

**Persistent somatic symptoms are key to individual illness perception at one year after COVID-19 in a cross-sectional prospective cohort study**

**Supplementary Material**

CovILD study team

2023-02-22

## Supplementary Methods

### Study cohort and approval

The CovILD longitudinal observation cohort (ClinicalTrials.gov: NCT04416100) includes adult ( $\geq 18$  year) patients with a symptomatic, PCR-confirmed SARS-CoV-2 infection. The follow-up visits were scheduled at two, three, six months and one year after COVID-19 diagnosis. In the current report, a subset of the original CovILD cohort was investigated, which displayed (1) persistent somatic symptoms (PSS) or (2) any abnormality in chest computed tomography or (4) any deficits in lung function testing or (5) any cardiological abnormality in trans-thoracic echocardiography at the one-year follow-up. The final analysis inclusion criterion was availability of complete Brief Illness Perception Questionnaire (BIPQ) [1] and the complete set of variables used in explorative data analysis, modeling and clustering (**Supplementary Table S1 and S2**).

The decision to analyze the study participant subset with COVID-19 sequelae was motivated by the design of the illness perception-measuring tool used in our report, the BIPQ, which was initially developed for and validated in somatic conditions [1].

A total of 190 individuals were screened for participation. The screened subset size was limited by the number of individuals diagnosed with COVID-19 at the study centers, their potential availability for follow-up visits (residency in the study region) and available resources for the follow-up assessments. No statistical estimation of the sample size was performed, as the kinetic of COVID-19 recovery was entirely unknown at the pandemic onset. Out of the positively screened individuals, 145 participants were recruited between March and June 2020 at three centers located in Tyrol, Austria: the Medical University of Innsbruck (Innsbruck), St. Vinzenz Hospital (Zams) and Karl-Landsteiner Rehabilitation Facility (Muenster). The reasons for screening failure were denied informed consent and self-declared logistic and temporary incompatibility with the scheduled study visits. Out of 108 participants, who completed the one-year follow-up, 74 met the inclusion criteria and were analyzed here. The recruitment process scheme is presented in **Figure 1**. Significant and near significant differences between the CovILD participants included in the analysis and excluded due to lacking COVID-19 sequelae at the one-year follow-up or missing BIPQ data are listed in **Supplementary Table S3**. Baseline characteristic of the study participants included in the analysis is presented in **Table 1**, follow-up features are summarized in **Supplementary Tables S4** (symptoms, cardiopulmonary findings) and **S5** (laboratory parameters). BIPQ item scores, the sum BIPQ score, and the emotion/concern/consequences and lacking control/coherence IP component scores are summarized in **Table 2**.

The study was performed in accordance with the Declaration of Helsinki and the European Data Policy. All participants gave written informed consent. The study protocol was approved by the ethics committee at the Medical University of Innsbruck (approval number 1103/2020).

## Procedures and variables

The protocol of each follow-up visit included a survey of PSS, blood laboratory parameter assessment, lung function testing, trans-thoracic echocardiography and chest computed tomography as described in more detail before [2–4]. Comorbidities and data on acute COVID-19 course (symptoms, laboratory parameters, functional tests and imaging) were recorded retrospectively at the two-month follow-up visit based on information provided by the participant and electronic patient records.

Variables used in the exploratory data analysis and multi-parameter modeling are listed in **Supplementary Table S1** and **Supplementary Table S2**.

### Baseline demographic and clinical status, acute COVID-19 severity

The baseline demographic variables included sex (male/female), age at COVID-19 diagnosis, smoking history, body weight and height. The following pre-existing comorbidities were surveyed: obesity, metabolic disease, cardiovascular disease, hypertension, type II diabetes, hypercholesterolemia, gastrointestinal disease, malignancy, chronic kidney disease, immune deficiency, pulmonary and respiratory conditions. Among the surveyed comorbidities, asthma, chronic obstructive pulmonary disease, interstitial lung disease and other chronic pulmonary disorders were subsumed under the ‘respiratory disease’ variable included in the independent modeling variable set. Study participants were stratified according to the acute COVID-19 severity as ambulatory (outpatient, WHO ordinal scale for clinical improvement 1 - 2), moderate (hospitalized at normal infection ward, no oxygen therapy, WHO 3 - 4) and severe COVID-19 survivors (hospitalized with oxygen therapy or mechanical ventilation or intensive care, WHO 5 - 7).

