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

Supplementary Material

CovILD study team

2022-09-24

Supplementary Methods

Software

The study variables were transformed, analyzed and visualized with R version 4.2.0. General data import and transformation tasks were accomplished with the *tidyverse* [1] and *rlang* [2] packages. Analysis results were visualized with the *ggplot2* [3], *cowplot* [4], *rmarkdown* [5], *knitr* [6] and *bookdown* [7] packages and the development package *figur* (https://github.com/PiotrTymoszuk/figur). Tables were generated with the package *flextable* [8].

Procedures, variables and variable stratification

Numeric variables are presented as medians with interquartile ranges (IQR), categorical variables are presented as percentages of complete answers.

Participant age during acute COVID-19 was stratified for modeling tasks with the 50 and 65 year cutoffs. Participants were stratified according to the severity of acute COVID-19 as ambulatory (outpatients), moderate (hospitalized at COVID-19 ward, no mechanical ventilation) and severe (mechanical ventilation and/or intensive care unit stay).

Investigated COVID-19 symptoms encompassed self-reported complaints (fever, night sweating, cough, smell disorders, sleep disorders, hair loss, dermatological and gastrointestinal symptoms, surveyed by single yes/no questions), dyspnoea (Modified Medical British Research Council [mMRC], dyspnea: mMRC ≥ 1), reduced physical performance (Eastern Cooperative Oncology Group [ECOG], reduced performance: ECOG ≥1) [9] and fatigue (bimodal and likert 11-item Chalder fatigue score [CFS]; fatigue: bimodal CFS ≥ 4) [10].

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 according to the Fleischner society glossary terms [11] and rated with the CT severity score summed over all lung lobes [12–14]. The CT score for each lobe was defined as follows: 0: none; 1: minimal (subtle ground-glass opacities [GGO]); 2: mild (several GGOs, subtle reticulation); 3: moderate (multiple GGOs, reticulation, small consolidation); 4: severe (extensive GGOs, consolidation, reticulation with distortion); and 5: massive (massive findings, parenchymal destruction). The total CT severity score was the sum overa all lung lobes. Any chest CT abnormalities were defined

as CT severity score ≥ 1, moderate-to-severe CT abnormalities were defined as CT severity score > 5.

TTE was performed according to the ESC and EACVI recommendations [15]. Diastolic dysfunction was determined by assessing left ventricular filling by trans-mitral PW Doppler and measuring the peak of early filling (E) and peak of late atrial filling (A). For the assessment of pseudo-normal E/A ratio, peak early diastolic velocity (e') was measured at the lateral mitral annulus with tissue Doppler to calculate E/e'. Grading of diastolic dysfunction was performed as follows: normal diastolic function: E/A ratio = 1 - 2; diastolic dysfunction I°: E/A ratio < 0.8; diastolic dysfunction II°: E/A = 0.8 - 1.5. and diastolic dysfunction III°: E/A > 2.

Reference values of the six-minute walking distance (6MWD) [16] were calculated with the participants sex, age, weight and height as described previously [17]. Low 6MWD was defined as a value below the patient's specific reference value.

Sub-scores of the EQ-5D-5L questionnaire (European quality of life 5 dimensions, 5 levels) [18] 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 [19]. Elevated stress levels were defined by the median split of the 4-item PSS score (Perceived Stress Scale, > 5: elevated stress) [20]. Resilient coping classes: low (4 - 13 points), medium (14 - 16) and high (17 - 21 points) were defined with the Brief Resilient Coping Scale (BRCS) [21].

Laboratory parameters were determined by the ISO-certified central laboratory of the University Hospital of Innsbruck (Zentralinstitut für medizinische u. chemische Labordiagnostik, ZIMCL). C-reative protein (CRP), interleukin 6 (IL6), procalcitonin (PCT), N-terminal pro-brain natriuretic peptide (NT-proBNP) and serum ferritin (FT) were determined with a Roche Cobas 8000 analyser (Basel, Switzerland). D-dimer was measured with a Siemens BCS-XP instrument using the Siemens D-Dimer Innovance reagent (Erlangen, Germany). Anti-S1/S2 protein SARS-CoV-2 immunoglobulin gamma (IgG) was quantified with LIAISON chemoluminescence assay (DiaSorin, Italy) and expressed as arbitrary units. Anemia was defined as hemoglobin < 140 g/dL (male) or < 120 g/dL (female). Elevated 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 NT-proBNP was defined as > 150 pg/mL. Elevated D-dimer, CRP, PCT and 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) [12].

The full list of the study variables and their stratification scheme is presented in **Supplementary Table1**.

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 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. Associations of the readouts of clinical, cardiopulmonary, mobility and quality of life deficits at the one-year follow-up visit were determined by pairwise Kendall's τ B correlation. In the correlation analysis, binary variables were recoded as 1 and 2 for absent and present, respectively. P values were corrected for multiple comparisons with Benjamini-Hochberg method [22]. R packages *rstatix* [23], *rcompanion* [24] and *Hmisc* [25] 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 parameter value in time

To model recovery kinetics for categorical variables, second-order mixed-effect logistic (categorical features) modeling was applied (packages: *lme4* and *lmerTest*) [26–28]. 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. The likelihood ratio λ statistic was used as an effect size measure. Results of the kinetic modeling were adjusted for multiple comparisons with Benjamini-Hochberg method [22].

Clustering analysis

To identify patterns of COVID-19 recovery defined by 19 binary symptom, cardiopulmonary and psychosocial variables recorded at the one-year follow-up visit (**Supplementary Tables S1**) 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 (packages *cluster* and *nomclust*) [29–31]. 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, packages *dbscan* and the development package *clustTools* [https://github.com/PiotrTymoszuk/clustTools]) [32, 33] 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 (package *factoextra*) [34] as presented in **Supplementary Figure S4B**. Permutation importance of the clustering factors was assessed by comparing the clustering variance (ratio of the total between-cluster to total sum of squares) of the original clustering structure with the clustering variance of the clustering object with the given variable reshuffled randomly. The difference in clustering variances (Δ clustering variance) served as the importance statistic. Determination of the clustering factor importance was done for n = 100 permutations for each variable using the development package *clustTools* (https://github.com/PiotrTymoszuk/clustTools) as presented in **Supplementary Figure 4C**.

Multi-parameter modeling

The risk of persistent COVID-19 symptoms, LFT findings, CT abnormalities, and diastolic dysfunction at the one-year follow-up was modeled with multi-parameter logistic LASSO (least absolute shrinkage and selection operator) regression [35] and 34 independent variables recorded during acute COVID-19 and at the 60-day follow-up (**Supplementary Table S1**). The LASSO modeling was accomplished with the packages *glmnet* [36] and *caret* [37]. 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 [36]. The model performance at predicting the one-year event in the training data set and in 10-fold CV was investigated by receiver-operating characteristic (ROC) [37], Cohen's κ [38] and R-squared statistics (package *caret*) [37]. Model assumptions were checked by a visual control of model residuals plots (residuals versus fitted and quantile-quantile plots, development package *caretExtra* [https://github.com/PiotrTymoszuk/caretExtra]). ROC plots were generated with the *plotroc* package [39] and the 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. The table is available as a supplementary Excel file.

Supplementary Table S2: Extended baseline characteristics, post-acute steroid therapy and rehabilitation status of the COVID-19 recovery clusters.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
N number	36	33	18		
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 = 0.33)	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 = 0.19)	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 = 0.061)	V = 0.29
Cardiovascular disease	36% (n = 13)	61% (n = 20)	22% (n = 4)	p = 0.044	V = 0.3
Pulmonary disease	11% (n = 4)	24% (n = 8)	17% (n = 3)	ns (p = 0.52)	V = 0.15
Malignancy	8.3% (n = 3)	15% (n = 5)	0% (n = 0)	ns (p = 0.35)	V = 0.19
Immune deficiency	2.8% (n = 1)	3% (n = 1)	11% (n = 2)	ns (p = 0.5)	V = 0.16
Chronic kidney disease[b]	5.6% (n = 2)	12% (n = 4)	5.6% (n = 1)	ns (p = 0.67)	V = 0.12
Gastrointestinal disease	11% (n = 4)	15% (n = 5)	11% (n = 2)	ns (p = 0.9)	V = 0.059
Steroid therapy	8.3% (n = 3)	30% (n = 10)	22% (n = 4)	ns (p = 0.14)	V = 0.25
Rehabilitation	none: 86% (n = 31) inpatient: 11% (n = 4) outpatient: 2.8% (n = 1)	none: 58% (n = 19) inpatient: 30% (n = 10) outpatient: 12% (n = 4)	none: 44% (n = 8) inpatient: 50% (n = 9) outpatient: 5.6% (n = 1)	p = 0.027	V = 0.28

[a] χ^2 test with Cramer V effect size statistic. P values corrected for multiple testing with Benjamini-Hochberg method.

[b] Steroid therapy in cases of non-resolving pneumonia beginning from week four post diagnosis at the discretion of the physician.

Supplementary Table S3: Extended characteristic of the COVID-19 recovery clusters at the one-year follow-up.

