COVID-19 and its continuing burden after 12 months: a longitudinal observational prospective multicenter trial

Supplementary Material

CovILD study team

Supplementary Methods

Data transformation, variable stratification, descriptive statistic

The study variables were transformed, analyzed and visualized with R version 4.0.5 with *tidyverse* (1,2) and *cowplot* (3).

Participant age during acute COVID-19 was stratified with the 50 and 65 year cutoffs. Participants were stratified according to the severity of acuet COVID-19 as ambulatory (outpatients), moderate (hospitalized at COVID-19 ward, no mechanical ventilation) and severe (mechanical ventilation and/or intensive care unit stay). Lung function testing (LFT) parameters were stratified by 80% predicted value (FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; DLCO: diffusion lung capacity for carbon monoxide; TLC: total lung capacity) or 70% predicted value cutoffs (FEV1:FVC: FEV1 to FVC ratio). Abnormal LFT was diagnosed when at least one of FEV1, FVC, DLCO, RV, TLC or FEV1:FVC was reduced. Chest computed tomography was assessed and rated with the CT severity score as described (4,5); moderate-to-severe CT abnormalities were defined by CT severity score > 5. Anemia was defined as hemoglobin < 140 g/dL or < 120 g/dL for men and women, respectively. Elevated ferritin (FT) was defined as ≥ 300 µg/L or ≥ 150 µg/L for men and women, respectively. Reduced transferrin saturation (TSAT) was defined as < 20% or < 15% for men and women, respectively. Elevated N-terminal pro - brain natriuretic peptide (NT-proBNP) was defined as > 150 pg/mL. Elevated D-dimer, C-reactive protein (CRP), procalcitonin (PCT) and interleukin 6 (IL6) were defined with the 500 µg/L, 0.5 mg/L, 0.15 µg/L and 7 pg/mL cutoffs, respectively. Anti-S1/S2 SARS-CoV-2 immunoglubulin gamma (anti-S1/S2 IgG) was stratified by quartiles (cutoffs: 54.35, 109, 168 arbitrary units) (4). For clustering, the sub-scores of the EQ-5D-5L questionnaire (6) addressing impairment of usual activities, mobility and self-care as well as pain/discomfort, depression/anxiety were stratified with the cutoff of 1 (1: no impairment or absence, > 1:impairment or presence). Visual analogue scale (VAS) of the EQ-5D-5L tool describing the self-perceived general health was binarized with the cutoff of 73.2 as published before for the general elderly German population (7). Elevated stress levels were defined by the median split of the 4-item PSS score (> 5: elevated stress). Reference values of the six-minute walking distance (SMWD) (8) were calculated with the participants sex, age, weight and height as described previously (9). Low SMWD was defined as a value below the patient's specific reference value. The full list of the study variables and their stratification scheme is presented in **Supplementary Tables 1** and **S2**.

Hypothesis testing and correlation

To compare differences in distribution of categorical features, χ^2 test with Cramer V effect size statistic was applied. Since multiple numeric variables were strongly non-normally distributed as identified by Shapiro-Wilk test, Mann-Whitney U test with Wilcoxon r effect

size statistic or Kruskal-Wallis test with η^2 effect size statistic was applied to assess differences between two or more groups, respectively. Association of the readouts of clinical, cardiopulmonary, mobility and quality of life deficits at the one-year follow-up visit were determined by Kendall's τB correlation. Binary variables were recoded as 1 and 2 for absent and present, respectively. P values were corrected for multiple comparisons with Benjamini-Hochberg method (10). R packages *rstatix*, *rcompanion* and *Hmisc* and the in-house developed package *ExDA* (https://github.com/PiotrTymoszuk/ExDA) were used for explorative data analysis, statistical hypothesis testing and correlation analysis.

Modeling of risk and paramater value in time

To model recovery kinetics for categorical variables, second-order mixed-effect logistic (categorical features) modeling was applied (packages: *lme4* and *lmerTest*) (11–13). Each model followed the general formula:

$$Response \sim time + time^2 + (1 \vee individual)$$

where $(1 \vee individual)$ indicates the random effect of the individual and $time$ and $time^2$ indicate the first- and second-order time effect terms. The first-order term estimate was interpreted as a measure of the recovery speed and the second-order term estimate was used to assess the plateau/rebound effect. Significance of the accuracy gain of the full second-order model compared with the nested null model was determined by likelihood ratio test (LRT) versus the nested first-order and null models, respectively. Likelihood ratio λ statistic was used as an effect size measure.

To assess the longitudinal changes of a numeric dependent variable in time, Friedman test with Kendall's W effect was used (packages: *rstatix* and *ExDA*: https://github.com/PiotrTymoszuk/ExDA). Results of the kinetic modeling and Friedman tests were adjusted for multiple comparisons with Benjamini-Hochberg method (10).

Clustering analysis

To identify patterns of COVID-19 recovery defined by 19 binary symptom, cardiopulmonary and psychosocial variables recorded at the 1-year follow-up visit (**Supplementary Tables S1** and **S2**) the subset of the study participants with the complete variable record was clustered using the PAM (partition around medoids) algorithm and the simple matching distance statistic (14,15). The data pre-processing included conversion of binary features to the numeric format (absent: 1, present: 1). The choice of the clustering procedure was motivated by the analysis of the clustering variance (ratio of the total between-cluster to total sum of squares) and clustering structure stability in 10-fold cross-validation (metric: rate of correct cluster assignment, cluster assignment predicted by k = 5 nearest neighbors label propagation algorithm) (16,17) for several clustering algorithms as presented in **Supplementary Figure S4A**. The optimal number of clusters was determined by the bend of the total within-cluster sum of squares curve (**Supplementary Figure S4B**).

The clustering analysis was accomplished with packages *cluster*, *dbscan* and the in-house developed package *clustTools* (https://github.com/PiotrTymoszuk/clustTools).

Univariable modeling

Association of 34 candidate co-variates recorded during acute COVID-19 or at the 60-day follow-up (**Supplementary Table S1**) with the risk of persistent COVID-19 symptoms, LFT or CT abnormalities or diastolic dysfunction at the 1-year follow-up was investigated with a series of univariable logistic models (**Supplementary Table S6**). Model estimate (odds ratio, OR) significance was assessed by Wald Z test and p values were corrected for multiple comparisons with Benjamini-Hochberg method (10). Model estimate extraction and quality control was accomplished with the in-house-developed package *lmqc* (https://github.com/PiotrTymoszuk/lmqc).

Multi-parameter modeling

The risk of persistent COVID-19 symptoms, LFT or CT abnormalities or diastolic dysfunction at the 1-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression (18) and 34 independent variables recorded during acute COVID-19 and at the 60-day follow-up (**Supplementary Table S1**). The value of the λ shrinkage parameter were determined by 10-fold cross-validation (CV) repeated 100 times; the optimal value for the model with the lowest deviance was chosen (19). The model performance at predicting the 1-year event in the training data set and in 10-fold CV was investigated by receiver-operating characteristic (ROC) (20), Cohen's κ (21) and R-squared statistics. Model assumption control was accomplished by a visual control of model residuals plots (residuals versus fitted and quantile-quantile plots). The multi-parameter modeling, model evaluation and model feature visualization tasks was accomplished with *glmnet* (19), *caret* (20), *plotroc* packages (22) and the in-house development package *caretExtra* (https://github.com/PiotrTymoszuk/caretExtra).

Source code availability

The raw study data will be made available upon request. The R analysis pipeline is available at https://github.com/PiotrTymoszuk/CovILD-Plus.

Supplementary Tables

Supplementary Table S1: Study variables.

R variable	Label	Collection time points	Independent modeling variable	Correlation variable	Clustering variable
smwd	SMWD, m	1-year FUP	no	no	no
smwd_dref	SMWD vs ref., m		no	yes	no
smwd_low	SMWD < ref., %		no	no	yes
BRCS	Resilience (BRCS)		no	no	no
BRCS_class	Resilience (BRCS)		no	no	no
EQ5DL_p	General health (EQ5D5L VAS), %		no	yes	no
EQ5DL_mobility	Mobility impairment (EQ5D5L)		no	yes	no
EQ5DL_selfcare	Self-care impairment (EQ5D5L)		no	yes	no
EQ5DL_activities	Activity impairment (EQ5D5L)		no	yes	no
EQ5DL_pain	Pain/discomfort (EQ5D5L)		no	yes	no
EQ5DL_anxiety	Anxiety/depression (EQ5D5L)		no	yes	no
EQ5DL_low	Imp. general health (VAS < 73, EQ5D5L), %		no	no	yes
EQ5DL_mobility_bi	Imp. mobility (EQ5D5L), %		no	no	yes
EQ5DL_selfcare_bi	Imp. self-care (EQ5D5L), %		no	no	yes
EQ5DL_activities_bi	Imp. usual activity (EQ5D5L), %		no	no	yes
EQ5DL_pain_bi	Pain/discomfort present (EQ5D5L), %		no	no	yes
EQ5DL_anxiety_bi	Anxiety/depression present (EQ5D5L), %		no	no	yes
Chalder_FS	Fatigue (likert CFS)		no	yes	no

R variable	Label	Collection time points	Independent modeling variable	Correlation variable	Clustering variable
Chalder_FS_bimodal	Fatigue (bimodal CFS ≥ 4), %		no	no	yes
SSD12	Somatic symptom disorder (SSD-12)		no	no	no
Stress	Stress (PSS)		no	yes	no
Stress_hi	Elevated stress (PSS > 5), %		no	no	yes
SES	Self-efficacy (GSES)		no	no	no
KW_IPQ	Illness perception (IPQ)		no	no	no
SOCL9	Sense of coherence loss (SOCL-9)		no	no	no
FVC_p	FVC, %	60- till 1 year FUP	no	no	no
FEV1_p	FEV1, %		no	no	no
FEV1_FVC_p	FEV1:FVC, %		no	no	no
DLCO_p	DLCO, %		no	yes	no
TLC_p	TLC, %		no	no	no
FVC_red	reduced FVC, %		no	no	no
FEV1_red	reduced FEV1, %		no	no	no
FEV1_FVC_red	reduced FEV1:FVC, %		no	no	no
TLC_red	reduced TLC, %		no	no	no
DLCO_red	reduced DLCO, %		no	no	no
lufo_red	LFT abnormality, %		no	yes	yes
Hb	Hb, g/dL		no	no	no
anemia	Anemia, %		yes	no	no
ferritin	FT, µg/L		no	no	no
FT_elv	elevated FT, %		yes	no	no
TSAT	TF-Sat, %		no	no	no
TSAT_red	reduced TF-Sat, %		no	no	no