### Cardiopulmonary assessment

Lung function testing abnormality was defined as forced vital capacity < 80% or forced expiratory volume in 1 second < 80% or total lung capacity < 80% or diffusion lung capacity for carbon monoxide < 80% of the predicted reference value or the ratio of forced expiratory volume in 1 second to forced vital capacity < 70% of the predicted reference value [3]. Computed tomography images were screened for ground glass opacities, consolidations, bronchiectasis and reticulations according to the Fleischner Society glossary [5]. Chest computed tomography abnormalities were scored separately for each lobe with the computed tomography severity score [2–4]: 0 - no abnormality, 1 - minimal

(subtle ground glass opacities), 2 - mild (several ground glass opacities, subtle reticulation), 3 - moderate (multiple ground glass opacities, reticulation, small consolidation), 4 - severe (extensive ground glass opacities, consolidation, reticulation with distortion), 5 - massive (multiple findings, parenchymal destruction). The sum computed tomography severity score for all five lobes was used in the analysis. Any chest computed tomography abnormality is defined as computed tomography severity score  $\geq 1$ . The most frequent cardiological abnormality in echocardiography was low grade diastolic dysfunction. No reduced left ventricular ejection fraction was observed in the participants included in the current analysis. Frequencies of cardiopulmonary findings at the one-year follow-up are presented in **Supplementary Table S4** and **Supplementary Figure S2**.

### Persistent somatic symptoms and exertional capacity

The following features were surveyed as PSS at the one-year follow-up:

- **reduced physical performance** defined as Eastern Cooperative Oncology Group (ECOG) scale  $\geq 1$
- **dyspnea** defined as Modified Medical British Research Council (mMRC) scale  $\geq 1$
- **cough**, self-reported, surveyed as a yes/no item
- **sleep problems**, self-reported, surveyed as a yes/no item
- **night sweating**, self-reported, surveyed as a yes/no item
- **hyposmia or anosmia**, self-reported, surveyed as a yes/no item
- **dermatological symptoms**, self-reported, surveyed as a yes/no item
- **gastrointestinal symptoms**, self-reported, surveyed as a yes/no item
- **hair loss**, self-reported, surveyed as a yes/no item
- **significant fatigue**, defined as bimodal Chalder's Fatigue Scale  $\geq 4$  [6,7]

Records of reduced physical performance, dyspnea, cough, sleep problems, night sweating, hyposmia/anosmia and gastrointestinal symptoms were available for acute disease and all scheduled study visits [3].

Since only few participants displayed values of Eastern Cooperative Oncology Group scale  $\geq 2$  or Modified Medical British Research Council scale  $\geq 2$ , the dichotomous 'reduced performance (ECOG  $\geq 1$ )' and 'dyspnea (mMRC  $\geq 1$ )' independent variables were used in modeling instead of the full numeric scales. Fatigue intensity was rated with the likert Chalder's Fatigue Scale [6,7]. Exertional capacity was assessed by six-minute walking test performed according to the American Thoracic Society guidelines [8]. The reference six-minute walking distance values were obtained as described by Crapo and colleagues [8] and differences between the actual and reference values were calculated. Frequencies of symptoms, ratings of fatigue and exertional capacity at the one-year follow-up are shown in **Supplementary Table S4**, **Supplementary Figures S1** and **S2**.

## Laboratory parameters

Laboratory parameters encompassing hemoglobin (total and glycated hemoglobin), biomarkers of iron turnover (ferritin, transferrin saturation, hepcidin and soluble transferrin receptor), inflammation and coagulation (C-reactive protein and D-dimer) and cardiovascular pathology (N-terminal pro-brain natriuretic peptide) were determined routinely at the certified laboratory of the Central Institute for Medical and Chemical Diagnostic at the University Hospital in Innsbruck. Other markers of systemic inflammation, interleukin 6 and procalcitonin, were within the normal ranges in the analysis collective. Values of laboratory parameters at the one-year follow-up are listed in **Supplementary Table S5**.