Variable	Cluster #1	Cluster #2	Cluster #3	Significance[a]	Effect size[a]
N number	36	33	18		
Number of 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	$p < 0.001$	$\eta^2 = 0.49$
Hair loss	0% (n = 0)	3% (n = 1)	33% (n = 6)	$p < 0.001$	V = 0.48
Dermatological symptoms	2.8% (n = 1)	21% (n = 7)	28% (n = 5)	ns ($p = 0.055$)	V = 0.29
CT severity score[b]	median: 0 [IQR: 0 - 0.25] range: 0 - 13	median: 2 [IQR: 1 - 5] range: 0 - 15	median: 0 [IQR: 0 - 8.8] range: 0 - 13	$p < 0.001$	$\eta^2 = 0.18$
6MWD, m[c]	median: 580 [IQR: 530 - 640] range: 460 - 740	median: 520 [IQR: 450 - 580] range: 270 - 760	median: 530 [IQR: 420 - 620] range: 310 - 710	$p = 0.027$	$\eta^2 = 0.085$
Mobility impairment score (EQ5D5L)[d]	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 1] range: 1 - 2	median: 1 [IQR: 1 - 2] range: 1 - 3	$p = 0.021$	$\eta^2 = 0.093$
Self-care impairment score (EQ5D5L)[d]	median: 1 [IQR: 1 - 1] range: 1 - 1	median: 1 [IQR: 1 - 1] range: 1 - 1	median: 1 [IQR: 1 - 1] range: 1 - 2	$p = 0.0091$	$\eta^2 = 0.12$

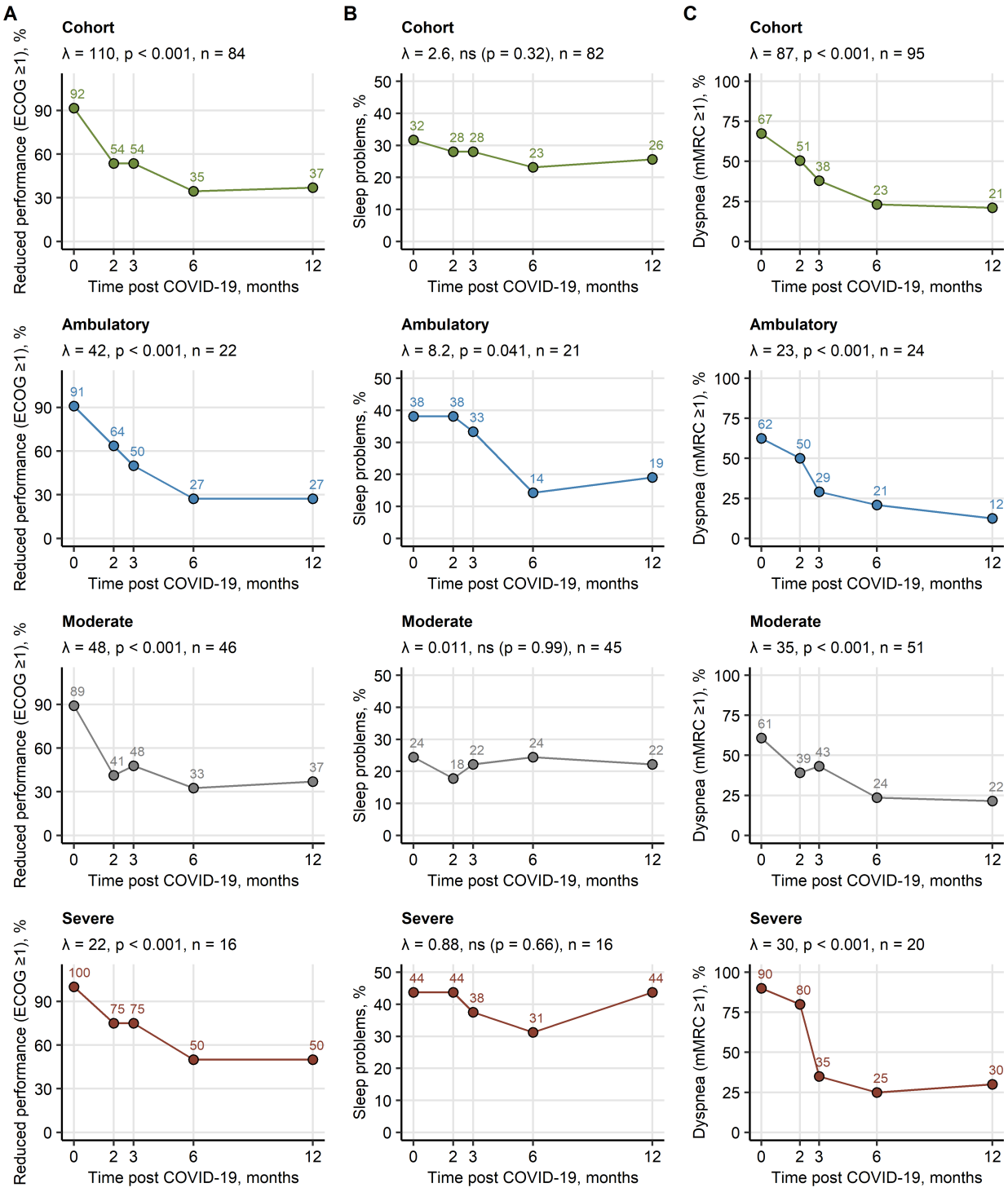
[a]Categorical variables: χ^2 test with Cramer V effect size statistic; numeric variables: Kruskal-Wallis test with η^2 effect size statistic. P values corrected for multiple testing with Benjamini-Hochberg method.

[b]CT severity score: chest computed tomography severity score

[c]6MWD, m: six-minute walking distance, meters

[d]EQ5D5L: European quality of life 5 dimensions, 5 levels

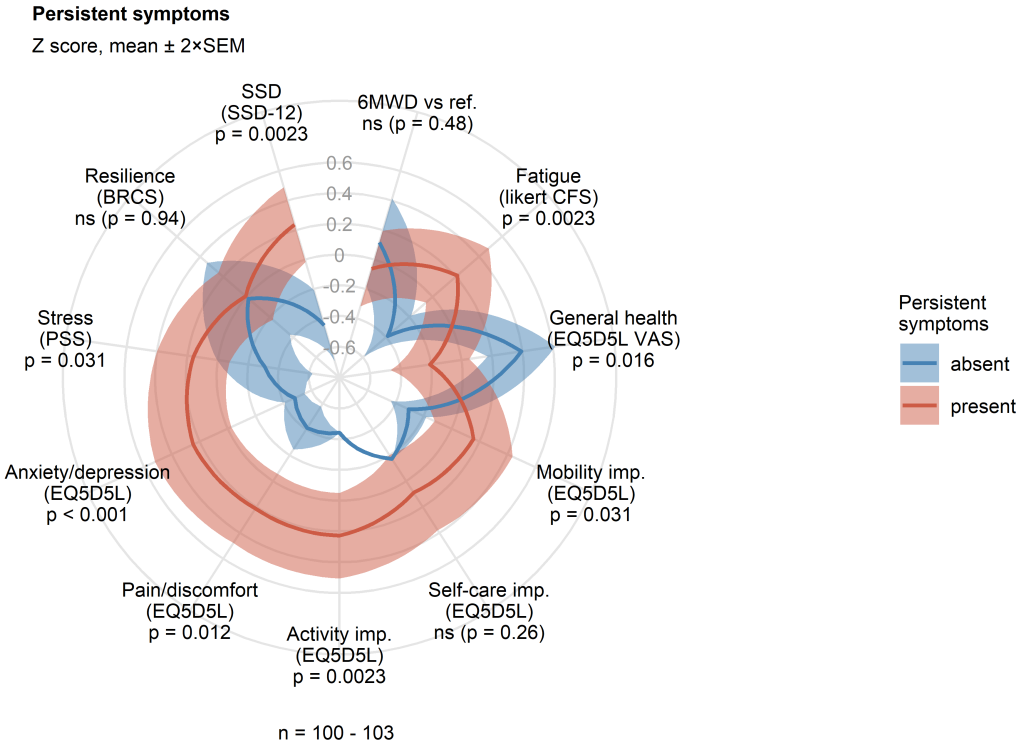
Supplementary Figures



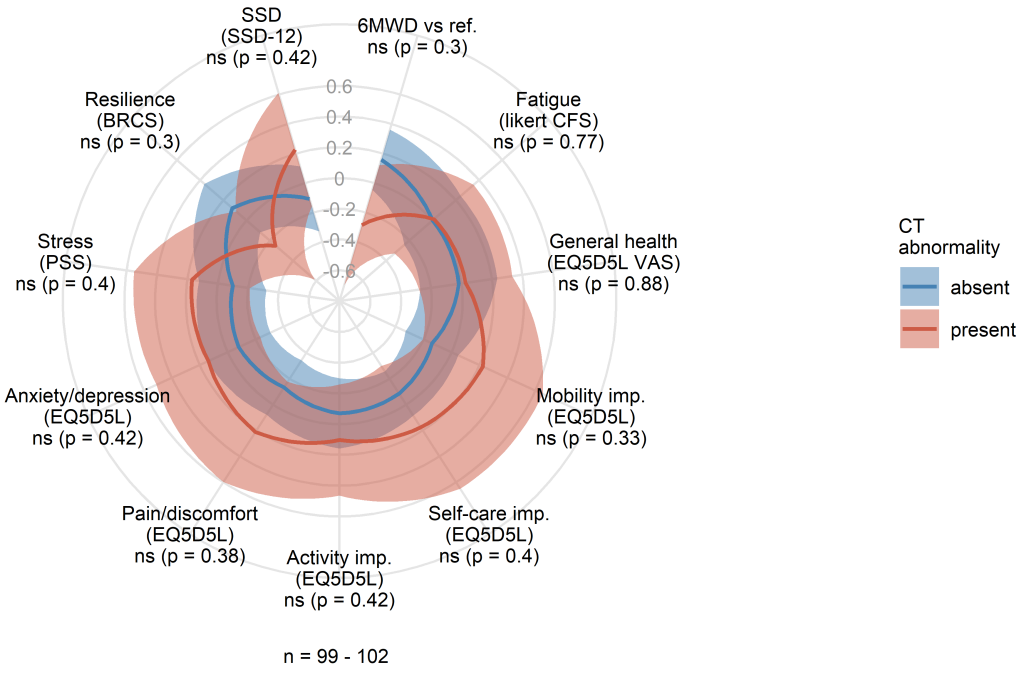
Supplementary Figure S1. Recovery of fatigue, sleep problems and dyspnea.

