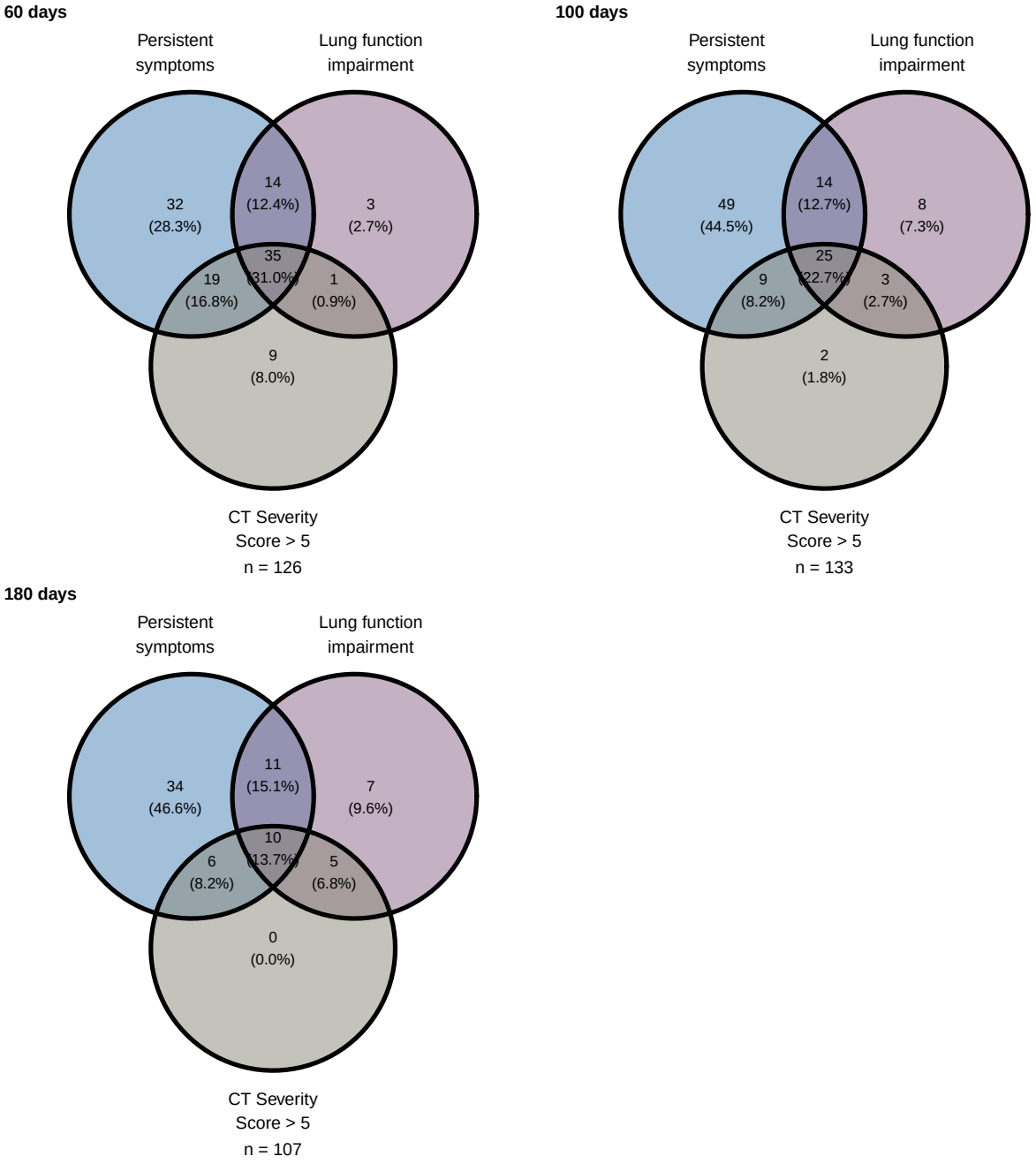

Figure S2: Co-occurrence of any lung CT abnormalities, functional lung impairment and persistent symptoms.

Supplementary Figure S2. Co-occurrence of moderate-to-severe lung CT abnormalities, functional lung impairment and persistent symptoms.

Numbers and percentages of the study participants with persistent symptoms, functional lung impairment or moderate-to-severe lung CT abnormalities (severity score > 5) at the consecutive follow-up visits presented in Venn diagrams. Numbers of complete observations are shown under the plots.

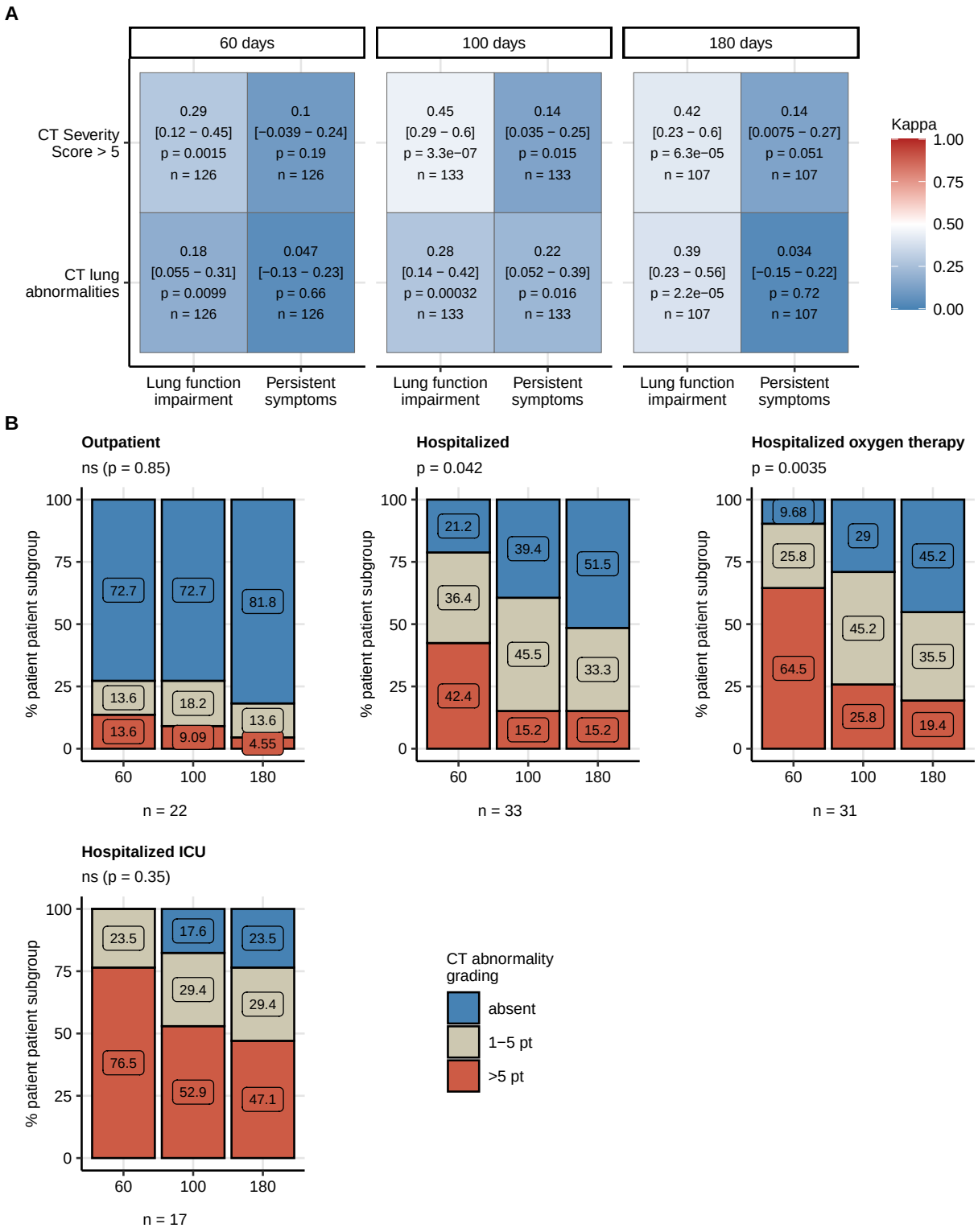

Figure S3: Frequency of mild and moderate-to-severe lung CT abnormalities. Prognostic value of functional lung impairment and persistent symptoms for prediction of radiological lung abnormalities.