## Illness perception

Illness perception (IP) was scored with a German translation of BIPQ [1]. Each of the 8 BIPQ items (questions Q1 - Q8) was scored in with an 11-point likert scale, the items Q3, Q4 and Q7 were conceptualized as negative items are were inverted prior to the analysis:

- **Q1 consequences** (0: no consequences, 10: very severe consequences)
- **Q2 timeline** (0: very short disease duration, 10: forever)
- **Q3 lacking personal control** (negative item, 0: full control, 10: no control at all)
- **Q4 lacking treatment control** (negative item, 0: full control, 10: no control at all)
- **Q5 identity** (0: no complaints, 10: multiple severe complaints)
- **Q6 concerns** (0: no concerns at all, 10: extreme concerns)
- **Q7 lacking coherence** (negative item, 0: very good disease understanding, 10: no understanding at all)
- **Q8 emotional representation** (0: no emotions at all, 10: extreme emotions)

The total IP score was calculated as an arithmetic sum of items Q1 - Q8. As shown by the results of factor analysis [9] of the BIPQ tool items (**Supplementary Figure S3**), two separate factors could be identified: the emotion/concern/consequences (items Q1, Q2, Q5, Q6, Q8) and the lacking control/coherence component (items Q3, Q4, Q7). For this reason, two IP component scores were calculated and analyzed: the emotion/concern/consequences score defined as the Q1, Q2, Q5, Q6, Q8 item sum and the lacking control/coherence score defined as the Q3, Q4, Q7 item sum. Values of the total IP score, the emotion/concern/consequences and lacking control/coherence IP component scores as well as values of the BIPQ items recorded at the one-year follow-up are presented in **Table 2**.

As measured by McDonald's total  $\omega$  values [10], the full BIPQ tool ( $\omega = 0.9$ ), the emotion, concern, consequence ( $\omega = 0.94$ ) and the lacking control/coherence ( $\omega = 0.71$ ) demonstrated acceptable-to-good internal consistency (**Supplementary Figure S4**).

## Software

Data import, transformation, statistical analysis and result visualization was done with R version 4.2.0 (R Foundation for Statistical Computing). For import, handling and transformation of tabular data, the *tidyverse* package bundle [11] and *rlang* package [12] was used. Handling of text variables (search, replacement) was done with *stringi* [13].

For descriptive statistic, correlation and statistical testing for differences between analysis groups the package *rstatix* [14] and the development package *ExDA* were employed. Factor analysis was accomplished with the *stats* package of base R and the development package *clustTools*. The BIPQ score consistency was investigated with the *psych* package [15].

For modeling and ANOVA, the development package *lmtree* and the packages *glmnet* (Elastic Net and LASSO) [16], *monomvn* [17] (Bayesian LASSO) and *caret* [18] were employed. Model quality control, cross-validation and fit statistics were computed with *caret* and the development package *caretExtra*.

Clustering was accomplished with tools provided by the packages *cluster* [19], *philentropy* [20], *factoextra* [21] and the development package *clustTools*.

The *ggplot2* [22] package were used for data visualization. Bar and violin plots were generated with *ExDA*. Quasi-proportional Venn diagrams were generated with the *nVennR* package [23]. Figures were constructed with *cowplot* [24]. Tables were generated with *flextable* [25]. The manuscript and supplementary material files were created with the *rmarkdown* [26], *knitr* [27], *bookdown* [28] packages and *Affiliations* Pandoc filter set. For management of figures, tables and R code in the RMarkdown documents, the development package *figur* was used.

## Descriptive statistic, distribution testing and hypothesis testing

Descriptive statistic: median with interquartile ranges for numeric variables, and percentages and counts for categorical variables were calculated with the *explore()* function (*ExDA* package). Statistical significance of differences in frequency of categorical variable levels was assessed by  $\chi^2$  test with Cramer V effect size statistic (*compare\_variables()*, the *ExDA* package). Normality was tested with Shapiro-Wilk test and quantile-quantile plots (*explore()*, *ExDA* package).