R variable	Label	Collection time points	Independent modeling variable	Correlation variable	Clustering variable
sTFR	sTFR, mg/L		no	no	no
Hepcidin	Hepcidin, ng/mL		no	no	no
NTproBNP	NT-proBNP, pg/mL		no	no	no
NtproBNP_elv	elevated NT-proBNP, %		yes	no	no
DDimer	D-dimer, µg/L		no	no	no
Ddimer_elv	elevated D-dimer, %		yes	no	no
CRP	CRP, mg/L		no	no	no
CRP_elv	elevated CRP, %		yes	no	no
PCT	PCT, µg/L		no	no	no
PCT_elv	elevated PCT, %		no	no	no
IL6	IL6, pg/mL		no	no	no
IL6_elv	elevated IL6, %		yes	no	no
HbA1c	HbA1c, %		no	no	no
HbA1c_elv	elevated HbA1c, %		no	no	no
EF_red	reduced LVEF, %		no	no	no
diastolic_dysf	Diastolic dysfunction, %		no	yes	yes
mmrc	mMRC		no	yes	no
ECOG	ECOG		no	no	no
ct_severity_score	CT severity score		no	yes	no
ct_severity_any	CT abnormality, %		no	no	yes
ct_severity_5	CT severity score > 5, %		no	no	no
ab_quant	anti-S1/S2 IgG	60-day FUP	yes	no	no
sex	Sex	acute COVID-19	yes	no	no
age	Age, years		yes	no	no

R variable	Label	Collection time points	Independent modeling variable	Correlation variable	Clustering variable
cat_WHO	COVID-19 severity		yes	no	no
weight_class	Weight class		yes	no	no
smoking	Smoking		yes	no	no
no_comorb	# comorbidities		yes	no	no
comorb_present	Comorbidity present, %		no	no	no
cardiovascular_comorb	CVD, %		yes	no	no
hypertension_comorb	Hypertension, %		yes	no	no
pulmonary_comorb	Pulmonary disease, %		yes	no	no
copd_comorb	COPD, %		yes	no	no
asthma_comorb	Asthma, %		yes	no	no
intenst_lung_comorb	ILD, %		no	no	no
endometabolic_comorb	Metabolic disease, %		yes	no	no
hyperchol_comorb	Hypercholesterolemia, %		yes	no	no
diabetes_comorb	Diabetes, %		yes	no	no
ckd_comorb	CKD, %		yes	no	no
gastro_comorb	GID, %		yes	no	no
cldis_comorb	CLD, %		no	no	no
malingancy_comorb	Malignancy, %		yes	no	no
immdef_comorb	Immune deficiency, %		yes	no	no
sympt_present	Symptoms present, %	acute COVID-19 till 1-year FUP	no	no	yes
sympt_number	# symptoms		yes	yes	no

R variable	Label	Collection time points	Independent modeling variable	Correlation variable	Clustering variable
sleep_sympt	Sleep problems, %		yes	yes	yes
dyspnoe_sympt	Dyspnea, %		yes	no	yes
cough_sympt	Cough, %		yes	yes	yes
fever_sympt	Fever, %		no	no	no
night_sweat_sympt	Night sweat, %		yes	yes	yes
gastro_sympt	Gastrointestinal, %		yes	no	no
anosmia_sympt	Hypo/anosmia, %		yes	yes	yes
fatigue_sympt	Reduced performance, %		yes	yes	yes
pain_sympt	Pain, %		yes	no	no
hair_loss_sympt	Hair loss, %		no	no	no
derma_sympt	Dermatological, %		no	no	no
rehabilitation	Rehabilitation, %		no	no	no

Supplementary Table S2: Physical performance, fatigue, quality of life, psychosocial and mental health assessment battery.

Evaluated parameter	Tool	Stratification scheme
Dyspnea	Modified Medical British Research Council (mMRC)	Dyspnea: > 0
Physical performance	Eastern Cooperative Oncology Group (ECOG)	Reduced performance: > 0
Mobility	Six-minute walking distance (SMWD)	
	Six-minute walking distance versus reference (SMWD vs ref.)	Impaired mobility: < 0
Fatigue	Likert 11-item Chalder Fatigue Score (CFS)	
	Bimodal 11-item Chalder Fatigue Score (CFS)	Increased fatigue: ≥ 4
Self-perceived general health	European quality of life 5 dimensions, visual analogue scale (EQ5D5L VAS)	Impaired general health: < 73
Mobility impairment	European quality of life 5 dimensions, mobility sub-score (EQ5D5L mobility)	Impaired mobility: > 1
Self-care impairment	European quality of life 5 dimensions, self-care sub-score (EQ5D5L self-care)	Impaired self care: > 1
Usual activity impairment	European quality of life 5 dimensions, activity sub-score (EQ5D5L activity)	Impaired activity: > 1
Pain/discomfort	European quality of life 5 dimensions, pain/discomfort sub-score (EQ5D5L pain/discomfort)	Pain/discomfort present: > 1
Anxiety/depression	European quality of life 5 dimensions, anxiety/depression sub-score (EQ5D5L anxiety/depression)	Anxiety/depression present: > 1
Resilience	Brief resilient coping scale (BRCS)	
Somatic symptom disorder	Somatic symptom disorder – B criteria scale, 12 items (SSD-12)	
Psychosocial stress	4-item perceived stress score (PSS)	Elevated stress: > 5

Supplementary Table S3: Demographic and clinical characteristics of the COVID-19 recovery clusters.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
N number	36	33	18		
Sex	female: 42% (n = 15) male: 58% (n = 21)	female: 18% (n = 6) male: 82% (n = 27)	female: 72% (n = 13) male: 28% (n = 5)	p = 0.0025	V = 0.41
Age, years	median: 55 [IQR: 53 - 61] range: 19 - 77	median: 63 [IQR: 53 - 71] range: 36 - 80	median: 50 [IQR: 42 - 58] range: 32 - 87	ns (p = 0.059)	$\eta^2 = 0.061$
Weight class	normal: 22% (n = 8) overweight: 56% (n = 20) obesity: 22% (n = 8)	normal: 36% (n = 12) overweight: 39% (n = 13) obesity: 24% (n = 8)	normal: 44% (n = 8) overweight: 44% (n = 8) obesity: 11% (n = 2)	ns (p = 0.54)	V = 0.15
Smoking	never: 64% (n = 23) ex: 36% (n = 13) active: 0% (n = 0)	never: 61% (n = 20) ex: 36% (n = 12) active: 3% (n = 1)	never: 72% (n = 13) ex: 17% (n = 3) active: 11% (n = 2)	ns (p = 0.32)	V = 0.19
Comorbidity present	72% (n = 26)	82% (n = 27)	67% (n = 12)	ns (p = 0.58)	V = 0.14
Metabolic disease	39% (n = 14)	58% (n = 19)	28% (n = 5)	ns (p = 0.18)	V = 0.23
Diabetes	14% (n = 5)	15% (n = 5)	17% (n = 3)	ns (p = 0.96)	V = 0.029
Hypercholesterolemia	31% (n = 11)	30% (n = 10)	0% (n = 0)	ns (p = 0.058)	V = 0.29
CVD[b]	36% (n = 13)	61% (n = 20)	22% (n = 4)	p = 0.042	V = 0.3
Pulmonary disease	11% (n = 4)	24% (n = 8)	17% (n = 3)	ns (p = 0.51)	V = 0.15
Malignancy	8.3% (n = 3)	15% (n = 5)	0% (n = 0)	ns (p = 0.34)	V = 0.19
Immune deficiency	2.8% (n = 1)	3% (n = 1)	11% (n = 2)	ns (p = 0.49)	V = 0.16
CKD[c]	5.6% (n = 2)	12% (n = 4)	5.6% (n = 1)	ns (p = 0.68)	V = 0.12
GID[d]	11% (n = 4)	15% (n = 5)	11% (n = 2)	ns (p = 0.9)	V = 0.059

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
Rehabilitation	no: 86% (n = 31) inpatient: 11% (n = 4) outpatient: 2.8% (n = 1)	no: 58% (n = 19) inpatient: 30% (n = 10) outpatient: 12% (n = 4)	no: 44% (n = 8) inpatient: 50% (n = 9) outpatient: 5.6% (n = 1)	p = 0.025	V = 0.28
COVID-19 severity	Ambulatory: 33% (n = 12) Moderate: 50% (n = 18) Severe: 17% (n = 6)	Ambulatory: 12% (n = 4) Moderate: 67% (n = 22) Severe: 21% (n = 7)	Ambulatory: 33% (n = 6) Moderate: 17% (n = 3) Severe: 50% (n = 9)	p = 0.012	V = 0.3

[a]Categorical variables: χ^2 test with Cramer V effect size statistic; numeric variables: Kruskal-Vallis test with η^2 effect size statistic. P values corrected for multiple testing with Benjamini-Hochberg method.

[b]CVD: Cardiovascular disease.