Supplementary Figure S3. Frequency of mild and moderate-to-severe lung CT abnormalities. Prognostic value of functional lung impairment and persistent symptoms for prediction of

radiological lung abnormalities.

(A) Percentages of mild (severity score ≤ 5) and moderate-to-severe lung CT abnormalities at the consecutive follow-up visits in the study participants stratified by the severity of acute COVID-19. Statistical significance of frequency differences was determined by χ^2 test for trend corrected for multiple testing with the Benjamini-Hochberg method. Numbers of complete observations are indicated under the plots.

(B) Relevance of functional lung impairment and persistent COVID-19 symptoms at predicting any lung CT abnormalities and moderate-to-severe lung CT abnormalities (severity score > 5) at the consecutive follow-up visits. The concordance of the outcome variables was determined by Cohen's κ coefficient. Statistical significance ($\kappa \neq 0$) was assessed by two-tailed T test corrected for multiple testing with the Benjamini-Hochberg method. Kappa with 95 confidence intervals and p values are presented as a heat map. Numbers of complete observations are indicated in the plot.

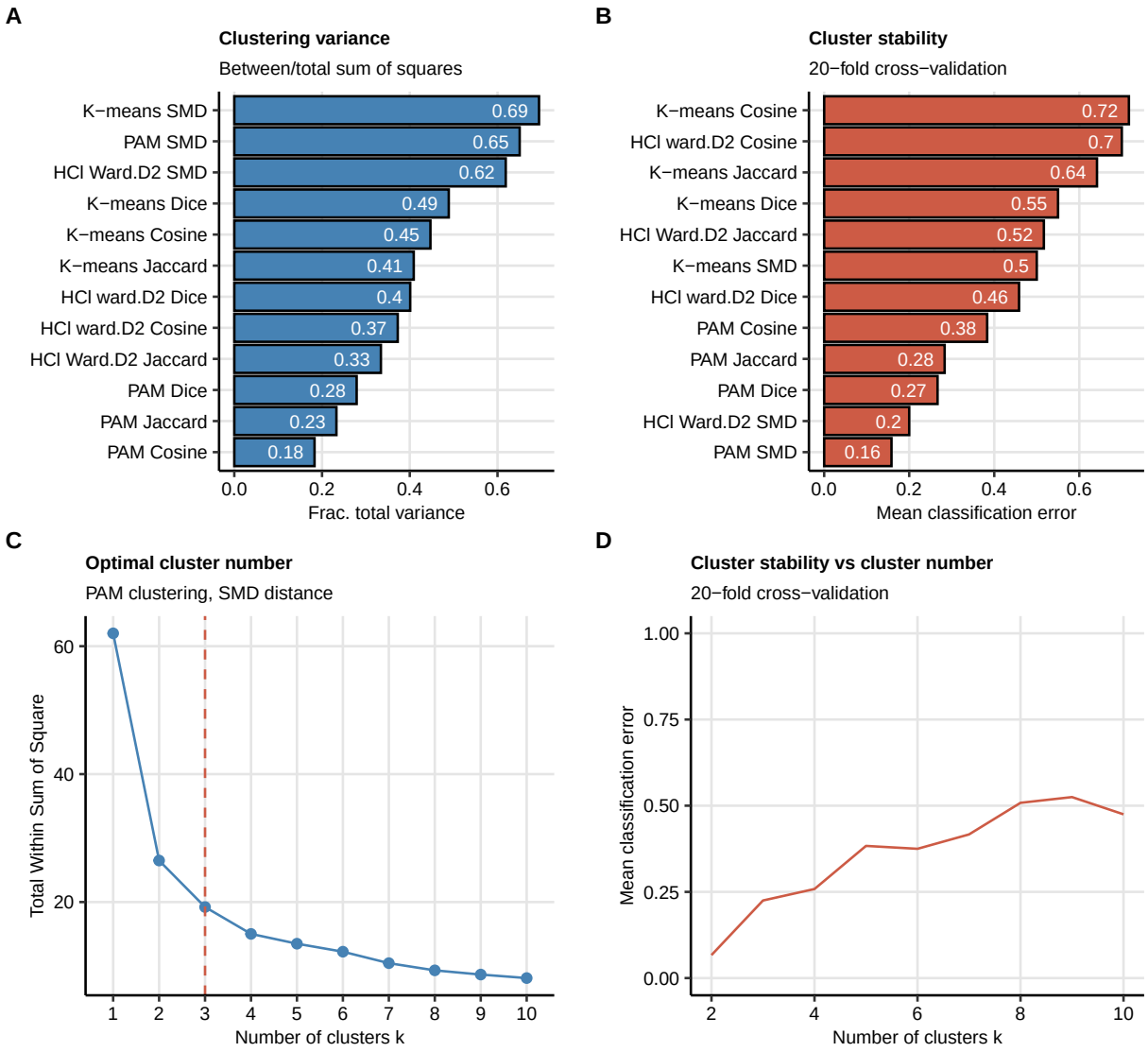

Figure S4: Study feature clustering algorithm.

Supplementary Figure S4. Study feature clustering algorithm.

Clustering of 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 (V0) or at the early 60-day follow-up visit (V1) (**Supplementary Table S1**).

(A, B) Comparison of 'explained' variances (between-cluster to total sum-of-squares ratio, **A**) and cluster stability (mean classification error in 20-fold cross-validation, **B**) in clustering of the data set with several algorithms with k = 3 centers/branches (algorithms: K-means, PAM: partitioning around medoids, HCl ward.D2: hierarchical clustering with ward.D2 method, distances: SMD: simple matching distance, Jaccard, Dice and Cosine).

(C, D) Optimal number of the feature clusters in clustering with the optimally performing PAM algorithm with SMD dissimilarity measure was determined by the bend of the total within-cluster sum-of-squares curve (**C**) and confirmed by good stability (low mean classification error) in 20-fold cross-validation (**D**).