Since multiple study variables were non-normally distributed, differences in numeric variables were assessed by non-parametric Mann-Whitney (effect size: r) or Kruskal-Wallis

test (effect size:  $\eta^2$ ), as appropriate. Group comparisons were done with the `compare_variables()` function (package *ExDA*). Correlations were analyzed with Spearman test (`correlate_variables()` package *ExDA*) and visualized by fitting a second order trend (`geom_smooth()`, package *ggplot*, formula:  $y \sim x + I(x^2)$ ). Statistical significance for changes in frequency of PSS in time was investigated with Cochran Q test (function `cochran_qtest()`, package *rstatix*) [14].

## BIPQ tool factor analysis and consistency assessment

BIPQ dimensionality was assessed by maximum likelihood factor analysis [9] (`reduce_data(red_fun = 'fa')`, package *clustTools*, a wrapper around `factanal()` from *stats*) and visual investigation of the BIPQ item loadings (method `plot(type = 'loadings')` called for the factor analysis object). The factor analysis revealed presence of two components of the BIPQ tool: the emotion/concern/consequences and the lacking control/coherence factor (**Figure 3**). Spearman's pair-wise correlations between the BIPQ items were calculated (function `correlate_variables()`, package *ExDA*) and visualized in a bubble plot. Due to multi-dimensionality of the BIPQ tool McDonald's  $\omega$  was calculated with the function `omega()` from the package *psych* [10,15] was used as a reliability measure for the global BIPQ and the emotion/concern/consequences and lacking control/coherence factors (**Supplementary Figure S4**).

## Multi-parameter modeling

In multi-parameter modeling, a set of 56 independent variables was used (**Supplementary Table S2**). To improve normality, ferritin and soluble transferrin receptor concentrations were transformed with the logarithm function, and transferrin saturation and hepcidin were transformed with square root. Strongly non-normally or bimodally distributed numeric explanatory variables: number of comorbidities (0, 1-3, 4+), number of symptoms (0, 1, 2-3, 4+), computed tomography severity score ('none': 0, 'mild': 1 - 5, 'moderate': 5 - 10, 'severe': 11+) [4], N-terminal pro-brain natriuretic peptide (normal, >125 pg/mL), D-Dimer (normal, >460 µg/L), C-reactive protein (normal, >460 µg/L) and glycated hemoglobin (normal, >460 µg/L) were stratified with clinically applicable cutoffs or cutoffs determined by visual analysis of scatter plots of modeling variable and the explanatory factor with fitted LOESS trends. For numeric explanatory variables, both first- and second-order terms were included in the model to account for possible non-linear relationship between the independent variable and response. During pre-processing, numeric variables (independent and responses) were normalized with mean centering (Z-scores).

The modeling responses: (1) the total IP score (all BIPQ item sum), (2) the emotion/concern/consequences (Q1, Q2, Q5, Q6 and Q8 BIPQ item sum) and (3) the lacking control/coherence (Q3, Q4 and Q7 BIPQ item sum) IP component scores were square root-transformed to guarantee normal distribution.

Three L1 regularized regression algorithms were utilized to model the total IP score and the IP component scores at one year after COVID-19: Elastic Net [29], LASSO (least absolute shrinkage and selection operator) [30] and Bayesian LASSO [31]. The choice of penalized regression over canonical backward elimination or forward selection was motivated by highly multi-dimensional nature of the data set [32]. As expected, modeling of the IP responses with canonical linear regression with Akaike information criterion-driven backward elimination (function `stepAIC()`, package MASS) [33] led to highly over-parameterized models as evident from huge expansion of the fit errors in cross-validation (not shown).