Frequencies of reduced physical performance (Eastern Cooperative Oncology score [ECOG] ≥ 1) (A), self-reported sleep problems (B) and dyspnea (modified Medical Research Council score [mMRC] ≥ 1) (C) during acute COVID-19 and at the 2-, 3-, 6-month and one-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 symptom kinetic was analyzed 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.

A


Persistent symptoms
Z score, mean ± 2×SEM
SSD (SSD-12) p = 0.0023
6MWD vs ref. ns (p = 0.48)
Fatigue (likert CFS) p = 0.0023
General health (EQ5D5L VAS) p = 0.016
Mobility imp. (EQ5D5L) p = 0.031
Self-care imp. (EQ5D5L) ns (p = 0.26)
Activity imp. (EQ5D5L) p = 0.0023
Pain/discomfort (EQ5D5L) p = 0.012
Anxiety/depression (EQ5D5L) p < 0.001
Stress (PSS) p = 0.031
Resilience (BRCS) ns (p = 0.94)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Persistent symptoms
absent
present
n = 100 - 103

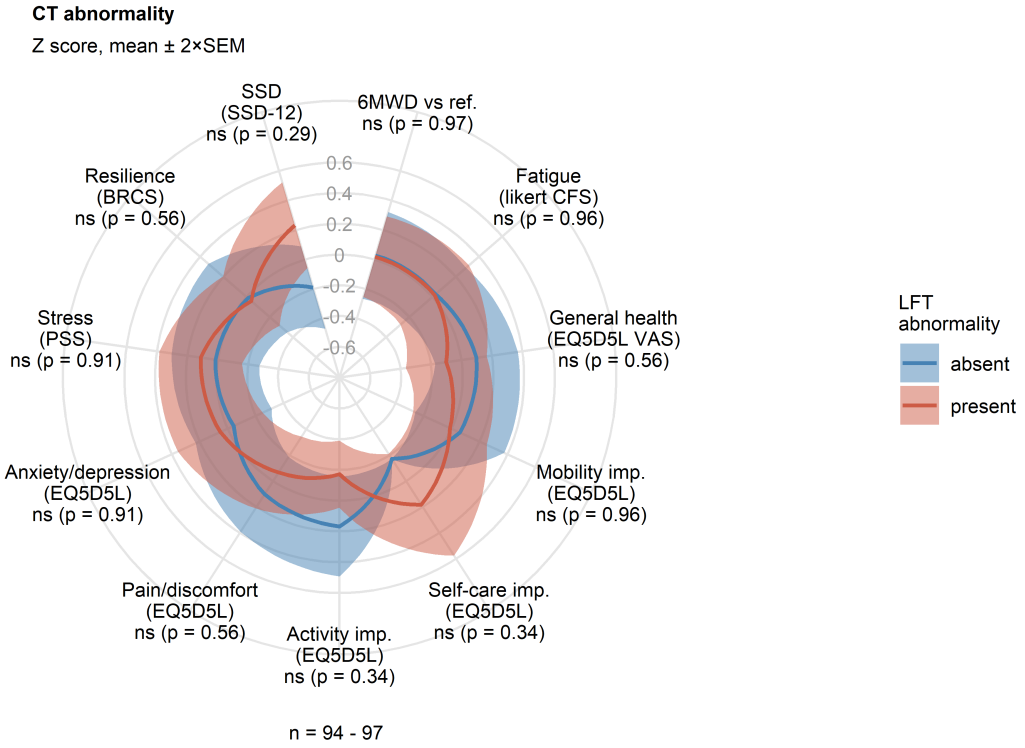
B


LFT abnormality
Z score, mean ± 2×SEM
SSD (SSD-12) ns (p = 0.42)
6MWD vs ref. ns (p = 0.3)
Fatigue (likert CFS) ns (p = 0.77)
General health (EQ5D5L VAS) ns (p = 0.88)
Mobility imp. (EQ5D5L) ns (p = 0.33)
Self-care imp. (EQ5D5L) ns (p = 0.4)
Activity imp. (EQ5D5L) ns (p = 0.42)
Pain/discomfort (EQ5D5L) ns (p = 0.38)
Anxiety/depression (EQ5D5L) ns (p = 0.42)
Stress (PSS) ns (p = 0.4)
Resilience (BRCS) ns (p = 0.3)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
CT abnormality
absent
present
n = 99 - 102

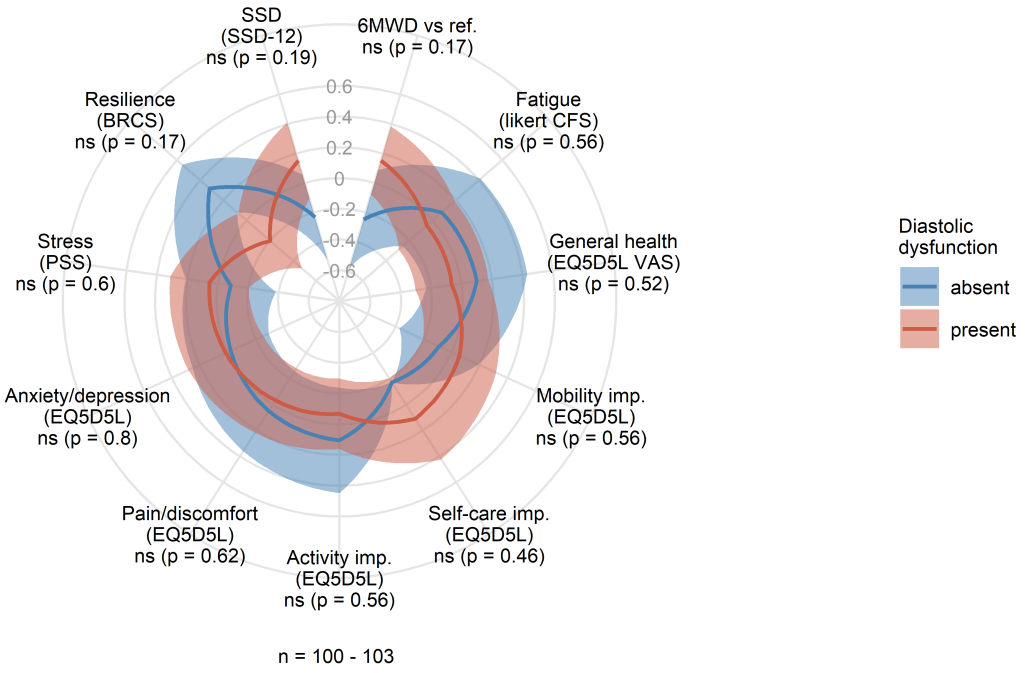
Supplementary Figure S2. Impact of persistent symptoms and LFT abnormality on physical and mental recovery from COVID-19.

Six-minute walking distance (6MWD vs ref.: difference versus the reference value), fatigue rating (CFS: Chalder fatigue score), self-perceived general health (EQ5D5L VAS: European quality of life 5 dimensions, 5 levels, visual analogue scale), self-reported mobility, self-care and activity impairment scores (EQ5D5L: European quality of life 5 dimensions, 5 levels, imp.: impairment), pain/discomfort and depression/anxiety ratings (EQ5D5L), stress (PSS: 4-item perceived stress scale), resilience (BRCS: brief resilient coping scale), and somatic symptom disorder scoring (SSD-12: somatic symptom disorder - B criteria scale) were compared between participants with or without persistent symptoms (A) and with or without lung function testing (LFT) abnormalities (B) at the one-year follow-up. Statistical significance was assessed by Mann-Whitney U test corrected for multiple testing with Benjamini-Hochberg method. Normalized (Z score) mean variable values with 2×SEM are presented in radial plots. Numbers of complete observations are indicated under the plots.