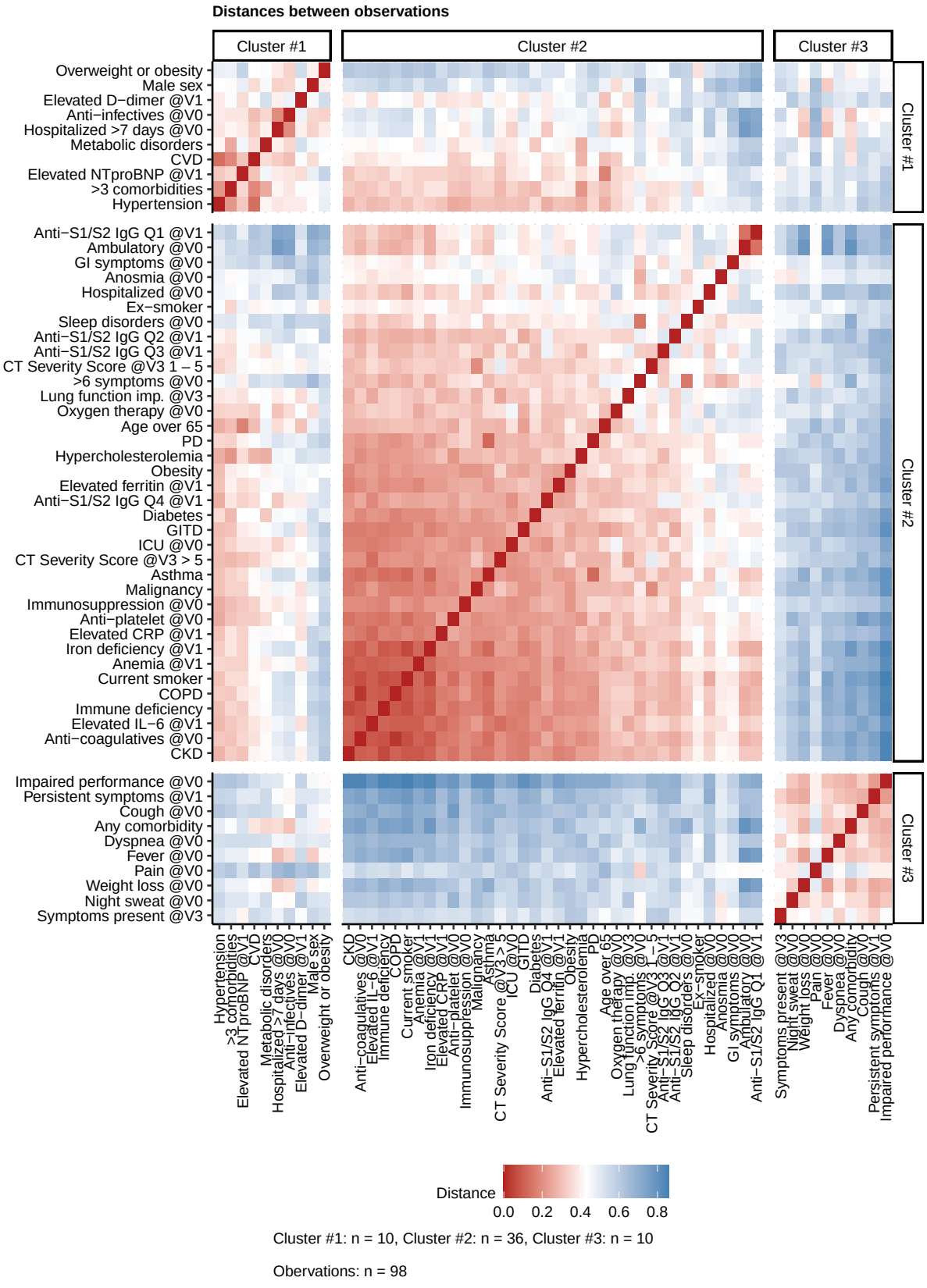

Figure S5: Semi-supervised clustering of mild and moderate-to-severe lung CT abnormalities, functional lung impairment and persistent symptoms at the 108-day follow-up with parameters of acute COVID-19 and early convalescence.

Supplementary Figure S5. Semi-supervised clustering of mild and moderate-to-severe lung CT abnormalities, functional lung impairment and persistent symptoms at the 108-day follow-up with parameters of acute COVID-19 and early convalescence.

Clusters of 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 (V0) or at the early 60-day follow-up visit (V1) (**Supplementary Table S1**) were defined by the optimally performing PAM algorithm and SMD (simple matching distance) dissimilarity measure (**Figure 6A**, **Supplementary Figure S4**, **Supplementary Table S3**). The cluster assignment for the outcome variables at the 180-day follow-up visit (V3, persistent symptoms, functional lung impairment, mild lung CT abnormalities [severity score ≤ 5] and moderate-to-severe lung CT abnormalities [severity score > 5]) was predicted by k-nearest neighbor (k-NN) label propagation procedure. SMD between the features and their cluster assignments are shown in a heat map. Numbers of the features in the clusters and the total number of observations are indicated under the plot.

CVD: cardiovascular disease, Q1, Q2, Q3, Q4: 1^{st}, 2^{nd}, 3^{rd} and 4^{th} quartile, GI: gastrointestinal, PD: pulmonary disease, GITD: gastrointestinal disease, ICU: intensive care unit, COPD: chronic obstructive pulmonary disease, CKD: chronic kidney disease.

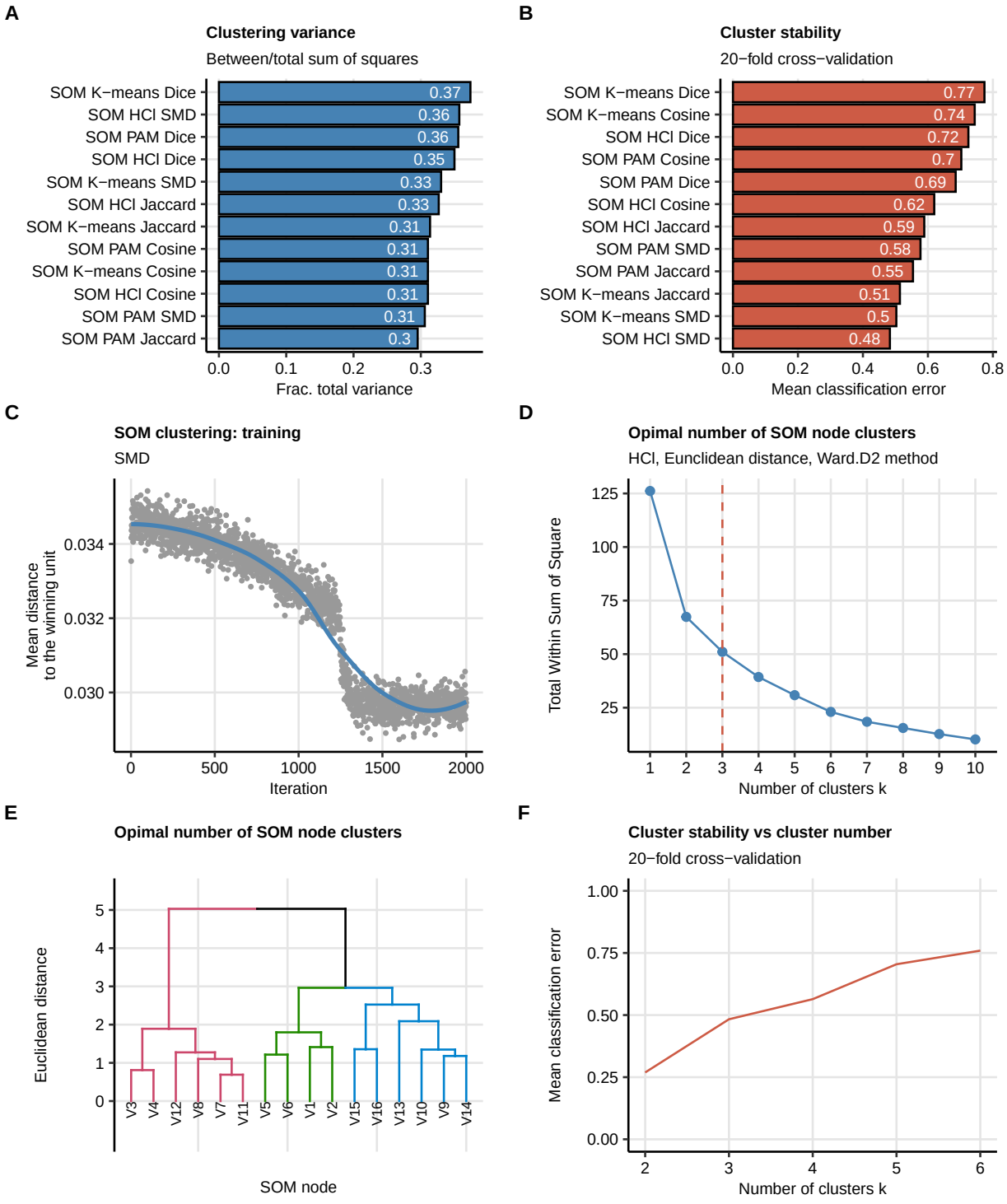

Figure S6: Study participant clustering algorithm.

Supplementary Figure S6. Study participant clustering algorithm.

Clustering of the study participants (n = 133 with the complete variable set) in respect to 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 (V0) or at the early 60-day follow-up visit (V1) (**Supplementary Table S1**). The procedure involved clustering of the observations with self-organizing maps (SOM, 4 × 4 hexagonal grid, distances: SMD: simple matching distance, Jaccard,

Dice or cosine) followed by clustering of the SOM nodes (algorithms: HCl ward.D2: hierarchical clustering with Ward.D2 method, K-means, PAM: partitioning around medoids, distance: euclidean). Different combinations of observation dissimilarity measures and SOM node clustering algorithms were tested at the search for the optimal clustering algorithm.

(A, B) Comparison of 'explained' variances (between-cluster to total sum-of-squares ratio, **A**) and cluster stability (mean classification error in 20-fold cross-validation, **B**) in clustering of the data set with different observation distance measures and SOM node clustering algorithms.

(C - E) Optimal number of the SOM node clusters in clustering with the optimally performing SOM HCl algorithm with SMD observation dissimilarity measure. The optimal cluster number was determined by the bend of the total within-cluster sum-of-squares curve (**C**) and confirmed by visual inspection of the HCl dendrogram (**E**) and good stability (low mean classification error) in 20-fold cross-validation (**D**).