The Elastic Net and LASSO algorithms are implemented in R by the *glmnet* package [16]. The optimal  $\lambda$  parameters for Elastic Net and LASSO regression were found as values associated with the minimum model deviance of the optimally regularized model in 200-repeats 10-fold CV (function `cv.glmnet()`, Elastic Net:  $\alpha = 0.5$ , LASSO:  $\alpha = 1$ ,  $\lambda$  parameter 'lambda.1se' corresponding to the optimally regularized model). The optimal  $\lambda$  values in Elastic Net were 0.42 for the total IP score, 0.24 for the emotion/concern/consequences and 0.62 for the lacking control/coherence IP component scores. The optimal  $\lambda$  values in LASSO were 0.25 for the total IP score, 0.1 for the emotion/concern/consequences and 0.31 for the lacking control/coherence IP component score. Elastic Net and LASSO models with the optimal  $\lambda$  values were constructed with *caret* package (function `train(method = 'glmnet')`) [18] and their performance in the training data set and 10-repeats 10-fold cross-validation was tested. As shown in **Supplementary Figure S5**, for the total IP score and the emotion/concern/consequences IP component score, meaningful, reproducible and acceptably parameterized models could be built with either of the algorithms as suggested by comparison of the model performance in the training and the cross-validation data sets. In turn, poorly performing, estimate-only models could be obtained for the lacking control/coherence IP component. Estimates of non-zero coefficients were extracted from the final models with the `coef()` method (the `s` parameter set to the optimal  $\lambda$  value) (**Supplementary Figures S6 - S11**).

The Bayesian LASSO algorithm is implemented in R by the *monomvn* package [17]. The Bayesian LASSO model was constructed with the *caret* package (function `train(method = 'blasso')`, iteration number:  $T = 1000$ ). The optimal value of the sparsity parameter, which controls the fraction of non-zero coefficients in the final model based on the posterior distributions, was found by 10-repeats 10-fold cross-validation with the root mean squared error as a selection criterion. The optimal sparsity values selected from the 0.1 - 0.7 range were 0.5, 0.5 and 0.4 for the total IP score, the emotion/concern/consequences and the lacking control/coherence IP component scores, respectively. Similar to the Elastic Net and LASSO algorithms, meaningful models could be obtained for the total IP score and the emotion/concern/consequences IP component score but not for the lacking control/coherence IP component score (**Supplementary Figure S5**). The model coefficient matrix X was extracted from the `finalModel` slot of the *caret* model. The coefficient values (**Supplementary Figures S8 and S11**) were calculated as matrix column medians over all algorithm iterations [31].

Extraction of caret model performance measures: root mean squared error and  $R^2$  defined as  $1 - \text{mean squared error} / \text{variance}(y)$  (**Supplementary Figure S5**) was accomplished by the in-house-developed package `caretExtra` (modeling objects created with `as_caretex()`). Model assumption testing (normality of residuals, Shapiro-Wilk test, method `residuals()`) and visual quality control based on plots of residuals versus fitted and residuals quantile-quantile plots was done with the `caretExtra` package (method `plot()` called for the `caretex` object).

The number of non-zero model coefficients was found to differ substantially between the regularized regression models (**Supplementary Figures S6 - S11**), which was reported for data sets of similar size [34]. Hence, the key factors impacting on the total IP score and the emotion/concern/consequences IP component at one year after COVID-19 (number of PSS, reduced performance (ECOG  $\geq 1$ ) and fatigue score (likert CFS)) were defined as variables with non-zero coefficients in all multi-parameter models. Since no meaningful models for the lacking control/coherence could be obtained, no influential factors for the lacking control/coherence component could be identified (**Figure 2**).

## Analysis of interaction with two-way ANOVA

Interaction between the key factors affecting the total IP score and the emotion/concern/consequences IP component score and presence of cardiopulmonary abnormalities (computed tomography, lung function findings or diastolic dysfunction in echocardiography) was analyzed by two-way ANOVA. Briefly, the interaction ANOVA models were constructed with the function `make_lm()` (employing `lm()` from base R, package `lmgc`). The dependent variables were square root-transformed to improve normality. The assumptions of normality and homogeneity of distribution of the model residuals were checked by Shapiro-Wilk and Levene tests, respectively (method `summary(type = 'assumptions')`) and standard diagnostic plots of model residuals (method `plot()`). ANOVA results with the  $\eta^2$  effect size statistic (fraction of variance explained by the specific ANOVA term) were extracted from the models with the `anova()` method (package `lmgc`). The significant and near significant ( $p < .1$ ) interactions are presented in **Supplementary Figures S14 - S16**.