[c]CKD: Chronic kidney disease

[d]GID: Gastrointestinal disease

Supplementary Table S4: Symptoms and cardiopulmonary abnormalities at the 1-year follow-up in the COVID-19 recovery clusters.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
N number	36	33	18		
Symptoms present	28% (n = 10)	85% (n = 28)	100% (n = 18)	$p < 0.001$	$V = 0.65$
# symptoms	median: 0 [IQR: 0 - 1] range: 0 - 4	median: 1 [IQR: 1 - 2] range: 0 - 5	median: 4 [IQR: 3.2 - 4.8] range: 1 - 9	$p < 0.001$	$\eta^2 = 0.49$
Sleep problems	11% (n = 4)	24% (n = 8)	67% (n = 12)	$p < 0.001$	$V = 0.47$
Dyspnea	5.6% (n = 2)	18% (n = 6)	72% (n = 13)	$p < 0.001$	$V = 0.59$
Cough	5.6% (n = 2)	9.1% (n = 3)	56% (n = 10)	$p < 0.001$	$V = 0.52$
Night sweat	11% (n = 4)	21% (n = 7)	28% (n = 5)	ns ($p = 0.43$)	$V = 0.17$
Gastrointestinal	2.8% (n = 1)	0% (n = 0)	22% (n = 4)	$p = 0.0091$	$V = 0.37$
Hypo/ anosmia	14% (n = 5)	15% (n = 5)	11% (n = 2)	ns ($p = 0.95$)	$V = 0.043$
Reduced performance	5.6% (n = 2)	48% (n = 16)	83% (n = 15)	$p < 0.001$	$V = 0.62$
Hair loss	0% (n = 0)	3% (n = 1)	33% (n = 6)	$p < 0.001$	$V = 0.48$
Dermatological	2.8% (n = 1)	21% (n = 7)	28% (n = 5)	ns ($p = 0.052$)	$V = 0.29$
LFT abnormality[b]	25% (n = 9)	42% (n = 14)	44% (n = 8)	ns ($p = 0.37$)	$V = 0.19$
CT abnormality (CT score ≥ 1)[c]	25% (n = 9)	85% (n = 28)	44% (n = 8)	$p < 0.001$	$V = 0.54$
Diastolic dysfunction	58% (n = 21)	70% (n = 23)	56% (n = 10)	ns ($p = 0.65$)	$V = 0.12$

[a]Categorical variables: χ^2 test with Cramer V effect size statistic; numeric variables: Kruskal-Vallis test with η^2 effect size statistic. P values corrected for multiple testing with Benjamini-Hochberg method.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]

[b]LFT abnormality: abnormality in lung function testing, > 80% predicted value (FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; DLCO: diffusion lung capacity for carbon monoxide; TLC: total lung capacity) or > 70% predicted value cutoffs (FEV1:FVC: FEV1 to FVC ratio.)

[c]CT abnormality: any abnormality in chest computed tomography, CT severity score ≥ 1.

Supplementary Table S5: Mobility, physical performance and psychosocial rating at the 1-year follow-up in the COVID-19 recovery clusters.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
N number	36	33	18		
SMWD, m[b]	median: 580 [IQR: 530 - 640] range: 460 - 740	median: 520 [IQR: 450 - 580] range: 270 - 760	median: 530 [IQR: 420 - 620] range: 310 - 710	p = 0.025	η^2 = 0.085
SMWD vs ref., m[c]	median: 38 [IQR: 5.1 - 62] range: -220 - 130	median: -44 [IQR: -79 - 19] range: -220 - 120	median: -72 [IQR: -100 - -9.3] range: -230 - 140	p = 0.0022	η^2 = 0.15
SMWD < ref.[d]	19% (n = 7)	73% (n = 24)	78% (n = 14)	p < 0.001	V = 0.54
Fatigue score (likert CFS)[e]	median: 11 [IQR: 11 - 12] range: 1 - 24	median: 13 [IQR: 10 - 15] range: 1 - 25	median: 20 [IQR: 16 - 26] range: 0 - 32	p < 0.001	η^2 = 0.24
Fatigue (bimodal CFS ≥ 4)[e]	22% (n = 8)	36% (n = 12)	72% (n = 13)	p = 0.0056	V = 0.38
General health score (EQ5D5L VAS)[f]	median: 90 [IQR: 80 - 95] range: 70 - 100	median: 80 [IQR: 80 - 90] range: 60 - 100	median: 70 [IQR: 61 - 75] range: 40 - 90	p < 0.001	η^2 = 0.26
Imp. general health (VAS < 73, EQ5D5L)[f]	8.3% (n = 3)	9.1% (n = 3)	61% (n = 11)	p < 0.001	V = 0.54
Mobility impairment score (EQ5D5L)	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 2] range: 1 - 3	p = 0.02	η^2 = 0.093
Imp. mobility (score > 1, EQ5D5L)[g]	2.8% (n = 1)	15% (n = 5)	33% (n = 6)	p = 0.023	V = 0.33
Self-care impairment score (EQ5D5L)	median: 1 [IQR: 1 - 1] range: 1 - 1	median: 1 [IQR: 1 - 1] range: 1 - 1	median: 1 [IQR: 1 - 1] range: 1 - 2	p = 0.0087	η^2 = 0.12
Imp. self-care (score > 1, EQ5D5L)[g]	0% (n = 0)	0% (n = 0)	17% (n = 3)	p = 0.0084	V = 0.37

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
Activity impairment score (EQ5D5L)	median: 1 [IQR: 1 - 1] range: 1 - 1	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 2 [IQR: 2 - 3] range: 1 - 3	$p < 0.001$	$\eta^2 = 0.73$
Imp. usual activity (score > 1, EQ5D5L)[g]	0% (n = 0)	3% (n = 1)	83% (n = 15)	$p < 0.001$	V = 0.86
Pain/ discomfort score (EQ5D5L)	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 1] range: 1 - 3	median: 2.5 [IQR: 2 - 3] range: 1 - 4	$p < 0.001$	$\eta^2 = 0.39$
Pain/ discomfort (score > 1, EQ5D5L)[g]	22% (n = 8)	21% (n = 7)	89% (n = 16)	$p < 0.001$	V = 0.57
Anxiety/ depression score (EQ5D5L)	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 2 [IQR: 2 - 3.8] range: 1 - 5	$p < 0.001$	$\eta^2 = 0.5$
Anxiety/ depression (score > 1, EQ5D5L)[g]	5.6% (n = 2)	15% (n = 5)	83% (n = 15)	$p < 0.001$	V = 0.69
Stress score (PSS)[h]	median: 5 [IQR: 1 - 8] range: 0 - 10	median: 5 [IQR: 3 - 8] range: 0 - 11	median: 9 [IQR: 8.2 - 10] range: 5 - 11	$p < 0.001$	$\eta^2 = 0.27$
Elevated stress (PSS > 5)[h]	39% (n = 14)	42% (n = 14)	94% (n = 17)	$p < 0.001$	V = 0.44
Somatic symptom disorder score (SSD-12)[i]	median: 4 [IQR: 1 - 7.5] range: 0 - 24	median: 5 [IQR: 2 - 10] range: 0 - 26	median: 20 [IQR: 12 - 24] range: 7 - 30	$p < 0.001$	$\eta^2 = 0.29$
Resilience score (BRCS)[j]	median: 16 [IQR: 12 - 19] range: 4 - 20	median: 16 [IQR: 12 - 18] range: 4 - 20	median: 16 [IQR: 14 - 18] range: 9 - 19	ns ($p = 0.96$)	$\eta^2 = -0.022$

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
Resilience (BRCS)[k]	low: 35% (n = 12) medium: 24% (n = 8) high: 41% (n = 14) n = 34	low: 31% (n = 10) medium: 19% (n = 6) high: 50% (n = 16) n = 32	low: 17% (n = 3) medium: 39% (n = 7) high: 44% (n = 8) n = 18	ns (p = 0.6)	V = 0.15

[a]Categorical variables: χ^2 test with Cramer V effect size statistic; numeric variables: Kruskal-Vallis test with η^2 effect size statistic. P values corrected for multiple testing with Benjamini-Hochberg method.

[b]SMWD: six-minute walking distance, meters.

[c]SMWD vs ref.: difference between the reference and observed SMWD value, meters.

[d]SMWD < ref.: SMWD below the reference value.

[e]CFS: 11-item Chalder fatigue score, incr.: increased.

[f]VAS: visual analogue scale, imp.: impaired.

[g]EQ5D5L subscore > 1, imp.: impaired.

[h]PSS: 4-item perceived stress scale.

[i]SSD-12: 12-item somatic syndrome disorder – B criteria scale.

[j]BRCS: brief resilient coping score.

[k]low: 4 - 13 points, medium: 14 - 16 points, high: 17 - 21 points in the BRCS scale.

Supplementary Table S6: Significant results of univariable modeling of the risk of symptom presence, lung function abnormalities, radiological chest abnormalities and diastolic dysfunction at the 1-year follow-up.