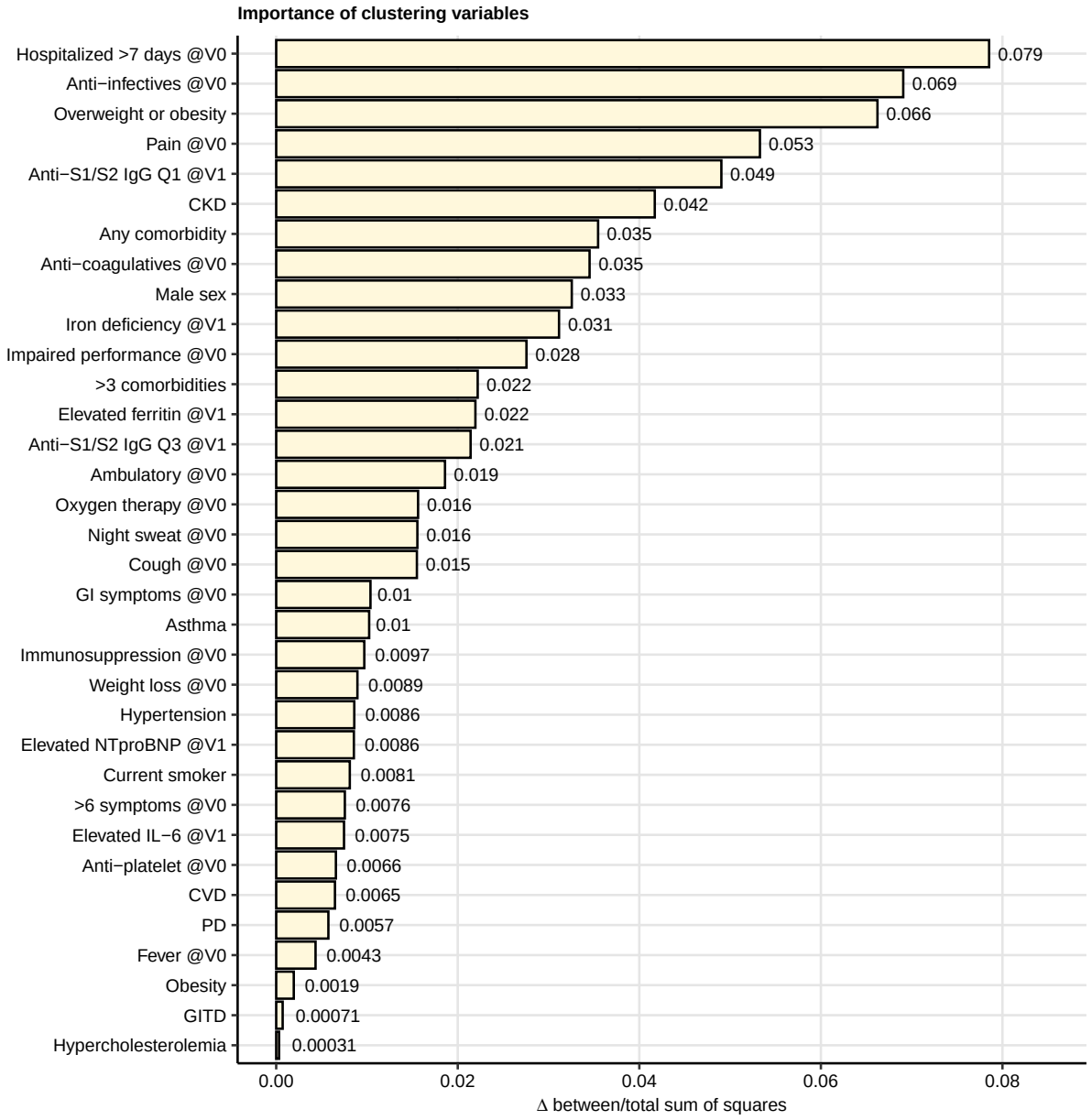

Figure S7: Impact of particular variables on the quality of participant clustering.

Supplementary Figure S7. Impact of particular variables on the quality of participant clustering.

Participants clusters were defined with the optimally performing HCl algorithm with SMD observation dissimilarity measure as presented in **Figure 7** and **Supplementary Figure S6**. The impact of a particular clustering variable was determined by comparing the 'explained' clustering variance (between-cluster to total sum-of-squares ratio) between the initial cluster structure and the structure wit random re-shuffling of the variable ('noising'). Differences in the clustering variances for the most influential clustering variables (Δ clustering variance > 0) are presented in the plot.

Q1, Q3: 1st, 3rd quartile, CKD: chronic kidney disease, GI: gastrointestinal, CVD: cardiovascular disease, PD: pulmonary disease, GITD: gastrointestinal disease, V0: acute COVID-19, V1: 60-day follow-up visit.

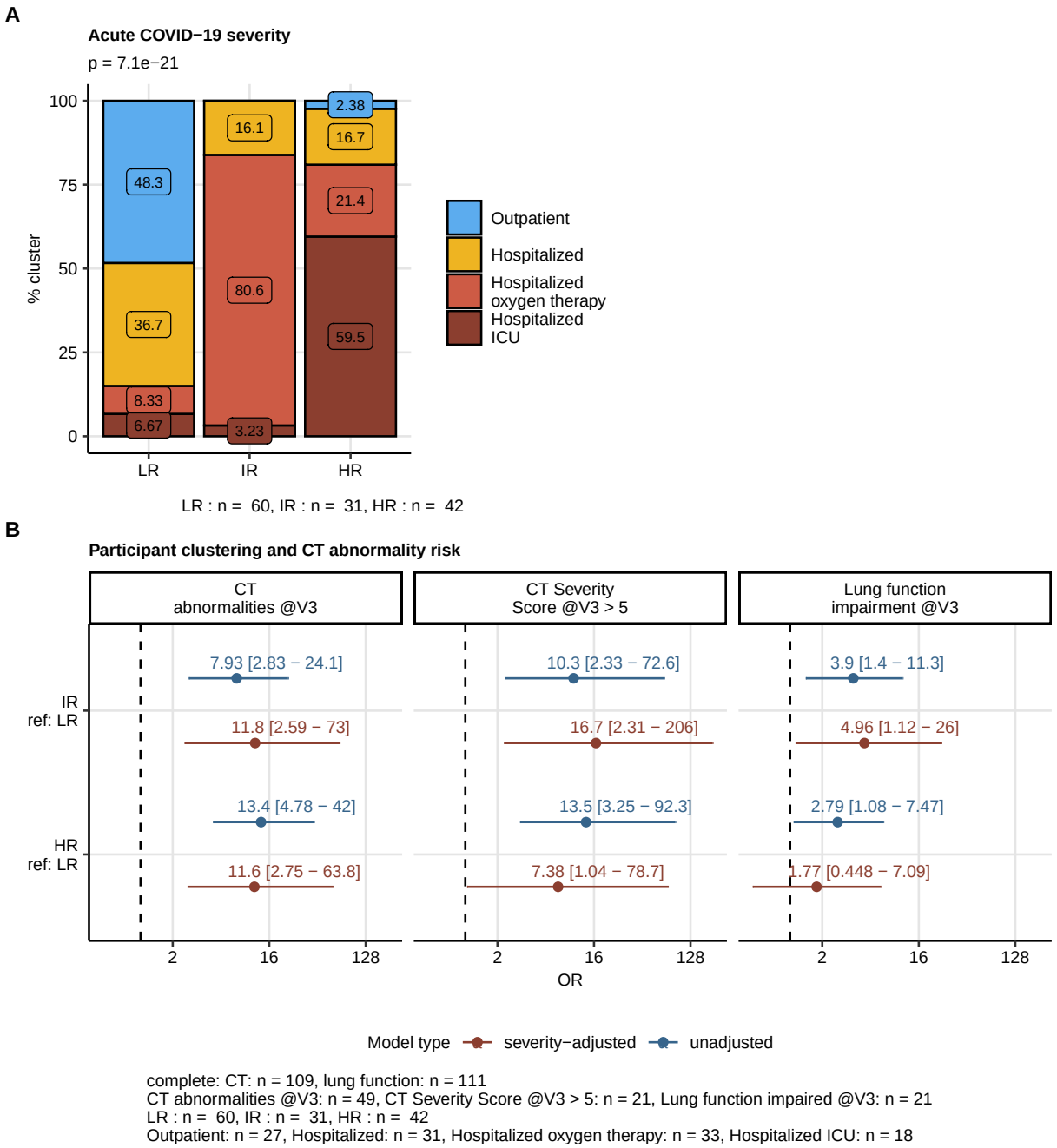

Figure S8: Risk of radiological lung abnormalities at the 180-day follow-up in the participant clusters.