## Clustering analysis

Study participants were clustered in respect to the BIPQ items by the partition around medoids algorithm (originally implemented by the package `cluster`) [19] with Euclidean distance between the observations (originally implemented by the package `philentropy`) [20]. The BIPQ item scores were not pre-processed, since they were measured with the same likert scale. Technically, the clustering object was constructed with the `kcluster(distance_method = 'euclidean', clust_fun = 'pam')` function from the development package `clustTools`.

The partition around medoids/Euclidean distance procedure was chosen based on the high explained clustering variance (ratio of the between-cluster sum of squares to the total sum of squares, method `var()`, package *clustTools*) and stability (cluster assignment accuracy in 10-fold cross-validation, cluster assignment by 7-nearest neighbors classifier, method `cv()`, package *clustTools*) [35,36] as compared with several other clustering algorithms (hierarchical clustering and KMEANS) (**Supplementary Figure S17A**). The choice of cluster number ( $k = 3$ ) was based on the bend of the curve of within-cluster sum of squares (method `plot()`, package *clustTools*, a wrapper around `fviz_nbclust()`, package *factoextra*) [21] and visual analysis of the distance heat map (**Supplementary Figure S17BC**).

## Data and code availability

An R data (RDa) file with anonymized patient data will be made available upon request to the corresponding author. The study analysis pipeline is available at <https://github.com/PiotrTymoszuk/CovILD-IPQ>.

## Supplementary Tables

**Supplementary Table S1: Study variables. The table is available as a supplementary Excel file.**

**Supplementary Table S2: Initial set of candidate explanatory variables in modeling of illness perception.**

Explanatory variable
Sex, Age, Age <sup>2</sup> , Smoking history, Weight class, Comorbidity present, # comorbidities, CVD, comorbidity, Hypertension, Metabolic comorbidity, Hypercholesterolemia, Diabetes, comorditity, Respiratory comorbidity, CKD, comorbidity, GID, comorbidity, Malignancy, comorbidity, Immune deficiency, comorbidity, COVID-19 severity, PSS present, # PSS, Reduced performance, Dyspnea, Cough, Sleep problems, Night sweat, Hypo/anosmia, Dermatological sympt., Gastrointestinal sympt., Hair loss, Fatigue score (likert CFS), (Fatigue score (likert CFS)) <sup>2</sup> , Fatigue (bimodal CFS $\geq$ 4), SMWD, SMWD <sup>2</sup> , SMWD vs ref., (SMWD vs ref.) <sup>2</sup> , LFT abnormality, CT abnormality, CT lesion severity, Diastolic dysfunction, Rehabilitation, Anemia, Hb, Hb <sup>2</sup> , log FT, log FT <sup>2</sup> , sqrt TF-Sat, TF-Sat, log sTFR, log sTFR <sup>2</sup> , sqrt Hepcidin, Hepcidin, NT-proBNP, D-dimer, CRP, HbA1c

**Supplementary Table S3: Significant and near-significant ( $p < .1$ ) factors differentiating between the CovILD cohort participants included in the analysis and excluded due to variable missingness or lacking long-term COVID-19 sequelae. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are shown as percentages and counts within the strata.**

Variable	Included in the analysis	Excluded from the analysis	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	74	34		
WHO COVID-19 severity <sup>b</sup>	4 [IQR: 3 - 4.8] range: 2 - 7	3 [IQR: 2 - 4] range: 2 - 7	ns (p = .05)	r = .19
Dermatological symptoms	9.5% (n = 7)	24% (n = 8)	ns (p = .10)	V = .19
Gastrointestinal symptoms	2.7% (n = 2)	15% (n = 5)	ns (p = .05)	V = .23
SMWD, m <sup>b</sup>	560 [IQR: 500 - 630] range: 270 - 760	520 [IQR: 450 - 580] range: 310 - 680	p = .03	r = .2
Hb, g/dL <sup>b</sup>	150 [IQR: 140 - 160] range: 110 - 180	140 [IQR: 130 - 150] range: 110 - 170	ns (p = .09)	r = .16
FT, µg/L <sup>b</sup>	160 [IQR: 110 - 260] range: 13 - 1300	110 [IQR: 60 - 180] range: 22 - 800	ns (p = .05)	r = .19

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Mann-Whitney test with r effect size statistic