A

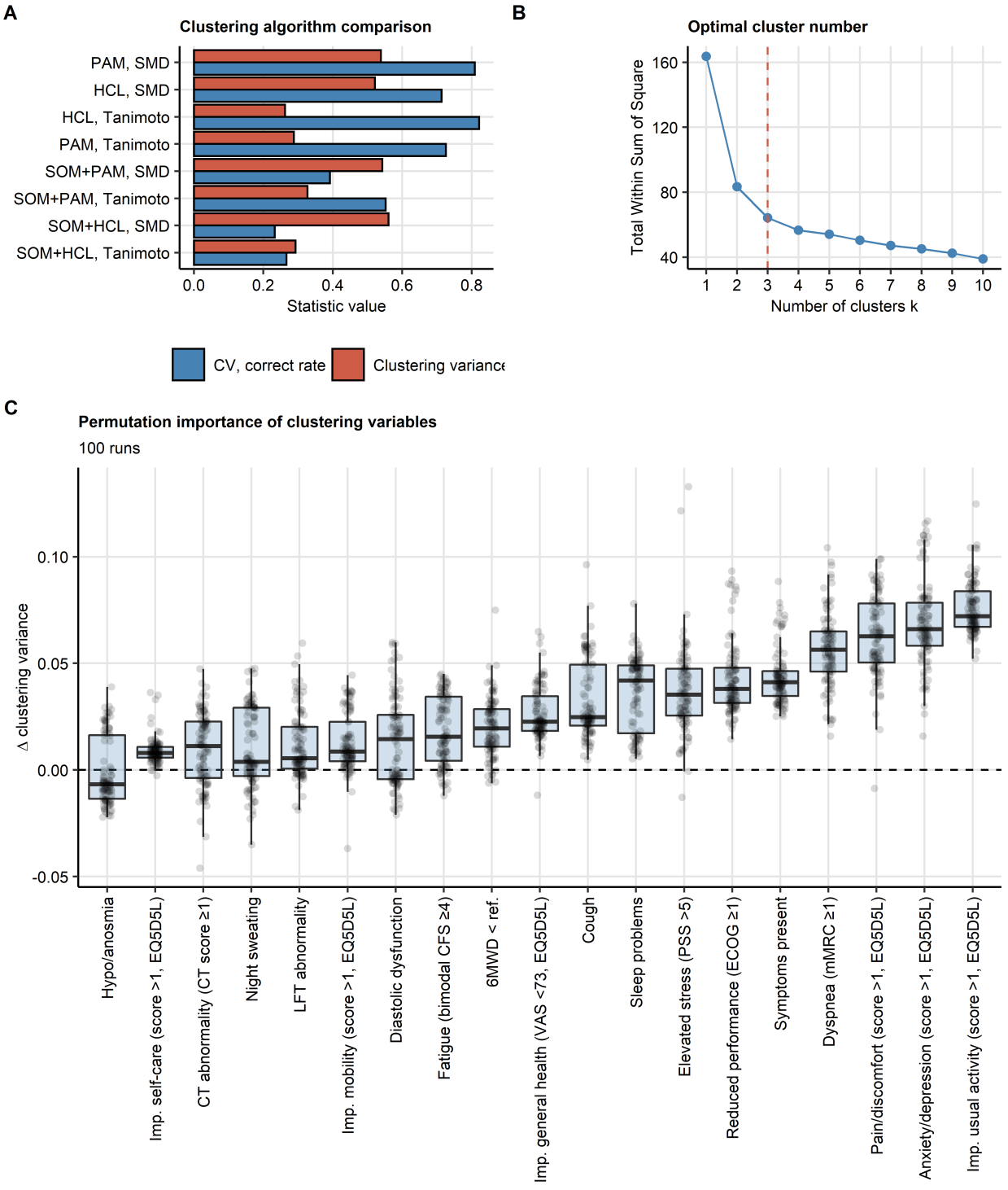

CT abnormality
Z score, mean ± 2×SEM
SSD (SSD-12) ns (p = 0.29)
6MWD vs ref. ns (p = 0.97)
Fatigue (likert CFS) ns (p = 0.96)
General health (EQ5D5L VAS) ns (p = 0.56)
Mobility imp. (EQ5D5L) ns (p = 0.96)
Self-care imp. (EQ5D5L) ns (p = 0.34)
Activity imp. (EQ5D5L) ns (p = 0.34)
Pain/discomfort (EQ5D5L) ns (p = 0.56)
Anxiety/depression (EQ5D5L) ns (p = 0.91)
Stress (PSS) ns (p = 0.91)
Resilience (BRCS) ns (p = 0.56)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
LFT abnormality
absent
present
n = 94 - 97

B


Diastolic dysfunction
Z score, mean ± 2×SEM
SSD (SSD-12) ns (p = 0.19)
6MWD vs ref. ns (p = 0.17)
Fatigue (likert CFS) ns (p = 0.56)
General health (EQ5D5L VAS) ns (p = 0.52)
Mobility imp. (EQ5D5L) ns (p = 0.56)
Self-care imp. (EQ5D5L) ns (p = 0.46)
Activity imp. (EQ5D5L) ns (p = 0.56)
Pain/discomfort (EQ5D5L) ns (p = 0.62)
Anxiety/depression (EQ5D5L) ns (p = 0.8)
Stress (PSS) ns (p = 0.6)
Resilience (BRCS) ns (p = 0.17)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Diastolic dysfunction
absent
present
n = 100 - 103

Supplementary Figure S3. Impact of lung CT and cardiological abnormality on physical and mental recovery from COVID-19.

Six-minute walking distance (6MWD vs ref.: difference versus the reference value), fatigue rating (CFS: Chalder fatigue score), self-perceived general health (EQ5D5L VAS: European quality of life 5 dimensions, 5 levels, visual analogue scale), self-reported mobility, self-care and activity impairment scores (EQ5D5L: European quality of life 5 dimensions, 5 levels, imp.: impairment), pain/discomfort and depression/anxiety ratings (EQ5D5L), stress (PSS: 4-item perceived stress scale), resilience (BRCS: brief resilient coping scale), and somatic symptom disorder scoring (SSD-12: somatic symptom disorder - B criteria scale) were compared between participants with or without lung CT abnormality (A) and with or without diastolic dysfunction (B) at the one-year follow-up. Statistical significance was assessed by Mann-Whitney U test corrected for multiple testing with Benjamini-Hochberg method. Normalized (Z score) mean variance values with 2×SEM are presented in radial plots. Numbers of complete observations are indicated under the plots.



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

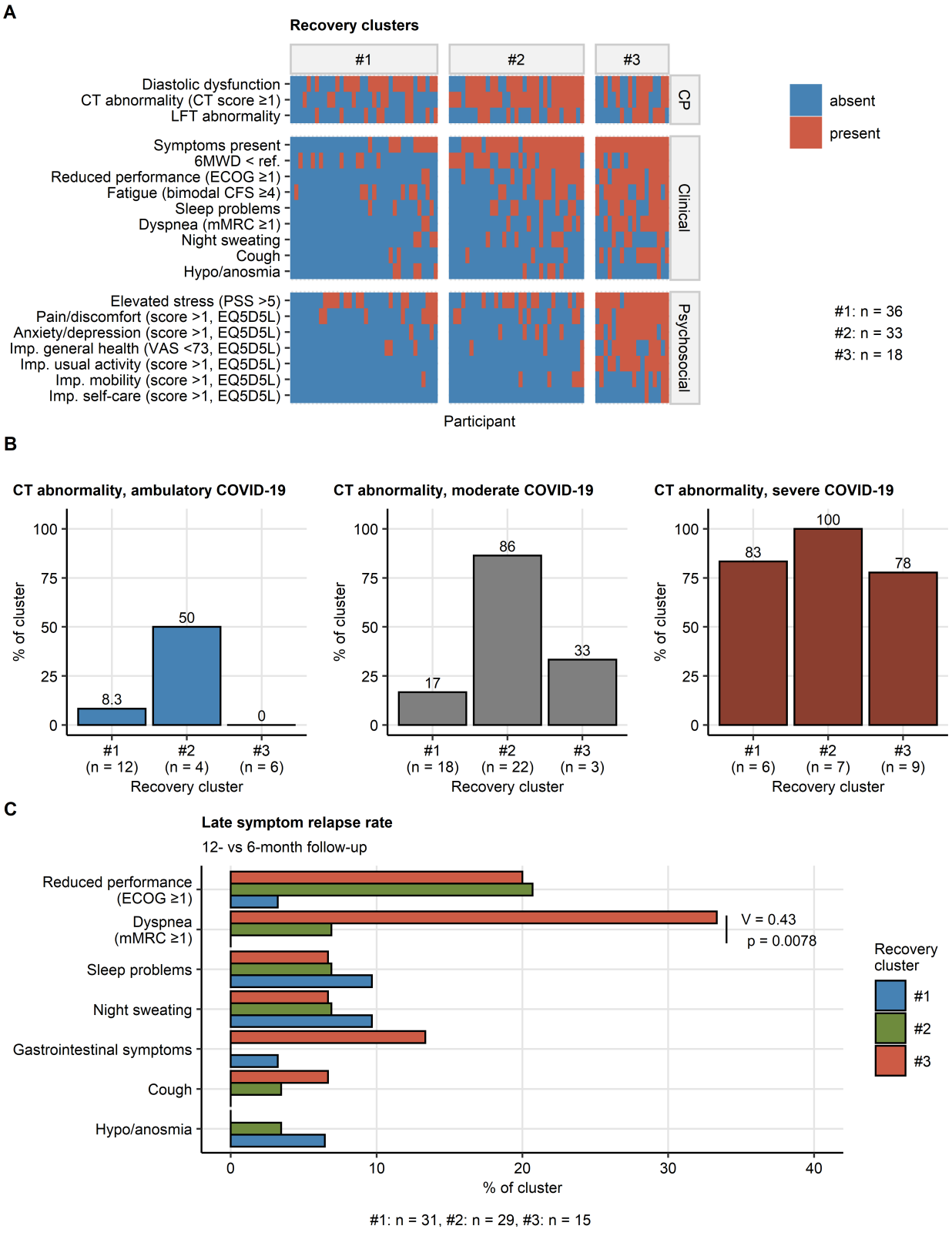
Clustering of the study participants in respect to symptoms (any symptom present, dyspnea: modified Medical Research Council score [mMRC] ≥ 1, reduced performance: Eastern