Supplementary Figure S8. Risk of radiological lung abnormalities at the 180-day follow-up in the participant clusters.

Participants clusters were defined by non-lung function and non-CT clinical features of acute COVID-19 and early convalescence (60-day follow-up visit, **Supplementary Table S1**) with the optimally performing HCl algorithm with SMD observation dissimilarity measure as presented in **Figure 7** and **Supplementary Figure S6**. Association of the participant cluster assignment (LR: low-risk, IR: intermediate-risk, HR: high-risk cluster) with the risk of any lung CT abnormalities and moderate-to-severe lung CT abnormalities (severity score $>$ 5) at the 180-day follow-up visit (V3) was investigated by logistic modeling with and without inclusion of the acute COVID-19 severity effect (severity-adjusted). Odds ratio (OR) significance

was determined by Wald Z test and corrected for multiple testing with the Benjamini-Hochberg method. OR wit 95 confidence intervals are presented in Forest plots. Numbers of complete observations, outcome events, participants in the clusters and the acute COVID-19 severity subsets are indicated under the plot.

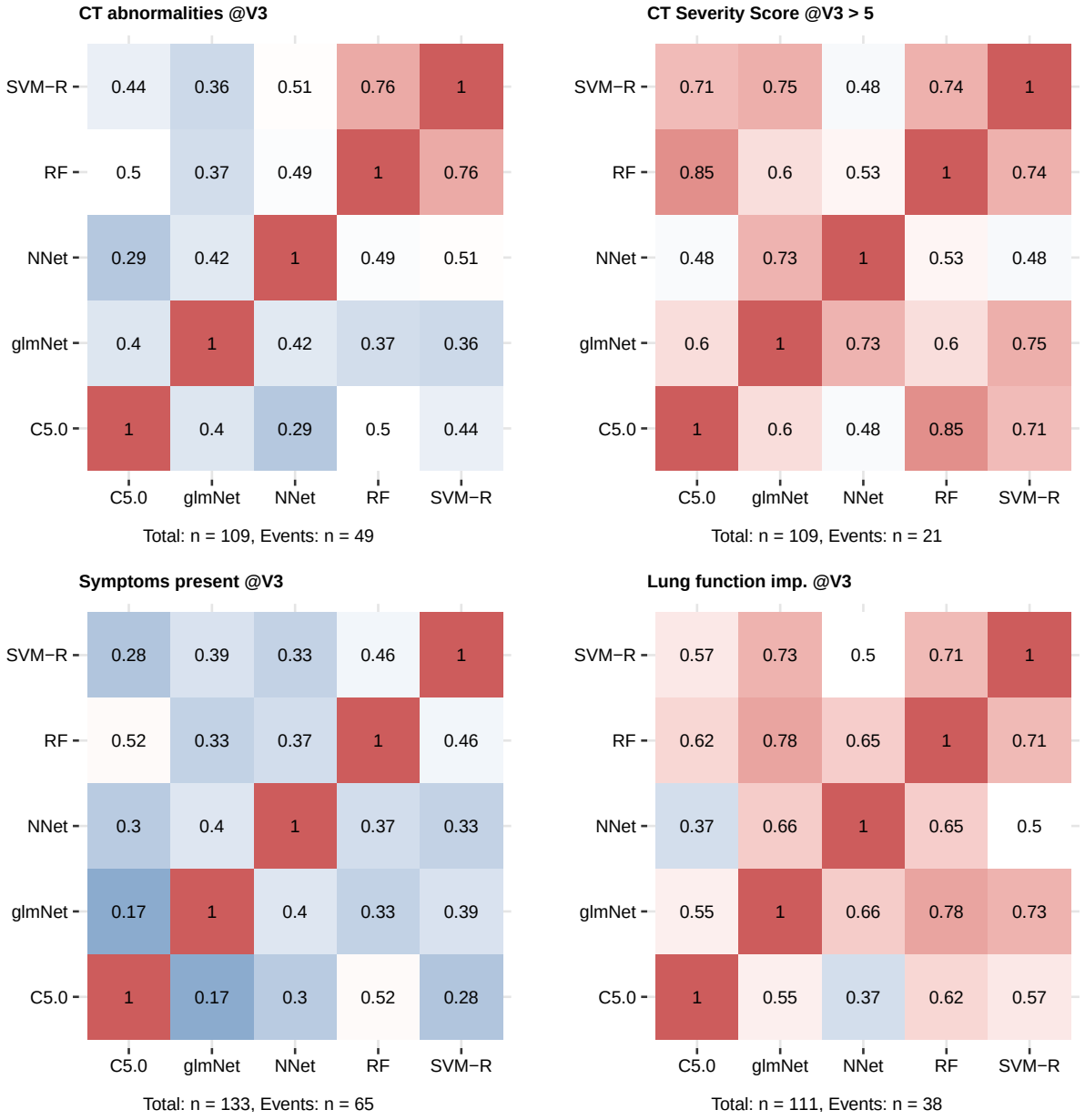

Figure S9: Correlation of the machine learning algorithm prediction accuracy.

Supplementary Figure S9. Correlation of the machine learning algorithm prediction accuracy.

Machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) were trained in the cohort data set with 52 non-CT and non-lung function binary explanatory variables recorded for acute COVID-19 (V0) or at the early 60-day follow-up visit (V1) (**Supplementary Table S1**) for predicting outcome variables at the 180-day follow-up visit (V3, any lung CT abnormalities, moderate-to-severe lung CT abnormalities [severity score > 5], functional lung impairment and persistent symptoms) (**Figure 9**, **Supplementary Table S4**). The prediction accuracy was verified by repeated 20-fold cross-validation (5 repeats). Pearson's correlation coefficients of the classifier prediction accuracy in the cross-validation are presented as heat maps. Numbers of complete observations and outcome events are indicated under the plots.