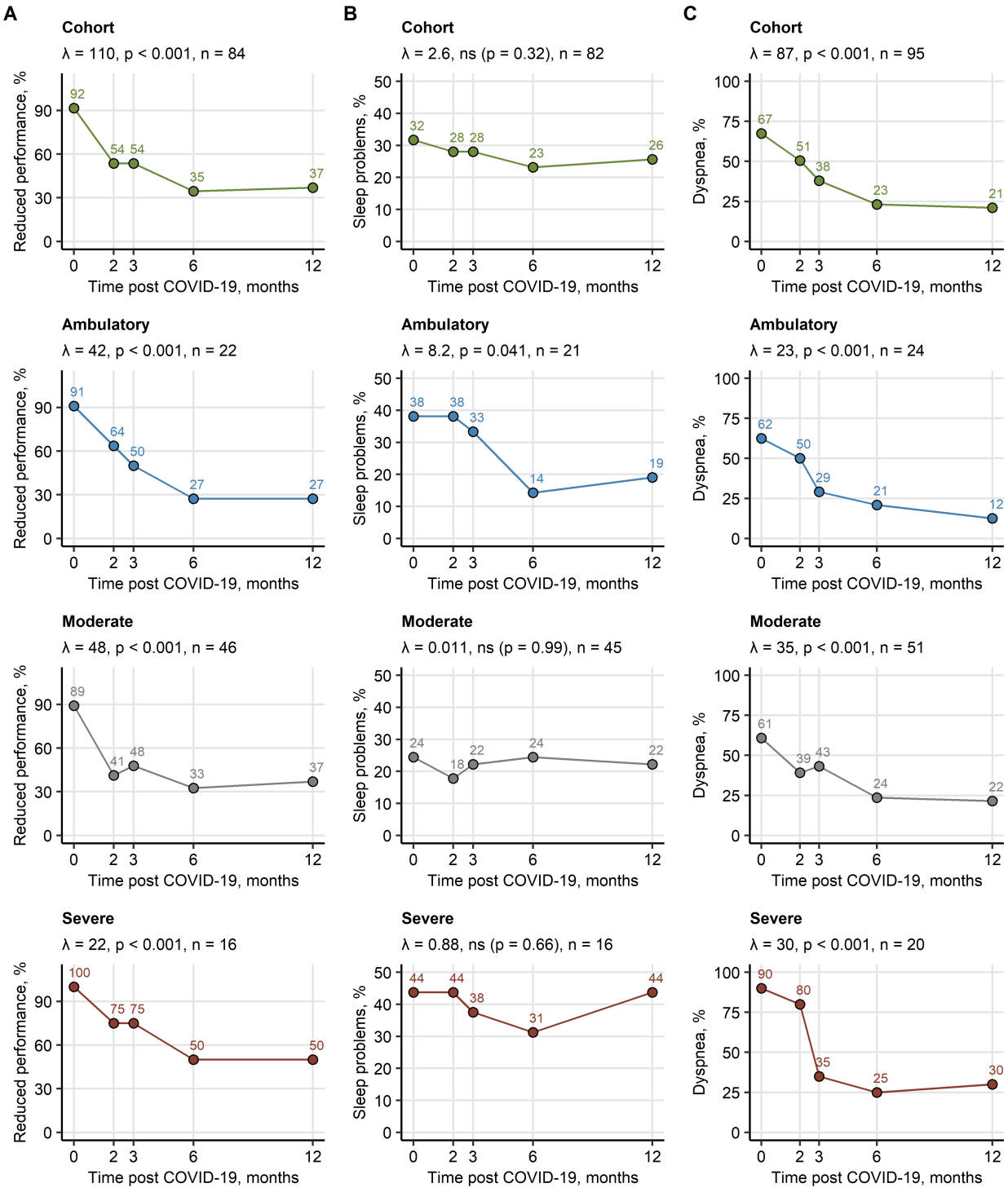
Response	Independent variable	Reference	Level	N level	N total	OR, 95% CI[1]	Significance[2]
CT abnormality, 1 year FUP	COVID-19 severity, acute COVID-19	Ambulatory	Moderate	52	101	7.2 [2.1 - 33]	p = 0.027
		Ambulatory	Severe	26	101	37 [8.3 - 220]	p < 0.001
	Sex, acute COVID-19	female	male	62	101	6.1 [2.6 - 15]	p = 0.0019
	Age, acute COVID-19	up to 50	51 - 65	45	101	3 [1.1 - 8.5]	ns (p = 0.13)
		up to 50	over 65	29	101	4.5 [1.5 - 15]	p = 0.043
	# comorbidities, acute COVID-19		per item	101	101	1.5 [1.2 - 2]	p = 0.0045
	Diabetes, acute COVID-19	no	yes	16	101	8.7 [2.2 - 57]	p = 0.032
	Hypertension, acute COVID-19	no	yes	28	101	4.1 [1.6 - 11]	p = 0.027
	Anemia, 60-day FUP	no	yes	25	101	5.5 [2 - 18]	p = 0.023
	anti-S1/S2 IgG, 60-day FUP	Q1	Q2	26	93	22 [3.7 - 420]	p = 0.027
		Q1	Q3	21	93	36 [5.8 - 710]	p = 0.02
		Q1	Q4	23	93	230 [29 - 5600]	p < 0.001
LFT abnormality, 1 year FUP	# comorbidities, acute COVID-19		per item	106	106	1.5 [1.2 - 2]	p = 0.0013
	elevated IL6, 60-day FUP	no	yes	10	106	10 [2.4 - 71]	p = 0.015
	elevated D-dimer, 60-day FUP	no	yes	38	106	3.2 [1.4 - 7.7]	p = 0.019
Diastolic dysfunction, 1 year FUP	COVID-19 severity, acute COVID-19	Ambulatory	Moderate	54	107	5.2 [1.9 - 15]	p = 0.02
		Ambulatory	Severe	26	107	13 [3.7 - 57]	p = 0.0046
	Age, acute COVID-19	up to 50	51 - 65	46	107	11 [3.9 - 33]	p < 0.001
		up to 50	over 65	29	107	11 [3.7 - 41]	p = 0.0026

Response	Independent variable	Reference	Level	N level	N total	OR, 95% CI[1]	Significance[2]
Symptoms present, 1 year FUP	# comorbidities, acute COVID-19		per item	107	107	1.5 [1.2 - 1.9]	p = 0.014
	anti-S1/S2 IgG, 60-day FUP	Q1	Q2	28	99	2.9 [0.98 - 9.2]	ns (p = 0.14)
		Q1	Q3	22	99	8.5 [2.4 - 37]	p = 0.024
		Q1	Q4	23	99	6.8 [2 - 27]	p = 0.035
	# symptoms, acute COVID-19		per item	104	104	1.4 [1.1 - 1.8]	p = 0.041
	Sleep problems, acute COVID-19	no	yes	40	104	4.2 [1.7 - 11]	p = 0.041
	elevated NT-proBNP, 60-day FUP	no	yes	38	105	3.6 [1.4 - 9.9]	p = 0.043

[1]Odds ratio with 95% confidence intervals.

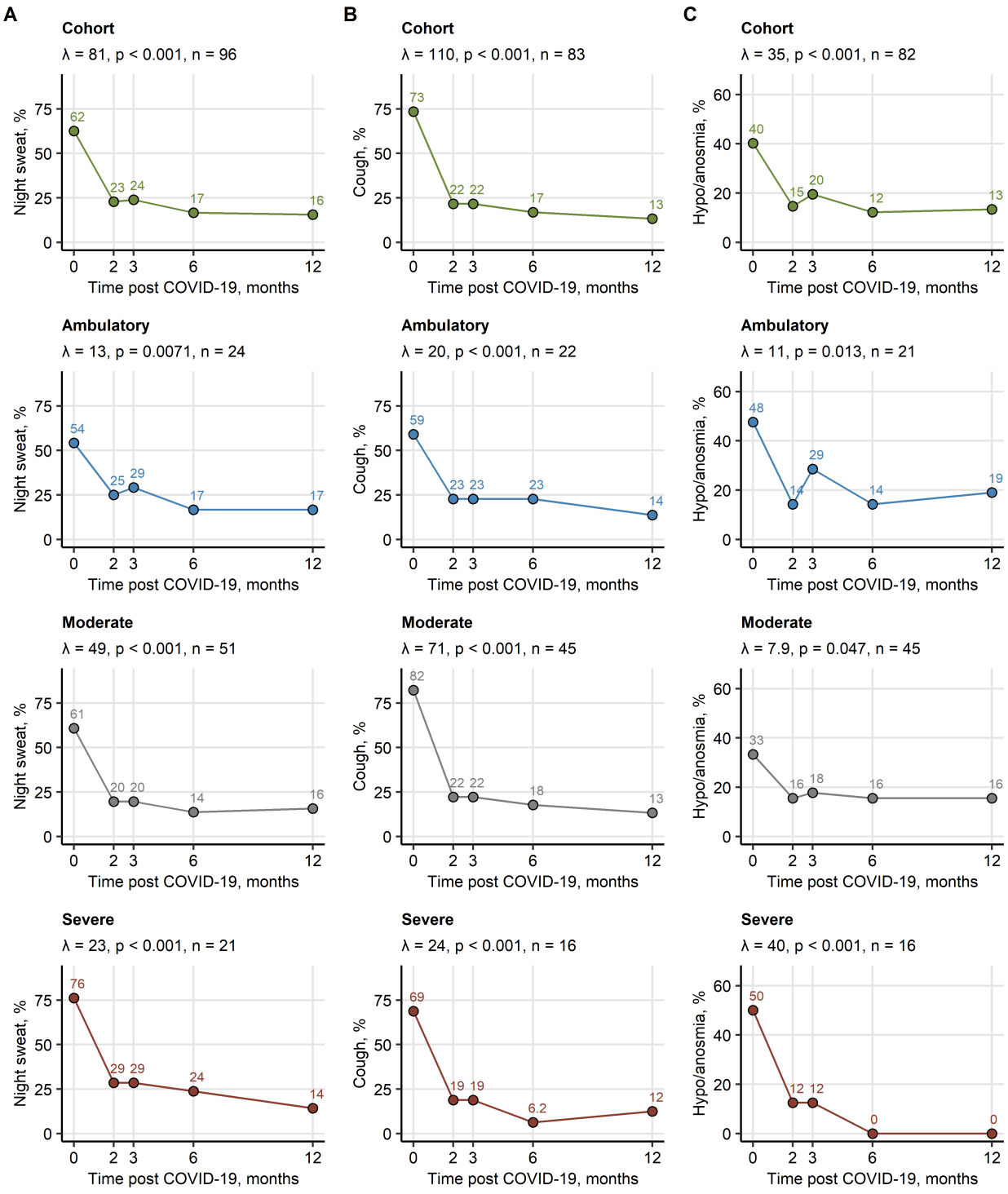
[2]Wald Z test. P values corrected for multiple testing with Benjamini-Hochberg method.

Supplementary Figures



Supplementary Figure S1. Recovery of fatigue, sleep problems and dyspnea.