Cooperative Oncology Group score [ECOG] ≥ 1, fatigue: bimodal Chalder fatigue score [CFS] ≥ 4, self-reported: sleep problems, cough, night sweating and hyposmia/anosmia), mobility (6MWD < ref.: six minute walking distance, difference versus the reference value), cardiopulmonary abnormalities (any chest computed tomography [CT] abnormality: CT severity score ≥ 1, any lung function testing [LFT] abnormality, diastolic dysfunction), significant stress (4-item perceived stress scale [PSS] >5), impaired self-perceived general health (European quality of life 5 dimensions, 5 levels [EQ5D5L] visual analogue scale [VAS] <73, imp.: impaired) as well as features of quality of life and mental health (EQ5D5L dimensions, cutoff: score >1) at the one-year follow-up. Clustering analysis was done with the PAM algorithm (PAM: partitioning around medoids, simple matching distance [SMD]).

(A) Performance comparison of various clustering algorithms and distance measures (algorithms: PAM, HCL: hierarchical clustering, SOM+PAM: combined self-organizing map/PAM, SOM+HCL: combined SOM/HCL algorithm, distances: Tanimoto and SMD). 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 SMD metric 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.

(C) Permutation importance of the clustering factors was determined by calculating differences in clustering variances of the original clustering structure with clustering objects generated with the particular clustering factors re-shuffled randomly (Δ clustering variance). Results for 100 random permutations (runs) are presented. Each point represents a separate run. Box plots represent median differences in clustering variance with interquartile ranges (IQR). whiskers span over 150% IQR.

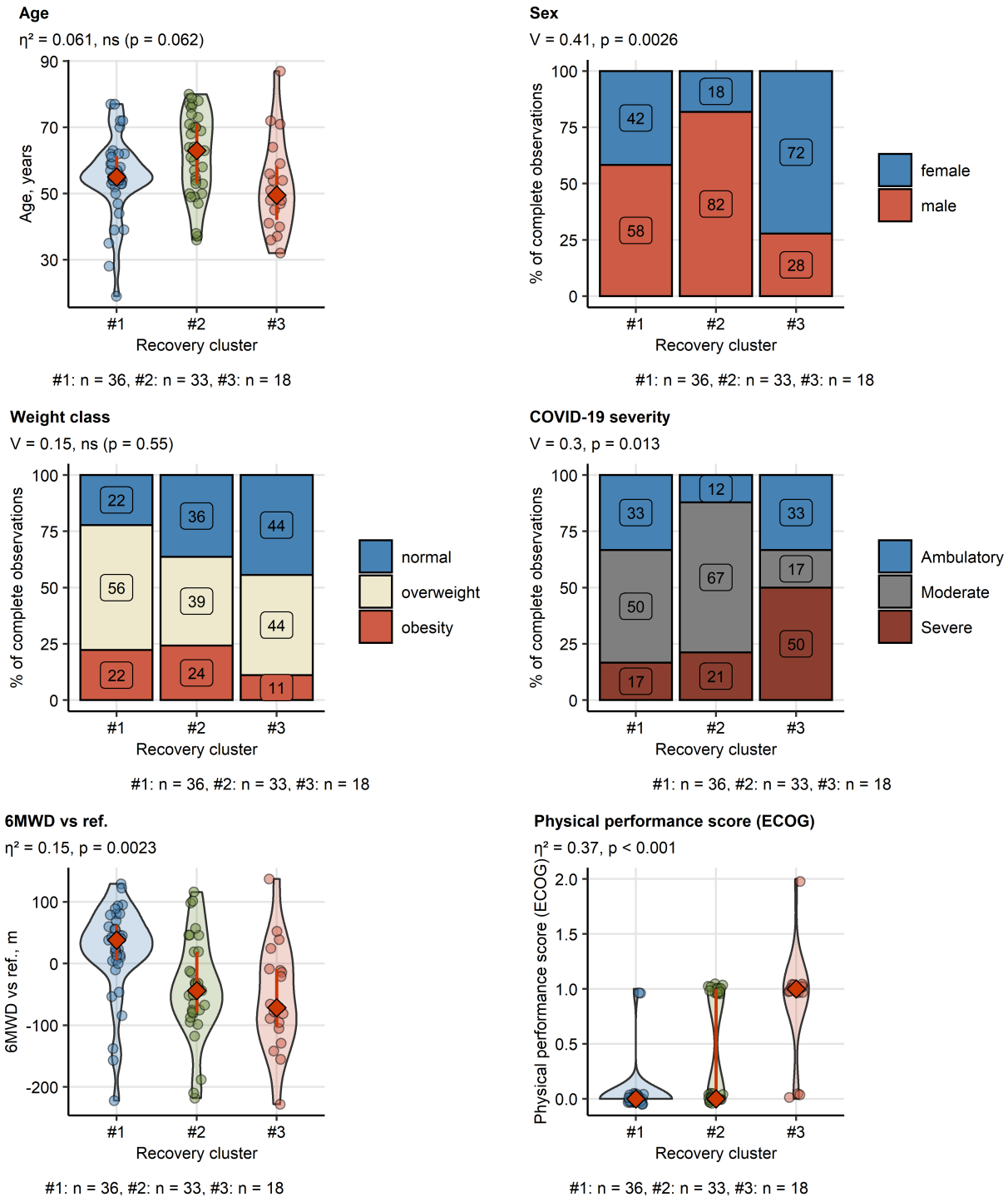

A
Recovery clusters
#1
#2
#3
Diastolic dysfunction
CT abnormality (CT score ≥1)
LFT abnormality
CP
Symptoms present
6MWD < ref.
Reduced performance (ECOG ≥1)
Fatigue (bimodal CFS ≥4)
Sleep problems
Dyspnea (mMRC ≥1)
Night sweating
Cough
Hypo/anosmia
Clinical
Elevated stress (PSS >5)
Pain/discomfort (score >1, EQ5D5L)
Anxiety/depression (score >1, EQ5D5L)
Imp. general health (VAS <73, EQ5D5L)
Imp. usual activity (score >1, EQ5D5L)
Imp. mobility (score >1, EQ5D5L)
Imp. self-care (score >1, EQ5D5L)
Psychosocial
Participant
absent
present
#1: n = 36
#2: n = 33
#3: n = 18
B
CT abnormality, ambulatory COVID-19
CT abnormality, moderate COVID-19
CT abnormality, severe COVID-19
% of cluster
8.3
50
0
17
86
33
83
100
78
#1 (n = 12)
#2 (n = 4)
#3 (n = 6)
#1 (n = 18)
#2 (n = 22)
#3 (n = 3)
#1 (n = 6)
#2 (n = 7)
#3 (n = 9)
Recovery cluster
C
Late symptom relapse rate
12- vs 6-month follow-up
Reduced performance (ECOG ≥1)
Dyspnea (mMRC ≥1)
Sleep problems
Night sweating
Gastrointestinal symptoms
Cough
Hypo/anosmia
V = 0.43
p = 0.0078
Recovery cluster
#1
#2
#3
% of cluster
#1: n = 31, #2: n = 29, #3: n = 15

Supplementary Figure S5. Heat map of clustering features, frequency of CT abnormalities and rates of symptom relapse at the 12-month follow-up in the COVID-19 recovery clusters.

(A) Cardiopulmonary (CP) abnormalities (CT abnormality: CT severity score ≥ 1, any LFT abnormality, diastolic dysfunction), mobility impairment (6MWD > ref.: six minute walking distance, difference versus the reference value), symptoms (any symptom present, dyspnea: modified Medical Research Council score [mMRC] ≥ 1, reduced performance: Eastern Cooperative Oncology Group score [ECOG] ≥ 1, fatigue: bimodal Chalder fatigue score [CFS] ≥ 4, self-reported: sleep problems, cough, night sweating and hyposmia/anosmia), significant stress (4-item perceived stress scale [PSS] >5), impaired self-perceived general health (European quality of life 5 dimensions, 5 levels [EQ5D5L] visual analogue scale [VAS] <73, imp.: impaired) as well as features of quality of life and mental health (EQ5D5L dimensions, cutoff: score >1) at the one-year follow-up in the COVID-19 recovery clusters. Presence/absence of the clustering features is presented in a heat map. Numbers of participants assigned to the clusters are displayed next to the plot.