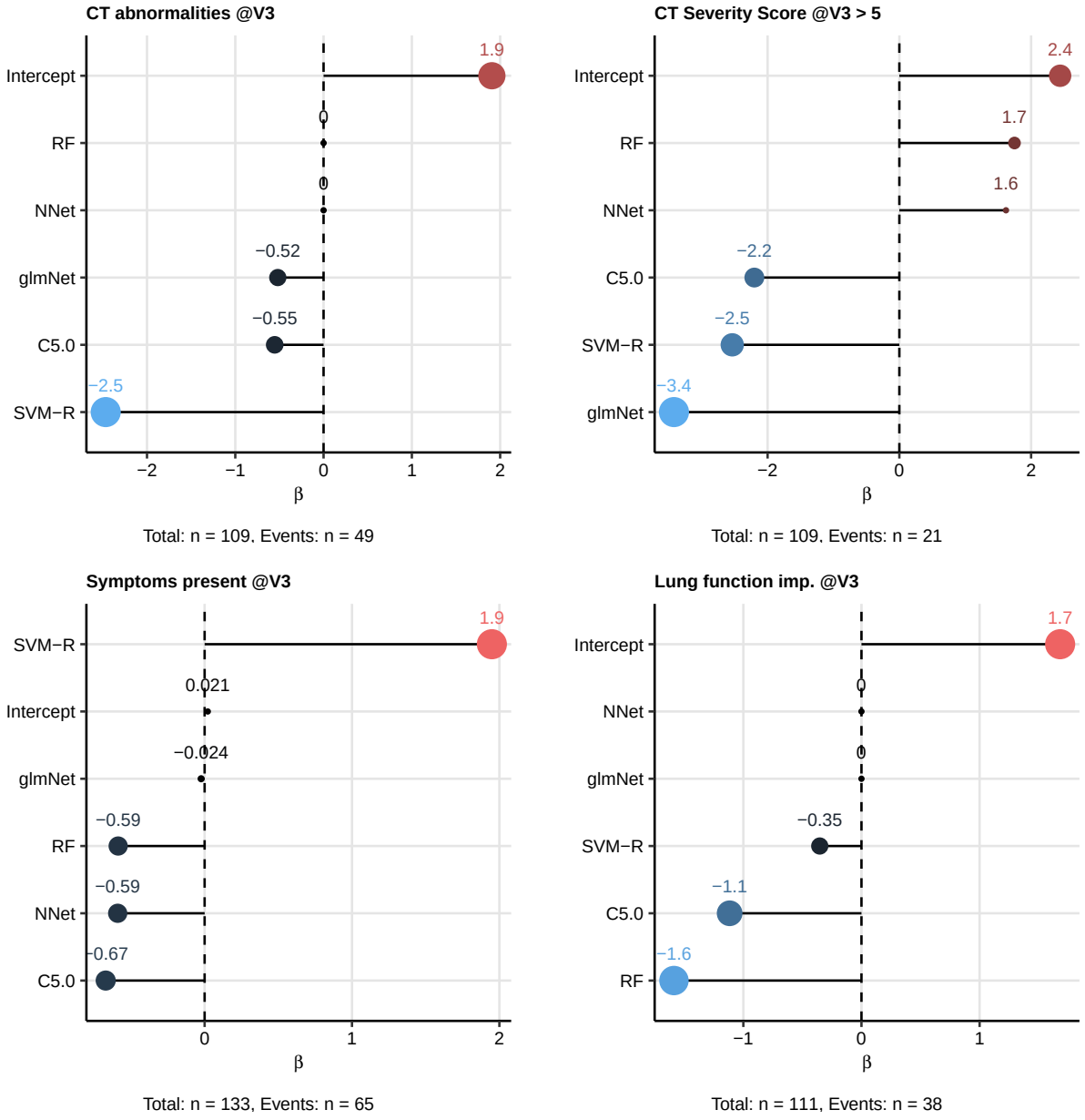

Figure S10: Machine learning model ensembles.

Supplementary Figure S10. Machine learning model ensembles.

Single machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) were trained as shown in **Figure 9** and **Supplementary Figure S9**. The model ensembles based on the single classifiers were constructed with the elastic net (glmNet) procedure (**Supplementary Table S4**). Elastic net regression coefficients (β) are presented in the plots. Point and text color corresponds to the β value. Numbers of complete observations and outcome events are indicated under the plots.

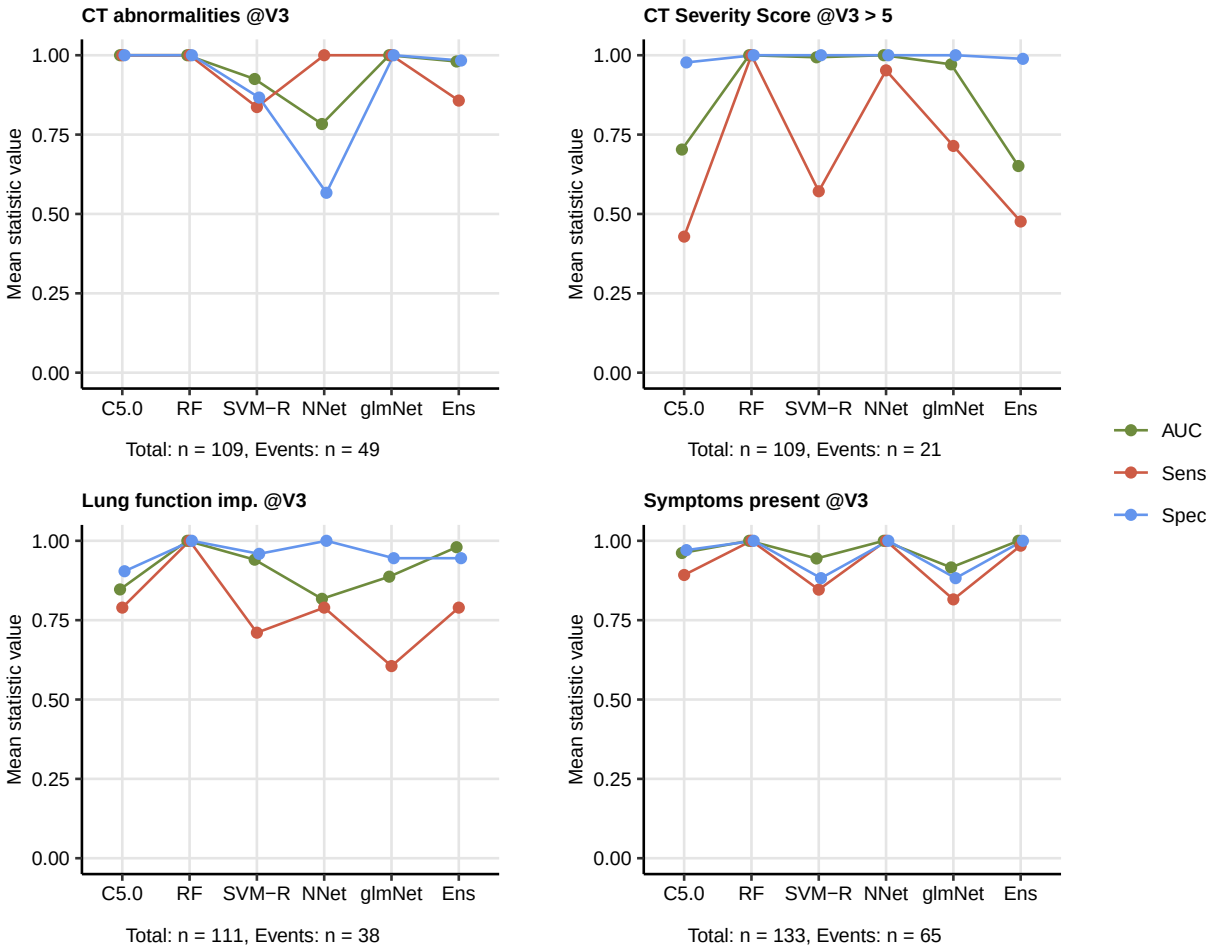

Figure S11: Prediction of persistent radiological lung abnormalities, functional lung impairment and symptoms by machine learning algorithms in the training data sets.

Supplementary Figure S11. Prediction of persistent radiological lung abnormalities, functional lung impairment and symptoms by machine learning algorithms in the training data sets.

Single machine learning classifiers (C5.0, RF: random forests, SVM-R: support vector machines with radial kernel, NNet: neural network, glmNet: elastic net) and their ensembles were trained as shown in **Figure 9** and **Supplementary Figure S9** and **S10**. Performance of the classifiers in the training data sets was investigated by receiver operating characteristic (ROC) of the algorithms (AUC: area under the curve, Sens: sensitivity, Spec: specificity, **Supplementary Table S5**). Numbers of complete observations and outcome events are indicated under the plots.