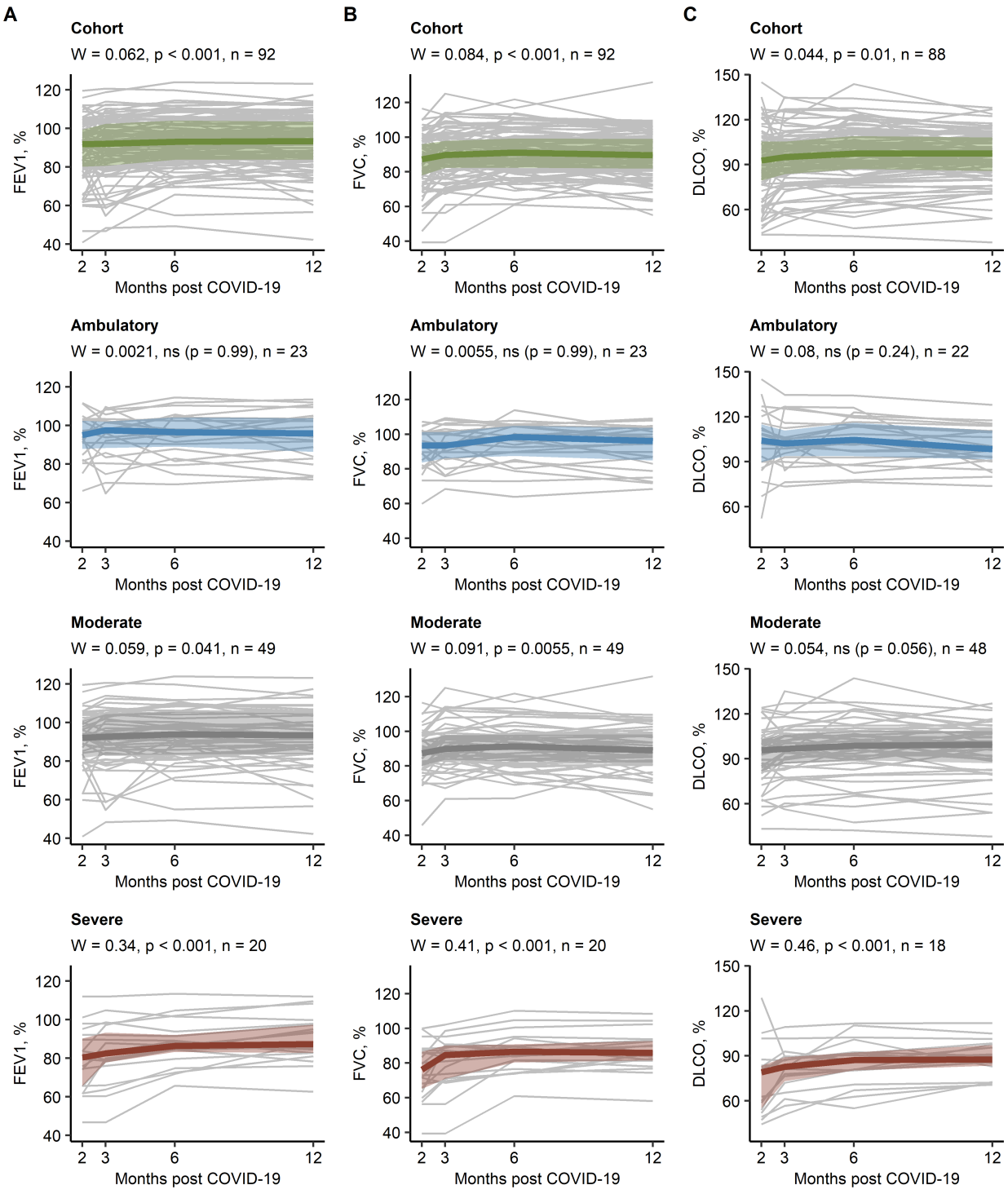
Frequencies of self-reported fatigue (**A**), sleep problems (**B**) and dyspnea (**C**) during acute COVID-19 and at the 2-, 3-, 6-month and 1-year follow-up were investigated in the entire study collective and in ambulatory, moderate and severe COVID-19 survivors. Participants with the complete longitudinal data set were included in the analysis. The recovery was modeled by second-order mixed-effect logistic modeling and likelihood ratio test (full vs null model). P values were corrected for multiple testing with the Benjamini-Hochberg method. Likelihood ratio (λ), p values and numbers of participants with the complete longitudinal data set are presented in the plot captions.



Supplementary Figure S2. Recovery from night sweating, cough and smell disorders.

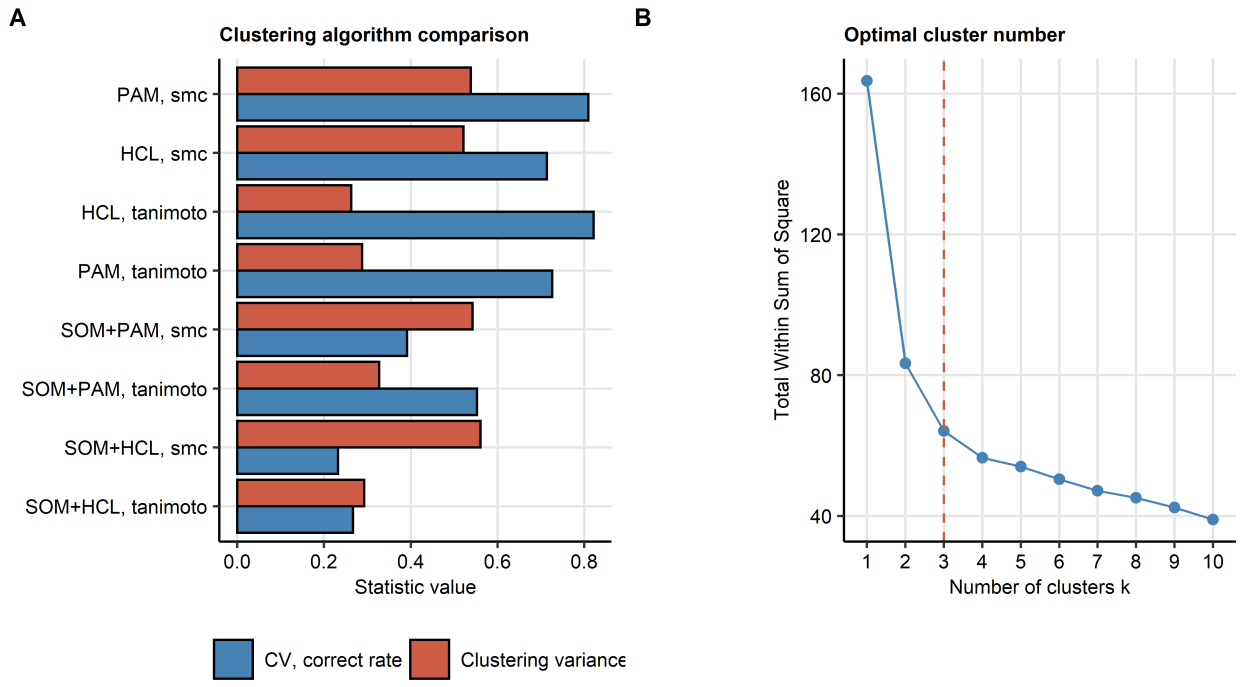
Frequencies of self-reported night sweating (**A**), cough (**B**) and hypo- or anosmia (**C**) during acute COVID-19 and at the 2-, 3-, 6-month and 1-year follow-up were investigated in

the entire study collective and in ambulatory, moderate and severe COVID-19 survivors. Participants with the complete longitudinal data set were included in the analysis. The recovery was modeled by second-order mixed-effect logistic modeling and likelihood ratio test (full vs null model). P values were corrected for multiple testing with the Benjamini-Hochberg method. Likelihood ratio (λ), p values and numbers of participants with the complete longitudinal data set are presented in the plot captions.



Supplementary Figure S3. Changes in FEV1, FVC and DLCO during COVID-19 convalescence.

FEV1 (forced expiratory volume in 1 second, **A**), FVC (forced vital capacity, **B**) and DLCO (diffusion lung capacity for carbon monoxide, **C**) were analyzed at the 2-, 3-, 6-month and 1-year follow-up in the entire study collective and in ambulatory, moderate and severe COVID-19 survivors. Participants with the complete longitudinal data set were included in the analysis. The recovery was assessed with Friedman test with Kendall's W effect size statistic. P values were corrected for multiple testing with the Benjamini-Hochberg method. W, p values and numbers of participants with the complete longitudinal data set are presented in the plot captions. Gray lines indicate parameter values for particular individuals, thick colored lines represented parameter medians, tinted regions indicate interquartile ranges.

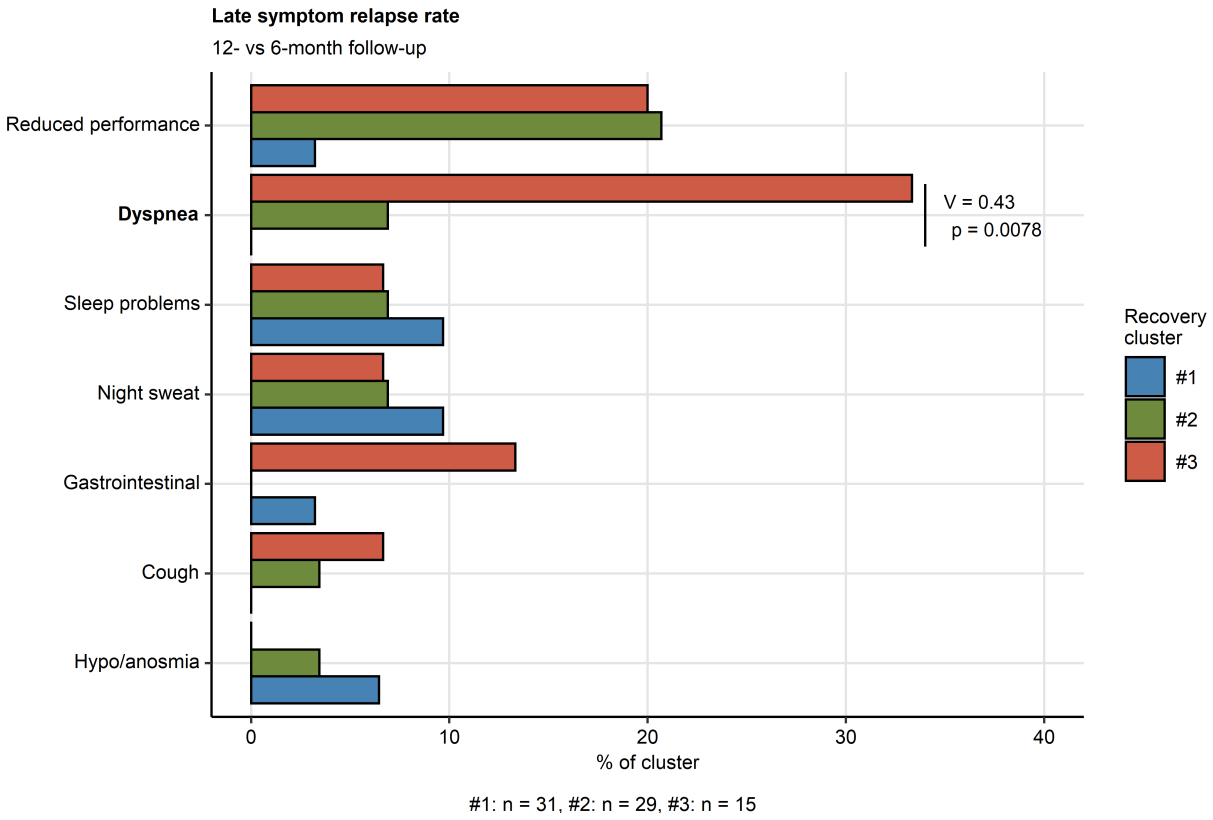


Supplementary Figure S4. Development of COVID-19 recovery clusters.