<sup>b</sup>WHO ordinal scale for clinical improvement

<sup>b</sup>Six minute walking distance

<sup>b</sup>Blood hemoglobin

<sup>b</sup>Serum ferritin

*Supplementary Table S4: COVID-19 symptoms, performance, fatigue, exertional capacity, cardiopulmonary abnormalities and rehabilitation status at the one-year follow-up. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are shown as percentages and counts within the strata.*

Variable	Cohort	Ambulatory COVID-19	Moderate COVID-19	Severe COVID-19	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	74	15	40	19		
PSS present <sup>c</sup>	72% (n = 53)	80% (n = 12)	72% (n = 29)	63% (n = 12)	ns (p = .55)	V = .13
Number of PSS <sup>c</sup>	2 [IQR: 0 - 3] range: 0 - 6	2 [IQR: 1 - 4] range: 0 - 6	1.5 [IQR: 0 - 3] range: 0 - 6	2 [IQR: 0 - 3] range: 0 - 6	ns (p = .70)	$\eta^2 = .018$
Reduced performance (ECOG ≥ 1) <sup>c</sup>	35% (n = 26)	33% (n = 5)	35% (n = 14)	37% (n = 7)	ns (p = .98)	V = .025
Dyspnea (mMRC ≥ 1) <sup>d</sup>	22% (n = 16)	13% (n = 2)	22% (n = 9)	26% (n = 5)	ns (p = .65)	V = .11
Cough	12% (n = 9)	6.7% (n = 1)	15% (n = 6)	11% (n = 2)	ns (p = .68)	V = .1
Sleep problems	32% (n = 24)	47% (n = 7)	25% (n = 10)	37% (n = 7)	ns (p = .28)	V = .19
Night sweat	18% (n = 13)	33% (n = 5)	15% (n = 6)	11% (n = 2)	ns (p = .18)	V = .21
Hypo/anosmia	14% (n = 10)	20% (n = 3)	18% (n = 7)	0% (n = 0)	ns (p = .13)	V = .23
Dermatological symptoms	9.5% (n = 7)	6.7% (n = 1)	7.5% (n = 3)	16% (n = 3)	ns (p = .55)	V = .13
Gastrointestinal symptoms	2.7% (n = 2)	6.7% (n = 1)	2.5% (n = 1)	0% (n = 0)	ns (p = .49)	V = .14
Hair loss	8.1% (n = 6)	13% (n = 2)	7.5% (n = 3)	5.3% (n = 1)	ns (p = .68)	V = .1
Fatigue score (likert CFS) <sup>e</sup>	12 [IQR: 11 - 16] range: 1 - 32	14 [IQR: 11 - 19] range: 2 - 26	12 [IQR: 11 - 15] range: 1 - 24	13 [IQR: 11 - 19] range: 1 - 32	ns (p = .54)	$\eta^2 = .011$
Fatigue (bimodal CFS ≥ 4) <sup>e</sup>	41% (n = 30)	53% (n = 8)	32% (n = 13)	47% (n = 9)	ns (p = .29)	V = .18
LFT abnormality <sup>f</sup>	32% (n = 24)	20% (n = 3)	30% (n = 12)	47% (n = 9)	ns (p = .21)	V = .2
CT abnormality (CT score ≥ 1)	54% (n = 40)	13% (n = 2)	52% (n = 21)	89% (n = 17)	p < .001	V = .52
CT severity score <sup>g</sup>	1 [IQR: 0 - 5] range: 0 - 14	0 [IQR: 0 - 0] range: 0 - 5	1 [IQR: 0 - 2.2] range: 0 - 8	5 [IQR: 2.5 - 10] range: 0 - 14	p < .001	$\eta^2 = .35$
Diastolic dysfunction	64% (n = 47)	27% (n = 4)	68% (n = 27)	84% (n = 16)	p = .002	V = .41
SMWD, m <sup>h</sup>	560 [IQR: 500 - 630] range: 270 - 760	620 [IQR: 560 - 650] range: 400 - 740	550 [IQR: 480 - 630] range: 270 - 760	560 [IQR: 500 - 640] range: 410 - 700	ns (p = .11)	$\eta^2 = .034$
SMWD vs reference, m <sup>i</sup>	0 [IQR: -61 - 46] range: -230 - 140	0 [IQR: -55 - 31] range: -230 - 120	0 [IQR: -56 - 46] range: -220 - 130	0 [IQR: -71 - 42] range: -210 - 140	ns (p = .85)	$\eta^2 = .024$