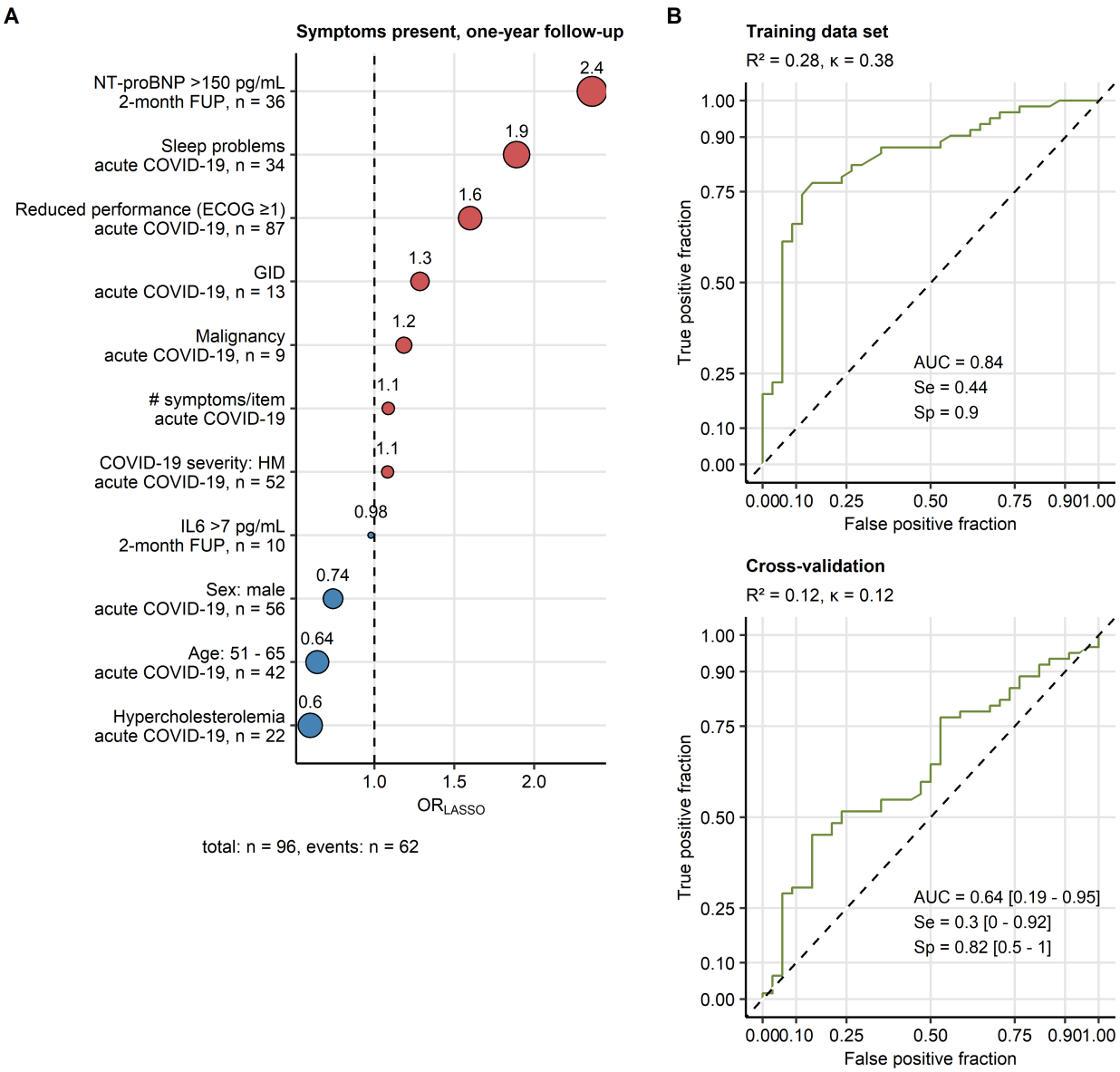
(B) Frequencies of any chest CT abnormality at the one-year follow-up in ambulatory, moderate and severe COVID-19 patients assigned to the COVID-19 recovery clusters. Percentages of complete observations in each cluster are presented in bar plots. Numbers of complete observations are indicated in the Y axis.

(C) Rates of relapse of COVID-19-related symptoms at the one-year follow-up as compared with the 6-month follow-up in the COVID-19 recovery clusters were assessed by χ^2 *test with Cramer V effect size statistic. Only participants 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. 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 plot.*



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

Age at COVID-19 diagnosis, sex, weight class distribution at COVID-19 diagnosis, acute COVID-19 severity, six-minute walking distance (6MWD vs ref.: difference between the observed and reference value) and physical performance scoring (ECOG: Eastern Cooperative Oncology Group score) at the one-year follow-up were compared between the COVID-19 recovery clusters. Statistical significance for numeric values was assessed by Kruskal-Wallis test with η^2 effect size statistic or by χ^2 test with Cramer V effect size statistic for categorical variables. 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.



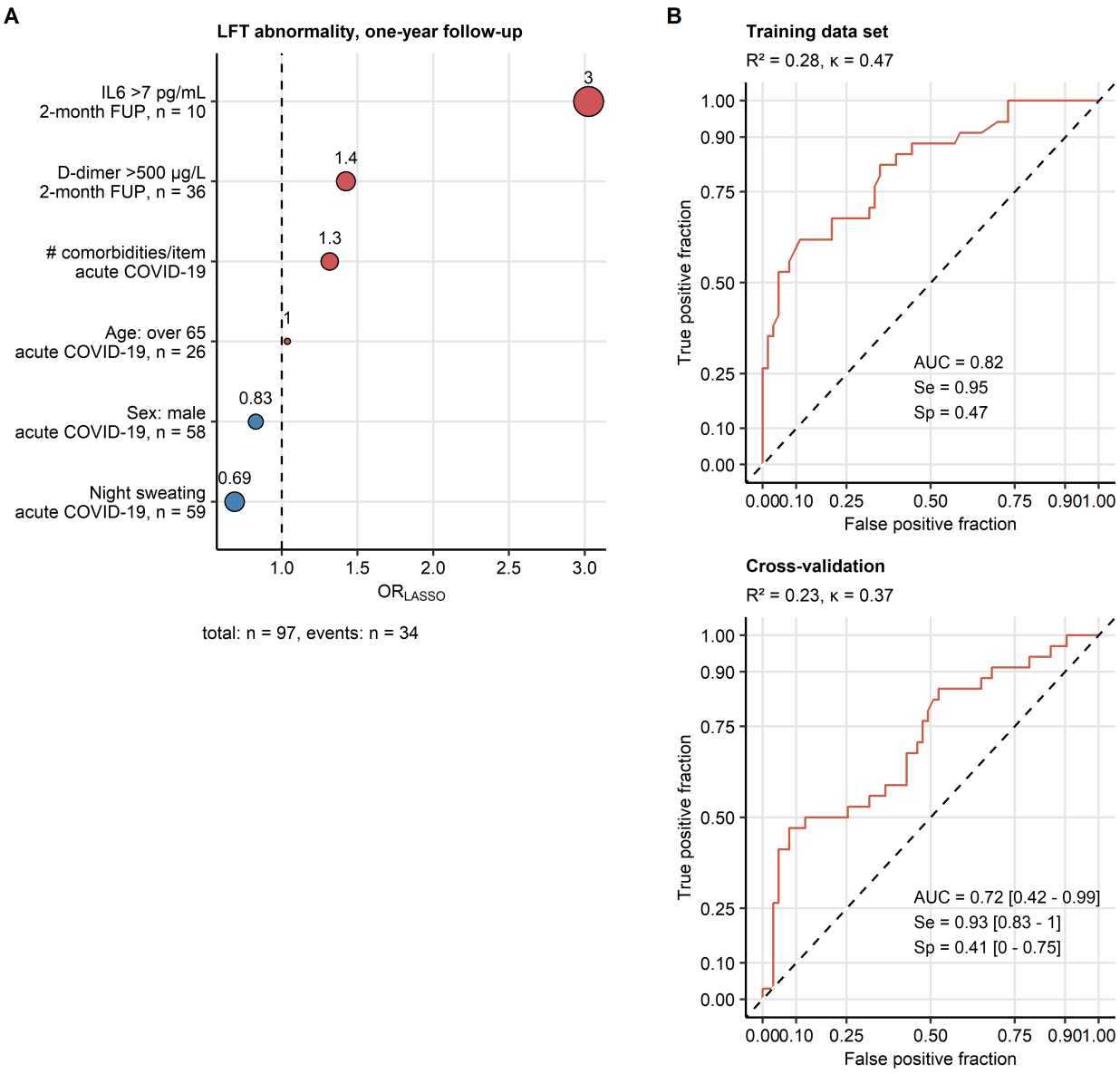
Supplementary Figure S7. Modeling of the persistent symptom risk at the one-year post-COVID-19 follow-up.

The risk of presence of COVID-19-related symptoms at the one-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 two-month follow-up (FUP). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with symptoms at the one-year follow-up 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). Time point of variable recording and number of participants in the variable strata are indicated in the Y axis.

(B) Performance of the LASSO model at predicting persistent symptoms at the one-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 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.