Clustering of the study participants in respect to 19 binary symptom, cardiopulmonary and psychosocial features recorded at the 1-year follow-up (**Supplementary Tables S1** and **S2**) was investigated with the PAM (partitioning around medoids) algorithm with simple matching distance (SMD) statistic. Study participants with the complete clustering data set were included in the analysis (n = 87).

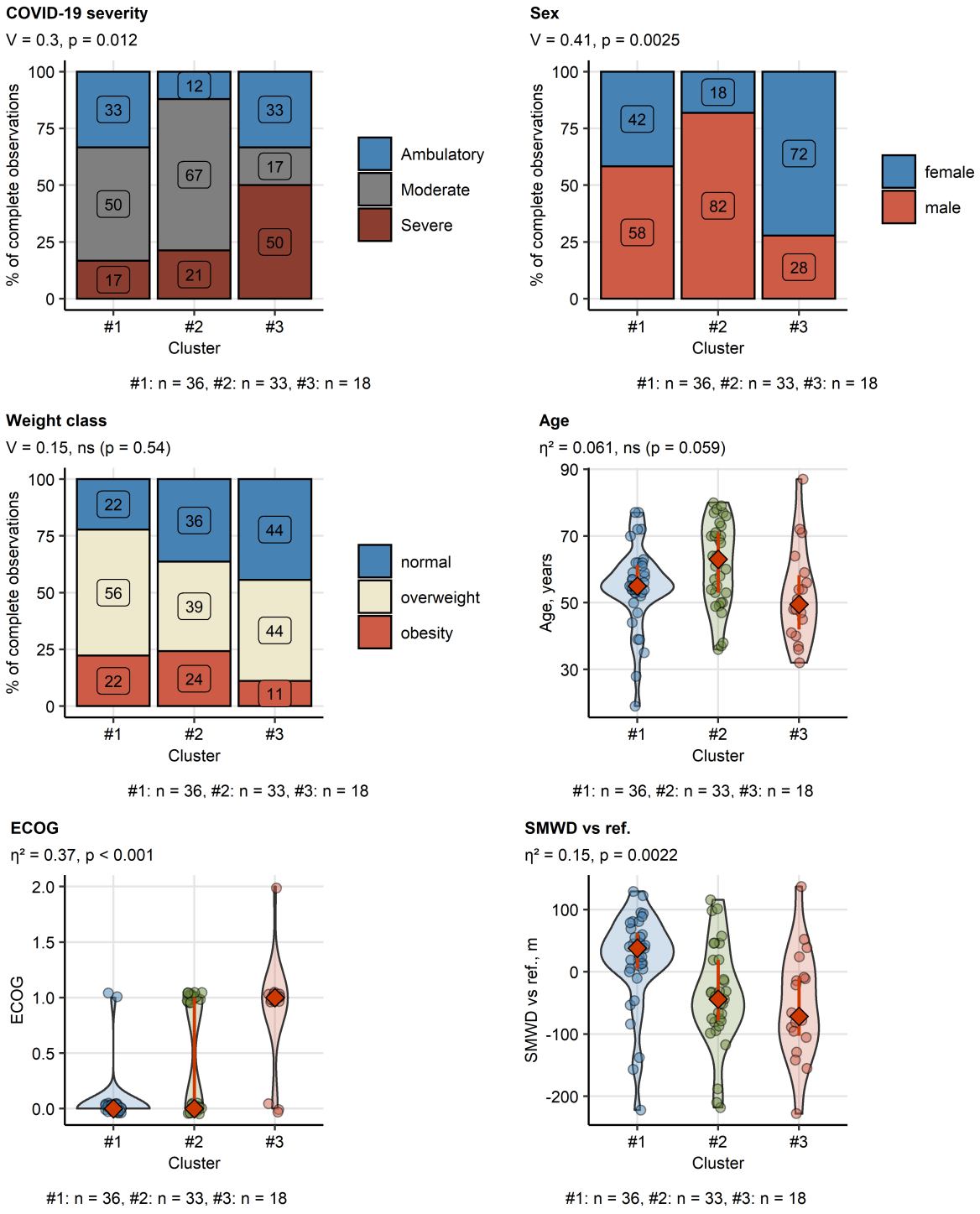
(A) Performance comparison of various clustering algorithms and distance measures. Clustering algorithm stability was assessed by the rate of correct cluster assignment in 10-fold cross-validation (CV). Clustering variance is expressed as a ratio of the total between-cluster sum of squares to total sum of squares. Based on the optimal stability, PAM (partitioning around medoids) algorithm with the Manhattan distance statistic was chosen for further analyses.

(B) The total within-cluster sum sum of squares as a function of the cluster number for the PAM/SMD clustering procedure. The optimal cluster number (dashed red line) was set at the bend of the curve.



Supplementary Figure S5. Rates of symptom relapse at the 12-month follow-up in the COVID-19 recovery clusters.

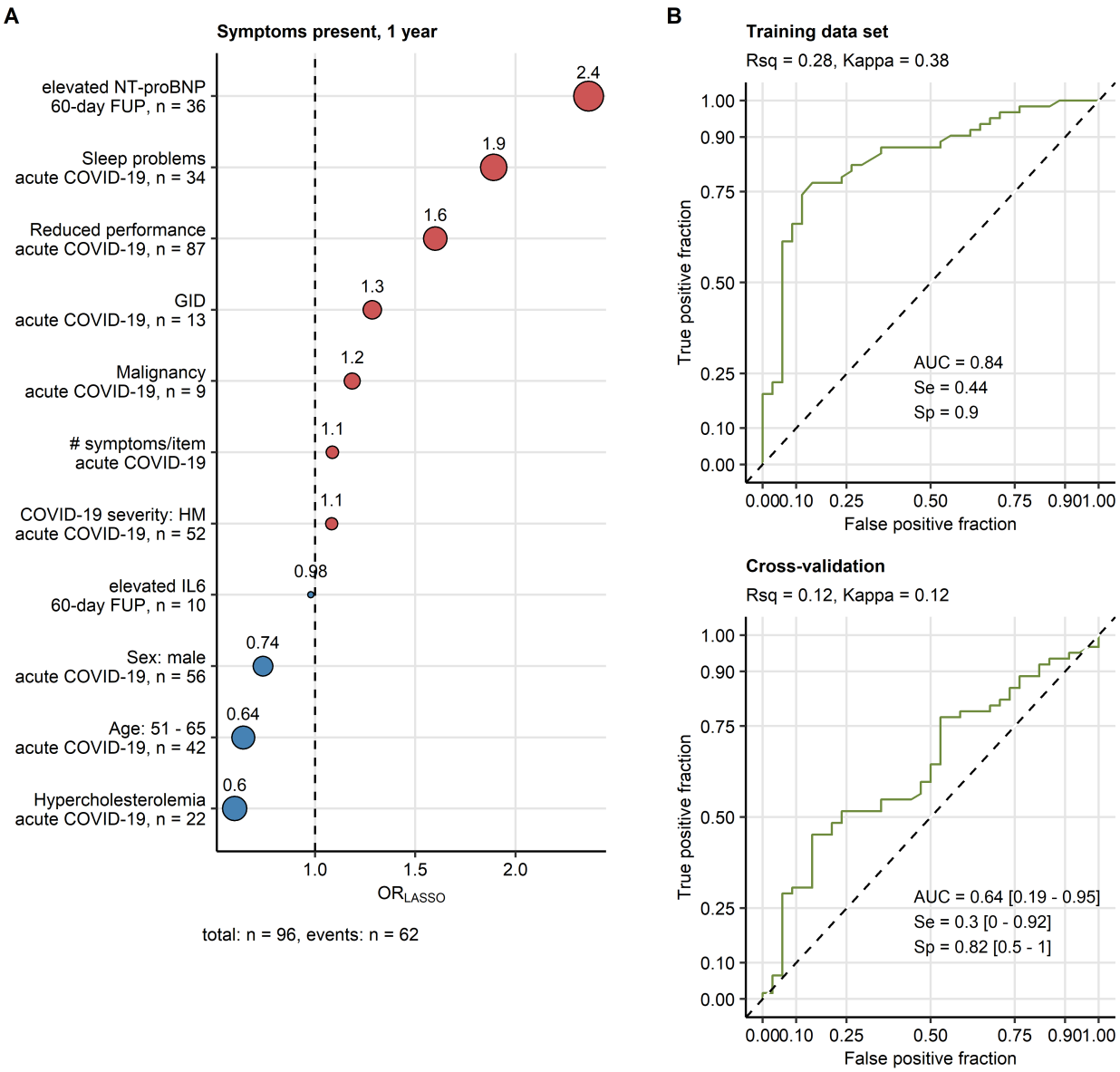
Rates of relapse for particular COVID-19-related symptoms at the 12-month follow-up as compared with the 6-month follow-up were compared between the COVID-19 recovery clusters (**Figure 7**, **Supplementary Figure S4**) with χ^2 test with Cramer V effect size statistic. Individuals with the complete symptom record for the investigated follow-up were included in the analysis. P values were corrected for multiple testing with Benjamini-Hochberg method. Percentages of the relapsing individuals are presented as a bar plot. Significant differences are highlighted in bold. Effect size statistics and p values for the significant comparisons are indicated in the plot. Numbers of participants assigned to the clusters are presented under the plots.