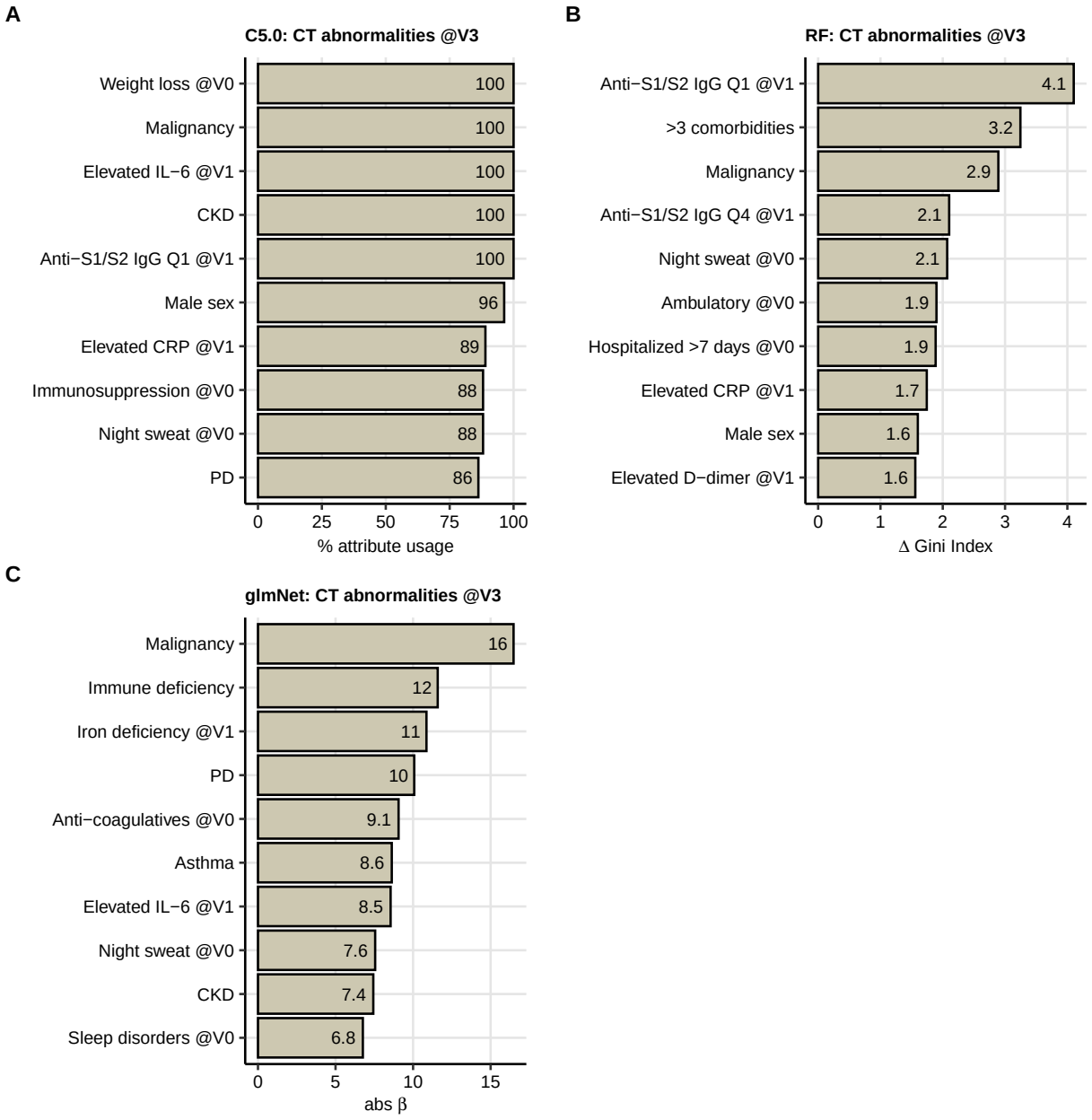

Figure S12: Variable importance statistics for prediction of any lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

Supplementary Figure S12. Variable importance statistics for prediction of any lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were were trained as presented in **Figure 9** and **Supplementary Figure S9** for prediction of any lung CT abnormalities at the 180-day follow-up visit (V3). Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented.

CKD: chronic kidney disease, Q1, Q4: 1^{st}, 4^{th} quartile, PD: pulmonary disease, CKD: chronic kidney disease, V0: acute COVID-19, V1: 60-day follow-up visit.

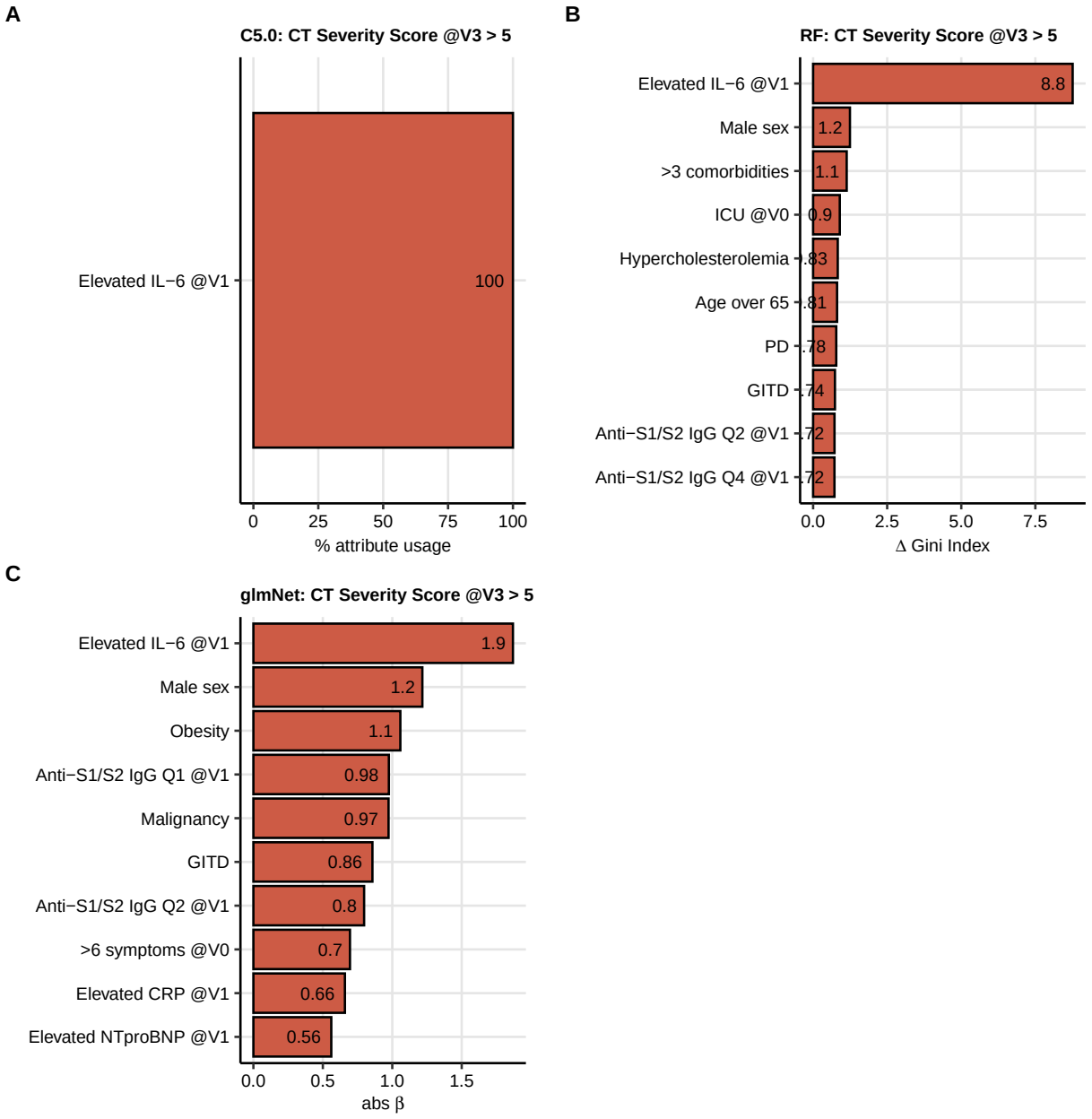

Figure S13: Variable importance statistics for prediction of moderate-to-severe lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

Supplementary Figure S13. Variable importance statistics for prediction of moderate-to-severe lung CT abnormalities at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were were trained as presented in **Figure 9** and **Supplementary Figure S9** for prediction of moderate-to-severe lung CT abnormalities (severity score > 5) at the 180-day follow-up visit (V3). Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented.

PD: pulmonary disease, GITD: gastrointestinal disease, Q1, Q2, Q4: 1^{st}, 2^{nd}, 4^{th} quartile, V0: acute COVID-19, V1: 60-day follow-up visit.