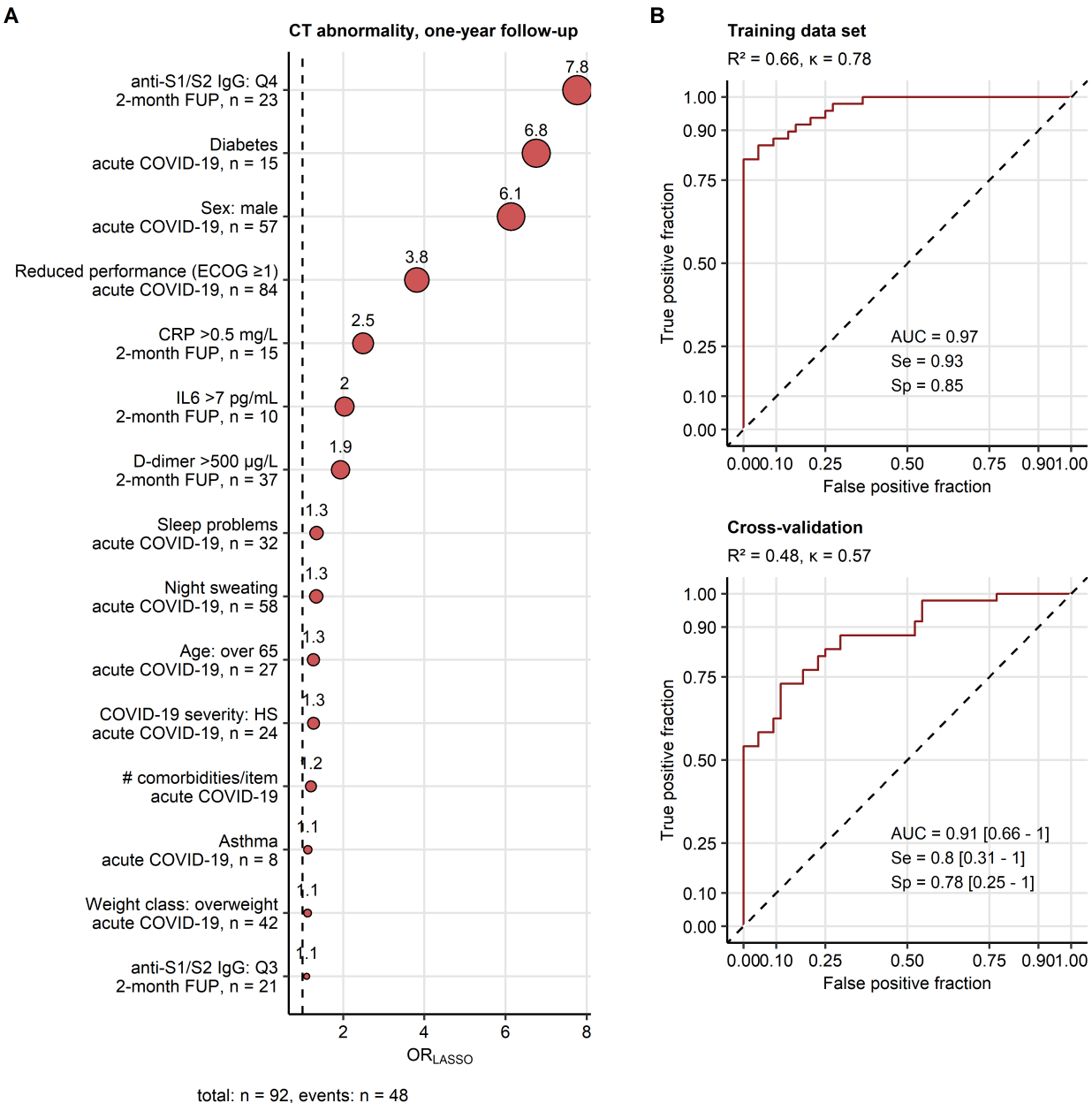
Supplementary Figure S8. Modeling of the persistent functional lung abnormality at the one-year post-COVID-19 follow-up.

The risk of any lung function testing (LFT) abnormality at the one-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 two-month follow-up (FUP). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with LFT abnormality at the one-year follow-up 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). Time point of variable recording and number of participants in the variable strata are indicated in the Y axis.

(B) Performance of the LASSO model at predicting LFT abnormality at the one-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 and Cohen's kappa statistic values are shown in the plot captions.

IL6: interleukin 6; #: number of.



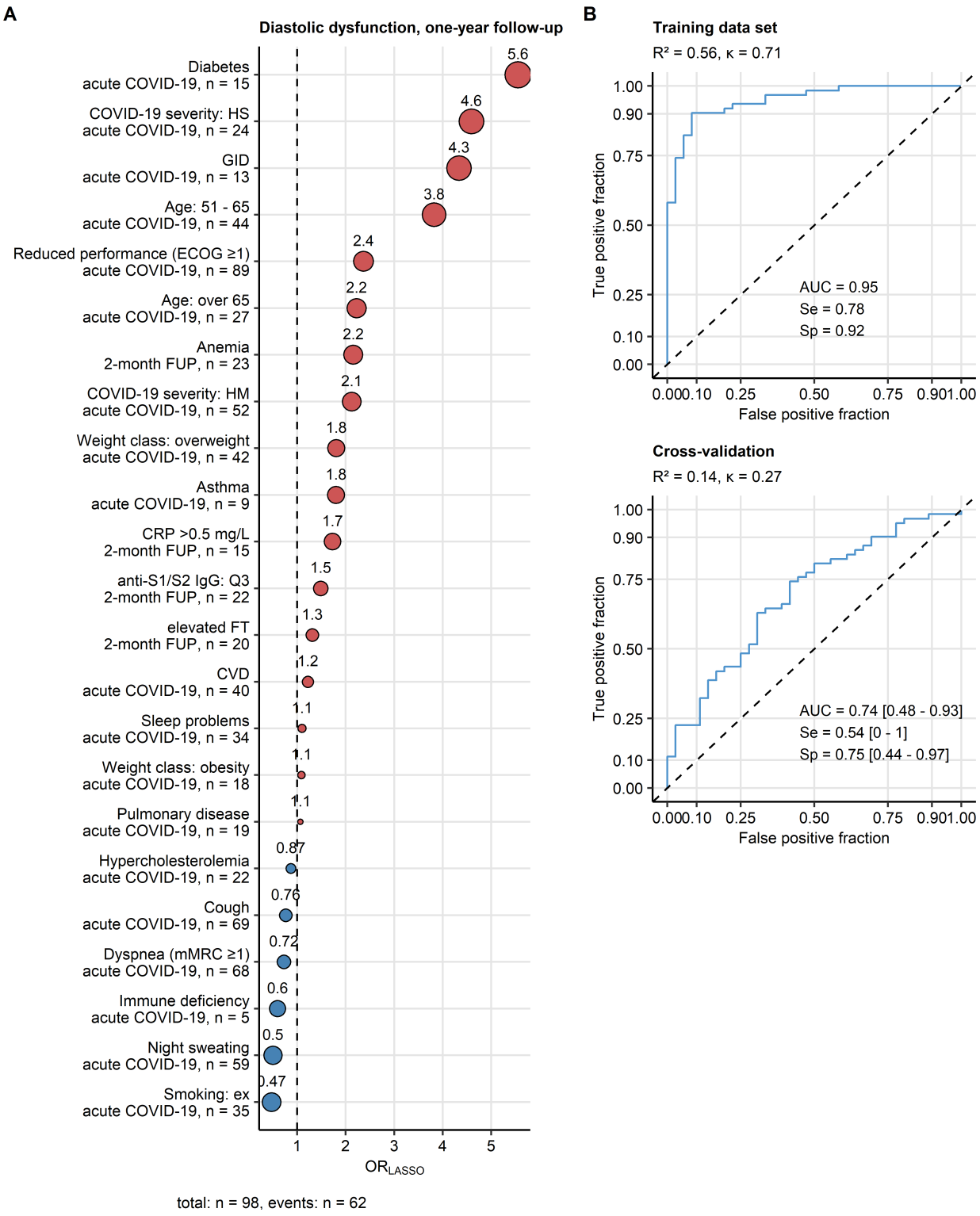
Supplementary Figure S9. Modeling of the persistent radiological lung abnormality at the one-year post-COVID-19 follow-up.

The risk of any chest computed tomography (CT) abnormality (CT severity score ≥ 1) at the one-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 two-month follow-up (FUP). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with CT abnormality at the one-year follow-up 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). Time point of variable recording and number of participants in the variable strata are indicated in the Y axis.

(B) Performance of the LASSO model at predicting CT abnormality at the one-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 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.


A
Diastolic dysfunction, one-year follow-up
Diabetes
acute COVID-19, n = 15
5.6
COVID-19 severity: HS
acute COVID-19, n = 24
4.6
GID
acute COVID-19, n = 13
4.3
Age: 51 - 65
acute COVID-19, n = 44
3.8
Reduced performance (ECOG ≥1)
acute COVID-19, n = 89
2.4
Age: over 65
acute COVID-19, n = 27
2.2
Anemia
2-month FUP, n = 23
2.2
COVID-19 severity: HM
acute COVID-19, n = 52
2.1
Weight class: overweight
acute COVID-19, n = 42
1.8
Asthma
acute COVID-19, n = 9
1.8
CRP >0.5 mg/L
2-month FUP, n = 15
1.7
anti-S1/S2 IgG: Q3
2-month FUP, n = 22
1.5
elevated FT
2-month FUP, n = 20
1.3
CVD
acute COVID-19, n = 40
1.2
Sleep problems
acute COVID-19, n = 34
1.1
Weight class: obesity
acute COVID-19, n = 18
1.1
Pulmonary disease
acute COVID-19, n = 19
1.1
Hypercholesterolemia
acute COVID-19, n = 22
0.87
Cough
acute COVID-19, n = 69
0.76
Dyspnea (mMRC ≥1)
acute COVID-19, n = 68
0.72
Immune deficiency
acute COVID-19, n = 5
0.6
Night sweating
acute COVID-19, n = 59
0.5
Smoking: ex
acute COVID-19, n = 35
0.47
1
2
3
4
5
OR_LASSO
total: n = 98, events: n = 62
B
Training data set
R² = 0.56, κ = 0.71
True positive fraction
False positive fraction
AUC = 0.95
Se = 0.78
Sp = 0.92
Cross-validation
R² = 0.14, κ = 0.27
True positive fraction
False positive fraction
AUC = 0.74 [0.48 - 0.93]
Se = 0.54 [0 - 1]
Sp = 0.75 [0.44 - 0.97]

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

The risk of diastolic dysfunction at the one-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 two-month follow-up (FUP). Study participants with the complete independent variable set were included in the analysis, numbers of complete observations and participants with diastolic dysfunction at the one-year follow-up 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). Time point of variable recording and number of participants in the variable strata are indicated in the Y axis.