Supplementary Figure S6. COVID-19 severity, demographic features, physical performance and mobility in the COVID-19 recovery clusters..

Acute COVID-19 severity, sex and weight class distribution, age at COVID-19 diagnosis as well as physical performance (ECOG: Eastern Cooperative Oncology Group performance status) and six-minute walking distance (SMWD, difference between the observed and reference value at the 1-year follow-up were compared between the COVID-19 recovery clusters (**Figure 7**, **Supplementary Figure S4**). Statistical significance for numeric values was assessed by Kruskal-Wallis test with η^2 effect size statistic or by χ^2 test with categorical variables with Cramer V effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. Effect size statistic and p values are presented in plot captions. Numbers of participants assigned to the clusters are presented under the plots.

SMWD vs ref.: six-minute walking distance, observed versus reference value.



Supplementary Figure S7. Modeling of the persistent symptom risk at the 1-year post-COVID-19 follow-up.

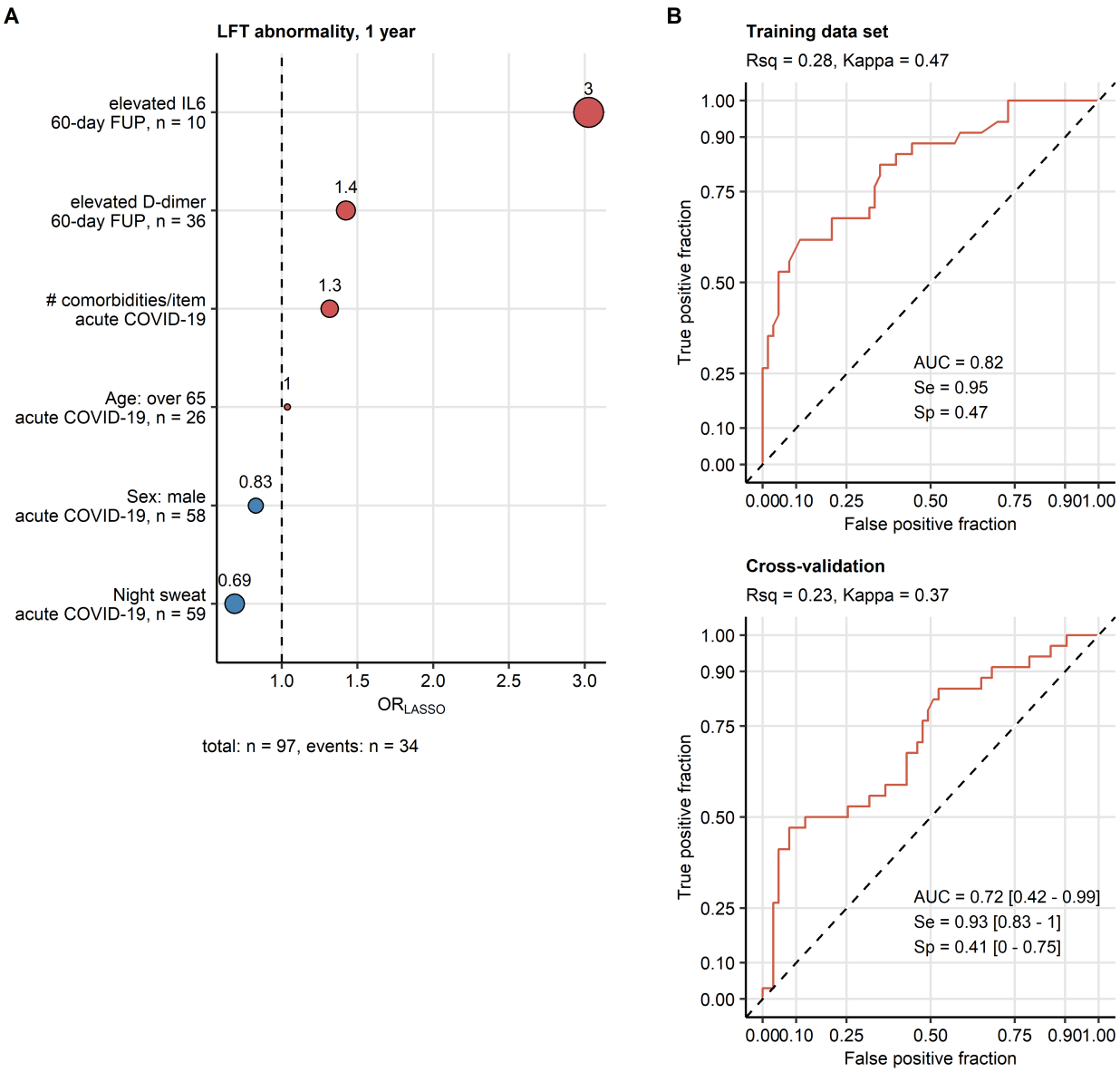
The risk of persistent COVID-19 symptoms at the 1-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression and 34 independent variables recorded during acute COVID-19 and at the 60-

day follow-up (**Supplementary Table S1**). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with symptoms are presented in **A**.

(A) Non-zero model coefficient values presented as odds ratios (OR). Point size codes for the absolute OR value, point color codes for the correlation with the risk (blue: favorable, red: unfavorable factor).

(B) Performance of the LASSO model at predicting the presence of symptoms at the 1-year follow-up in the training data set and 10-fold cross-validation (CV) assessed by receiver-operating characteristic (ROC). Area under the ROC curve (AUC), sensitivity (Se) and specificity (Sp) values with 95% confidence intervals (for CV) are presented in the plots. R-squared (Rsq) and Cohen's kappa statistic values are shown in the plot captions.

NT-proBNP: N-terminal pro - brain natriuretic peptide; GID: gastrointestinal disease; #: number of; HM: hospitalized moderate COVID-19; IL6: interleukin 6.



Supplementary Figure S8. Modeling of the persistent functional lung abnormality at the 1-year post-COVID-19 follow-up.

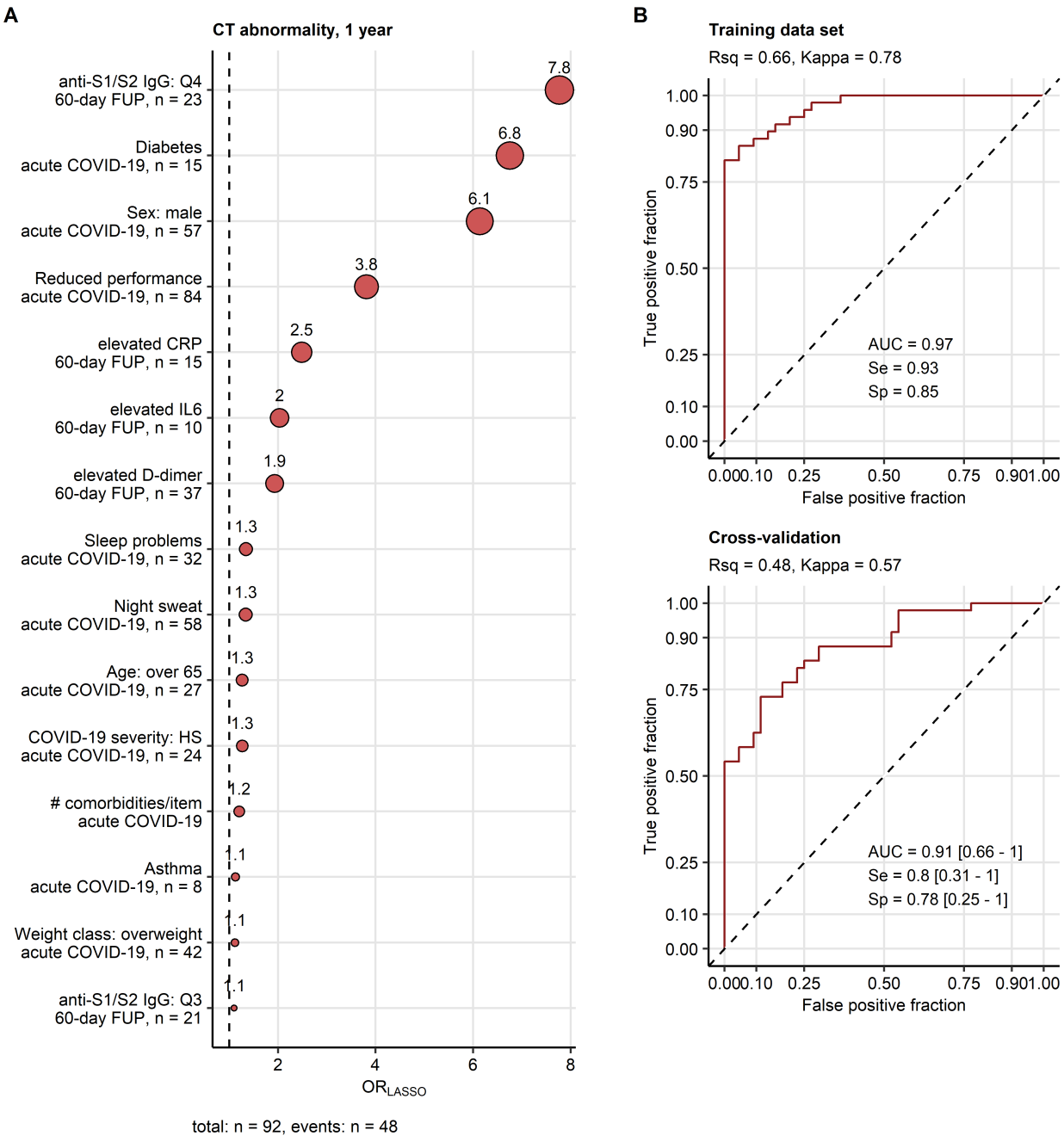
The risk of any lung function testing (LFT) abnormality at the 1-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression and 34 independent variables recorded during acute COVID-19 and at

the 60-day follow-up (**Supplementary Table S1**). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with LFT abnormalities are presented in **A**.

(A) Non-zero model coefficient values presented as odds ratios (OR). Point size codes for the absolute OR value, point color codes for the correlation with the risk (blue: favorable, red: unfavorable factor).