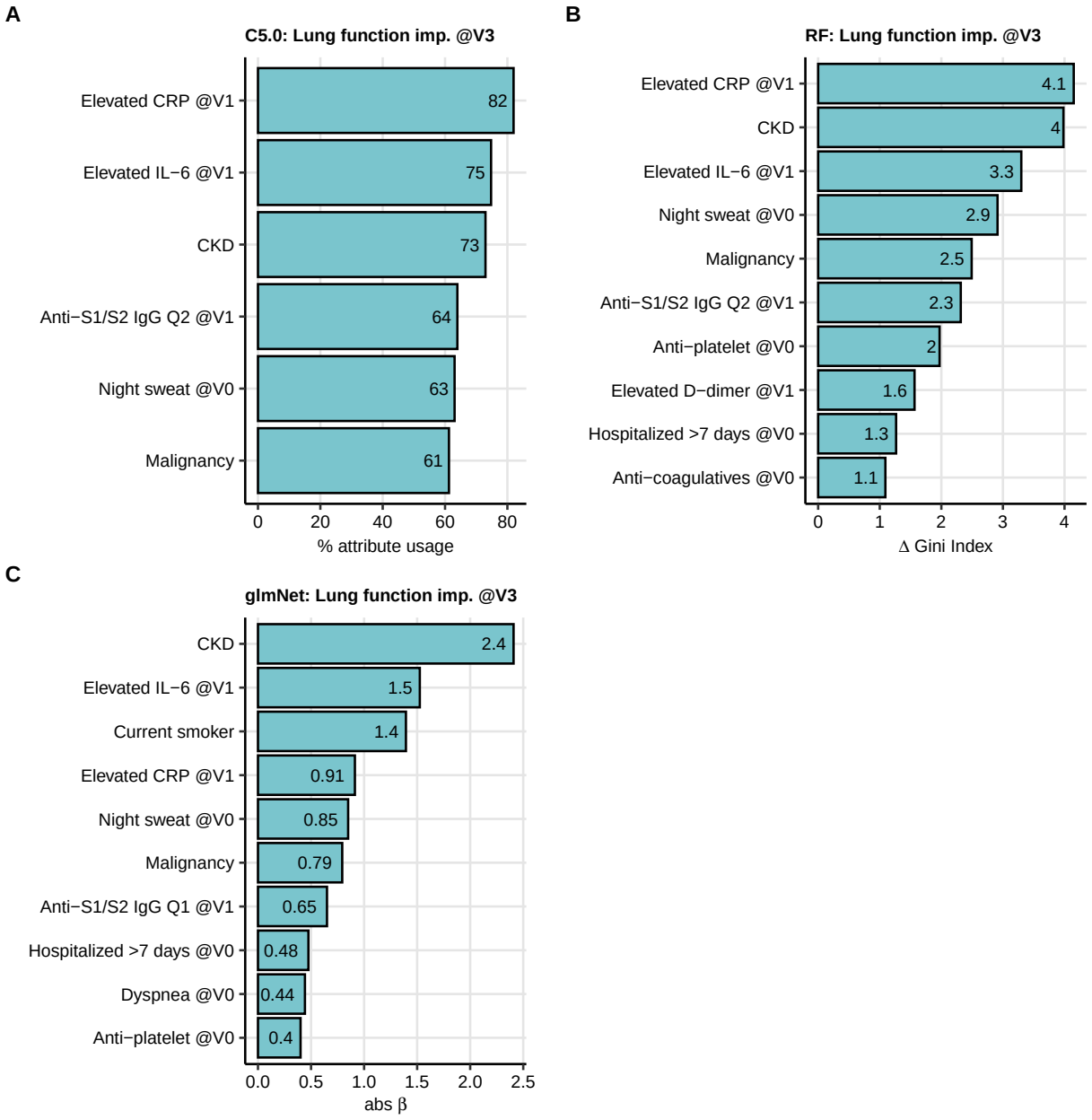

Figure S14: Variable importance statistics for prediction of functional lung impairment at the 180-day follow-up by machine learning classifiers.

Supplementary Figure S14. Variable importance statistics for prediction of functional lung impairment at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were were trained as presented in **Figure 9** and **Supplementary Figure S9** for prediction of functional lung impairment at the 180-day follow-up visit (V3). Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented.

CKD: chronic kidney disease, Q1, Q2: 1^{st}. 2^{nd} quartile, V0: acute COVID-19, V1: 60-day follow-up visit.

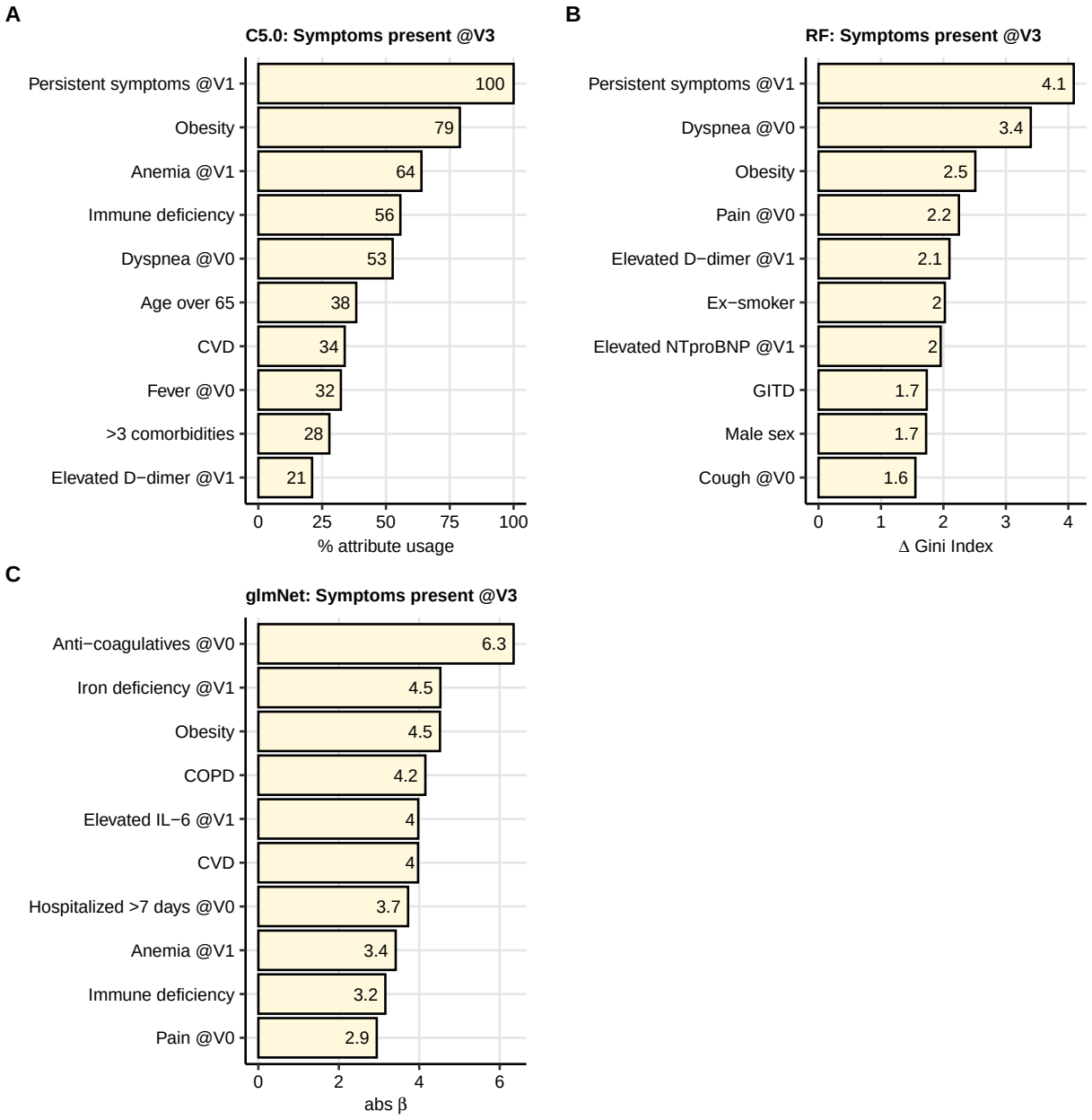

Figure S15: Variable importance statistics for prediction of persistent symptoms at the 180-day follow-up by machine learning classifiers.

Supplementary Figure S15. Variable importance statistics for prediction of persistent symptoms at the 180-day follow-up by machine learning classifiers.

C5.0, random forests (RF) and elastic net (glmNet) classifiers were were trained as presented in **Figure 9** and **Supplementary Figure S9** for prediction of persistent symptoms at the 180-day follow-up visit (V3). Variable importance measures (C5.0: % attribute/variable usage in the tree model, RF: difference in Gini index, glmNet: absolute value of the regression coefficient β) for the 10 most influential explanatory variables are presented.

CVD: cardiovascular disease, GITD: gastrointestinal disease, COPD: chronic obstructive lung disease, V0: acute COVID-19, V1: 60-day follow-up visit.