(B) Performance of the LASSO model at predicting diastolic dysfunction at the one-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 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; Q3 and Q4: 3rd and 4th quartile.

References

1. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. *Journal of Open Source Software* The Open Journal; 2019; 4: 1686.

2. Henry L, Wickham Hadley. rlang: Functions for Base Types and Core R and 'Tidyverse' Features [Internet]. 2022.Available from: https://cran.r-project.org/web/packages/rlang/index.html.

3. Wickham Hadley. ggplot2: Elegant Graphics for Data Analysis [Internet]. 1st ed. New York: Springer-Verlag; 2016.Available from: https://ggplot2.tidyverse.org.

4. Wilke CO. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. 1st ed. Sebastopol: O'Reilly Media; 2019. p. 389.

5. Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J. rmarkdown: Dynamic Documents for R [Internet]. 2022.Available from: https://cran.r-project.org/web/packages/rmarkdown/index.html.

6. Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R [Internet]. 2022.Available from: https://cran.r-project.org/web/packages/knitr/index.html.

7. Xie Y. Bookdown : authoring books and technical documents with R Markdown. 2016. p. 113.

8. Gohel D. flextable: Functions for Tabular Reporting [Internet]. 2022.Available from: https://cran.r-project.org/web/packages/flextable/index.html.

9. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology* [Internet] Ovid Technologies (Wolters Kluwer Health); 1982; 5: 649–655Available from: https://www.scienceopen.com/document?vid=1125a396-3edb-4eed-9421-39241e862ed3.

10. Morriss RK, Wearden AJ, Mullis R. Exploring the validity of the chalder fatigue scale in chronic fatigue syndrome. *Journal of Psychosomatic Research* Elsevier; 1998; 45: 411–417.

11. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: Glossary of Terms for Thoracic Imaging1. *https://doi.org/10.1148/radiol.2462070712* [Internet] Radiological Society of North

America; 2008; 246: 697–722Available from: https://pubs.rsna.org/doi/10.1148/radiol.2462070712.

12. Sonnweber T, Tymoszuk P, Sahanic S, Boehm A, Pizzini A, Luger A, Schwabl C, Nairz M, Grubwieser P, Kurz K, Koppelstätter S, Aichner M, Puchner B, Egger A, Hoermann G, Wöll E, Weiss G, Widmann G, Tancevski I, Löffler-Ragg J. Investigating phenotypes of pulmonary COVID-19 recovery - a longitudinal observational prospective multicenter trial. *eLife* [Internet] Elife; 2022; 11Available from: https://pubmed.ncbi.nlm.nih.gov/35131031/.

13. Sonnweber T, Sahanic S, Pizzini A, Luger A, Schwabl C, Sonnweber B, Kurz K, Koppelstätter S, Haschka D, Petzer V, Boehm A, Aichner M, Tymoszuk P, Lener D, Theurl M, Lorsbach-Köhler A, Tancevski A, Schapfl A, Schaber M, Hilbe R, Nairz M, Puchner B, Hüttenberger D, Tschurtschenthaler C, Aßhoff M, Peer A, Hartig F, Bellmann R, Joannidis M, Gollmann-Tepeköylü C, et al. Cardiopulmonary recovery after COVID-19 - an observational prospective multi-center trial. *The European respiratory journal* [Internet] Eur Respir J; 2020; Available from: http://www.ncbi.nlm.nih.gov/pubmed/33303539.

14. Luger AK, Sonnweber T, Gruber L, Schwabl C, Cima K, Tymoszuk P, Gerstner AK, Pizzini A, Sahanic S, Boehm A, Coen M, Strolz CJ, Wöll E, Weiss G, Kirchmair R, Feuchtner GM, Prosch H, Tancevski I, Löffler-Ragg J, Widmann G. Chest CT of Lung Injury 1 Year after COVID-19 Pneumonia: The CovILD Study. *Radiology* [Internet] Radiology; 2022; Available from: https://pubmed.ncbi.nlm.nih.gov/35348379/.

15. Lancellotti P, Cosyns B. The EACVI Echo Handbook [Internet]. 2015.Available from: https://academic.oup.com/esc/book/40751.

16. Crapo RO, Casaburi R, Coates AL, Enright PL, MacIntyre NR, McKay RT, Johnson D, Wanger JS, Zeballos RJ, Bittner V, Mottram C. ATS Statement. *https://doi.org/10.1164/ajrccm.166.1.at1102* [Internet] American Thoracic Society; 2012; 166: 111–117Available from: www.atsjournals.org.

17. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. *American Journal of Respiratory and Critical Care Medicine* [Internet] American Thoracic SocietyNew York, NY; 1998; 158: 1384–1387Available from: www.atsjournals.org.

18. 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* [Internet] Springer; 2011; 20: 1727–1736Available from: https://link.springer.com/article/10.1007/s11136-011-9903-x.

19. Marten O, Greiner W. EQ-5D-5L reference values for the German general elderly population. *Health and quality of life outcomes* [Internet] Health Qual Life Outcomes; 2021; 19Available from: https://pubmed.ncbi.nlm.nih.gov/33676523/.

20. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of health and social behavior* 1983; 24: 385–396.

21. Sinclair VG, Wallston KA. The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment* [Internet] SAGE Publications; 2004; 11: 94–101Available from: https://journals.sagepub.com/doi/10.1177/1073191103258144.

22. 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)* Wiley; 1995; 57: 289–300.

23. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests [Internet]. 2021.Available from: https://cran.r-project.org/package=rstatix.

24. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation [Internet]. Comprehensive R Archive Network (CRAN); 2022.Available from: https://cran.r-project.org/package=rcompanion.

25. Harrell FE, Dupont C. Hmisc: Harrell Miscellaneous [Internet]. Comprehensive R Archive Network (CRAN); 2022.Available from: https://cran.r-project.org/package=Hmisc.

26. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* [Internet] American Statistical Association; 2015; 67: 1–48Available from: https://arxiv.org/abs/1406.5823.

27. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software* Foundation for Open Access Statistic; 2017; 82: 1–26.

28. Box GE, Hunter SJ, Hunter WG. Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley-Interscience; 2005;

29. Schubert E, Rousseeuw PJ. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* [Internet] Springer; 2019. p. 171–187Available from: https://link.springer.com/chapter/10.1007/978-3-030-32047-8_16.

30. Boriah S, Chandola V, Kumar V. Similarity measures for categorical data: A comparative evaluation. *Society for industrial and applied mathematics - 8th SIAM international conference on data mining 2008, proceedings in applied mathematics 130* [Internet] 2008. p. 243–254Available from: https://experts.umn.edu/en/publications/similarity-measures-for-categorical-data-a-comparative-evaluation.

31. Sulc Z, Cibulkova J, Rezankova H. nomclust: Hierarchical Cluster Analysis of Nominal Data [Internet]. Comprehensive R Archive Network (CRAN); 2021.Available from: https://cran.r-project.org/package=nomclust.

32. Lange T, Roth V, Braun ML, Buhmann JM. Stability-Based Validation of Clustering Solutions. *Neural Computation* MIT Press; 2004; 16: 1299–1323.

33. Leng M, Wang J, Cheng J, Zhou H, Chen X. Adaptive semi-supervised clustering algorithm with label propagation. *Journal of Software Engineering* Academic Journals; 2014; 8: 14–22.

34. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses [Internet]. 2020.Available from: https://cran.r-project.org/web/packages/factoextra/index.html.

35. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* [Internet] [Royal Statistical Society, Wiley]; 1996; 58: 267–288Available from: http://www.jstor.org/stable/2346178.

36. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* [Internet] University of California at Los Angeles; 2010; 33: 1–22Available from: https://www.jstatsoft.org/index.php/jss/article/view/v033i01/v33i01.pdf https://www.jstatsoft.org/index.php/jss/article/view/v033i01.

37. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software* American Statistical Association; 2008; 28: 1–26.

38. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* 1969; 72: 323–327.

39. Sachs MC. Plotroc: A tool for plotting ROC curves. *Journal of Statistical Software* [Internet] American Statistical Association; 2017; 79: 1–19Available from: https://www.jstatsoft.org/index.php/jss/article/view/v079c02/v79c02.pdf https://www.jstatsoft.org/index.php/jss/article/view/v079c02.