(B) Performance of the LASSO model at predicting the lung function abnormalities at the 1-year follow-up in the training data set and 10-fold cross-validation (CV) assessed by receiver-operating characteristic (ROC). Area under the ROC curve (AUC), sensitivity (Se) and specificity (Sp) values with 95% confidence intervals (for CV) are presented in the plots. R-squared (Rsq) and Cohen's kappa statistic values are shown in the plot captions.

IL6: interleukin 6; #: number of.



Supplementary Figure S9. Modeling of the persistent radiological lung abnormality at the 1-year post-COVID-19 follow-up.

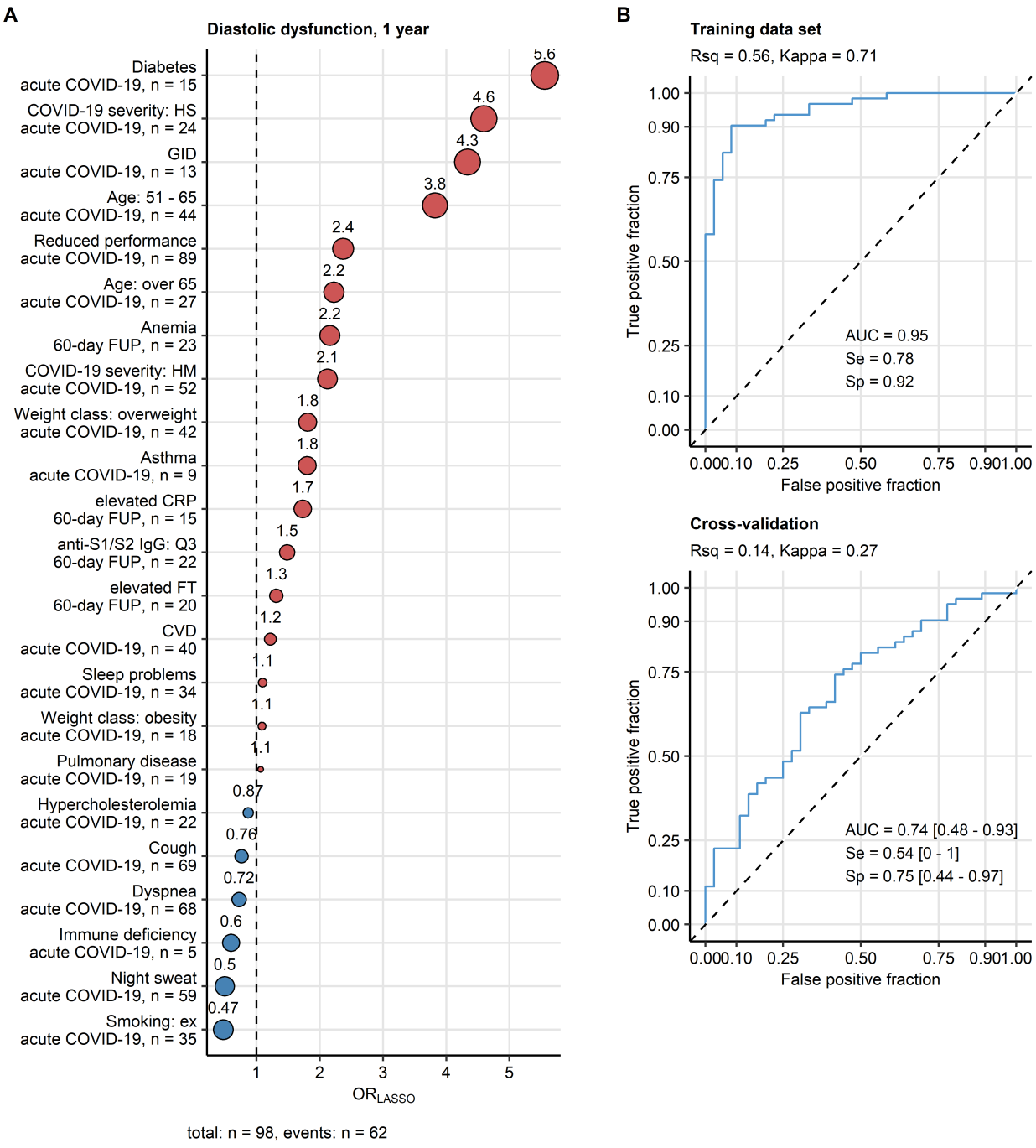
The risk of any chest computed tomography (CT) abnormality at the 1-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression and 34 independent variables recorded during acute COVID-19 and at

the 60-day follow-up (**Supplementary Table S1**). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with CT abnormalities are presented in **A**.

(A) Non-zero model coefficient values presented as odds ratios (OR). Point size codes for the absolute OR value, point color codes for the correlation with the risk (blue: favorable, red: unfavorable factor).

(B) Performance of the LASSO model at predicting the CT abnormalities at the 1-year follow-up in the training data set and 10-fold cross-validation (CV) assessed by receiver-operating characteristic (ROC). Area under the ROC curve (AUC), sensitivity (Se) and specificity (Sp) values with 95% confidence intervals (for CV) are presented in the plots. R-squared (Rsq) and Cohen's kappa statistic values are shown in the plot captions.

anti-S1/S2 IgG: anti-S1/S2 SARS-CoV-2 immunoglobulin; Q3, Q4: 3rd and 4th quartile; CRP: C-reactive protein; IL6: interleukin 6; HS: hospitalized severe COVID-19; #: number of.



Supplementary Figure S10. Modeling of the persistent diastolic dysfunction at the 1-year post-COVID-19 follow-up.

The risk of diastolic dysfunction at the 1-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression and 34 independent variables recorded during acute COVID-19 and at the 60-day follow-up

(**Supplementary Table S1**). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with diastolic dysfunction are presented in **A**.

(A) Non-zero model coefficient values presented as odds ratios (OR). Point size codes for the absolute OR value, point color codes for the correlation with the risk (blue: favorable, red: unfavorable factor).

(B) Performance of the LASSO model at predicting diastolic dysfunction at the 1-year follow-up in the training data set and 10-fold cross-validation (CV) assessed by receiver-operating characteristic (ROC). Area under the ROC curve (AUC), sensitivity (Se) and specificity (Sp) values with 95% confidence intervals (for CV) are presented in the plots. R-squared (Rsq) and Cohen's kappa statistic values are shown in the plot captions.

GID: gastrointestinal disease; HS: hospitalized severe COVID-19; CRP: C-reactive protein; HM: hospitalized moderate COVID-19; FT: ferritin; CVD: cardiovascular disease; anti-S1/S2 IgG: anti-S1/S2 SARS-CoV-2 immunoglobulin; Q4: 4^{th} quartile.

References

1. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. Welcome to the Tidyverse. *Journal of Open Source Software* (2019) 4:1686. doi: 10.21105/joss.01686

2. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. New York: Springer-Verlag (2016). pp. https://ggplot2.tidyverse.org

3. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st ed. Sebastopol: O'Reilly Media (2019). pp.

4. Sonnweber T, Tymoszuk P, Sahanic S, Boehm A, Pizzini A, Luger A, Schwabl C, Nairz M, Grubwieser P, Kurz K, et al. Investigating phenotypes of pulmonary COVID-19 recovery - a longitudinal observational prospective multicenter trial. *eLife* (2022) 11: doi: 10.7554/ELIFE.72500

5. Sonnweber T, Sahanic S, Pizzini A, Luger A, Schwabl C, Sonnweber B, Kurz K, Koppelstätter S, Haschka D, Petzer V, et al. Cardiopulmonary recovery after COVID-19 - an observational prospective multi-center trial. *The European respiratory journal* (2020) doi: 10.1183/13993003.03481-2020

6. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* (2011) 20:1727–1736. doi: 10.1007/S11136-011-9903-X/TABLES/5

7. Marten O, Greiner W. EQ-5D-5L reference values for the German general elderly population. *Health and quality of life outcomes* (2021) 19: doi: 10.1186/S12955-021-01719-7

8. Crapo RO, Casaburi R, Coates AL, Enright PL, MacIntyre NR, McKay RT, Johnson D, Wanger JS, Zeballos RJ, Bittner V, et al. ATS Statement. *https://doiorg/101164/ajrccm1661at1102* (2012) 166:111–117. doi: 10.1164/AJRCCM.166.1.AT1102

9. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. *American Journal of Respiratory and Critical Care Medicine* (1998) 158:1384–1387. doi: 10.1164/ajrccm.158.5.9710086

10. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

11. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* (2015) 67:1–48. doi: 10.18637/jss.v067.i01

12. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* (2017) 82:1–26. doi: 10.18637/jss.v082.i13

13. Box GE, Hunter SJ, Hunter WG. Statistics for experimenters: an introduction to design, data analysis, and model building. (2005)

14. Schubert E, Rousseeuw PJ. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer (2019). pp. 171–187 doi: 10.1007/978-3-030-32047-8_16

15. Boriah S, Chandola V, Kumar V. Similarity measures for categorical data: A comparative evaluation. In. *Society for industrial and applied mathematics - 8th siam international conference on data mining 2008, proceedings in applied mathematics 130*. (2008). pp. 243–254 doi: 10.1137/1.9781611972788.22

16. Lange T, Roth V, Braun ML, Buhmann JM. Stability-Based Validation of Clustering Solutions. *Neural Computation* (2004) 16:1299–1323. doi: 10.1162/089976604773717621

17. Leng M, Wang J, Cheng J, Zhou H, Chen X. Adaptive semi-supervised clustering algorithm with label propagation. *Journal of Software Engineering* (2014) 8:14–22. doi: 10.3923/JSE.2014.14.22

18. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* (1996) 58:267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

19. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* (2010) 33:1–22. doi: 10.18637/jss.v033.i01

20. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software* (2008) 28:1–26. doi: 10.18637/jss.v028.i05

21. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* (1969) 72:323–327. doi: 10.1037/h0028106

22. Sachs MC. Plotroc: A tool for plotting ROC curves. *Journal of Statistical Software* (2017) 79:1–19. doi: 10.18637/jss.v079.c02