<b>Variable</b>	<b>Cohort</b>	<b>Ambulatory COVID-19</b>	<b>Moderate COVID-19</b>	<b>Severe COVID-19</b>	<b>Significance<sup>a</sup></b>	<b>Effect size<sup>a</sup></b>
Rehabilitation	31% (n = 23)	6.7% (n = 1)	18% (n = 7)	79% (n = 15)	p < .001	V = .61

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic

---

<sup>c</sup>PSS: persistent somatic symptoms

---

<sup>c</sup>ECOG: Eastern Cooperative Oncology Group performance score

---

<sup>d</sup>mMRC: Modified Medical Research Council dyspnea score

---

<sup>e</sup>CFS: Chalder's Fatigue Score

---

<sup>f</sup>Lung function testing abnormality: FVC < 80% or FEV1 < 80% or TLC < 80% or DLCO < 80% predicted or FEV1:FVC < 70% predicted reference value

---

<sup>g</sup>Chest computer tomography severity score

---

<sup>h</sup>SMWD: Six minute walking distance

---

<sup>i</sup>SMWD, difference between the actual and reference value

*Supplementary Table S5: Laboratory parameters at the one-year follow-up. Values are presented as medians with interquartile ranges (IQR).*

Variable	Cohort	Ambulatory COVID-19	Moderate COVID-19	Severe COVID-19	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	74	15	40	19		
Anemia <sup>b</sup>	11% (n = 8)	0% (n = 0)	15% (n = 6)	11% (n = 2)	ns (p = .28)	V = .19
Hb, g/dL <sup>c</sup>	150 [IQR: 140 - 160] range: 110 - 180	140 [IQR: 140 - 150] range: 130 - 160	150 [IQR: 140 - 160] range: 110 - 170	160 [IQR: 150 - 160] range: 140 - 180	p = .003	$\eta^2 = .14$
FT, µg/L <sup>d</sup>	160 [IQR: 110 - 260] range: 13 - 1300	140 [IQR: 48 - 170] range: 13 - 450	170 [IQR: 130 - 270] range: 13 - 1300	180 [IQR: 110 - 250] range: 25 - 870	ns (p = .19)	$\eta^2 = .018$
TF-Sat, % <sup>e</sup>	26 [IQR: 20 - 33] range: 6 - 61	28 [IQR: 18 - 32] range: 14 - 49	26 [IQR: 21 - 33] range: 6 - 61	24 [IQR: 20 - 32] range: 10 - 44	ns (p = .82)	$\eta^2 = .022$
sTFR, mg/L <sup>f</sup>	3 [IQR: 2.4 - 3.4] range: 1.6 - 6.2	2.8 [IQR: 2.3 - 3.2] range: 2 - 4.1	3 [IQR: 2.4 - 3.3] range: 1.6 - 6	3 [IQR: 2.5 - 3.4] range: 1.9 - 6.2	ns (p = .49)	$\eta^2 = .008$
Hepcidin, ng/mL	10 [IQR: 6.5 - 17] range: 0 - 51	9.8 [IQR: 3 - 12] range: 0 - 18	12 [IQR: 7 - 21] range: 0 - 38	8.7 [IQR: 4.5 - 18] range: 0 - 51	ns (p = .28)	$\eta^2 = .0078$
NT-proBNP, pg/mL <sup>g</sup>	58 [IQR: 0 - 110] range: 0 - 1600	0 [IQR: 0 - 73] range: 0 - 140	68 [IQR: 0 - 150] range: 0 - 1600	0 [IQR: 0 - 90] range: 0 - 850	ns (p = .18)	$\eta^2 = .021$
D-dimer, µg/L	320 [IQR: 210 - 530] range: 0 - 4000	520 [IQR: 250 - 970] range: 0 - 4000	320 [IQR: 210 - 590] range: 170 - 1800	290 [IQR: 210 - 350] range: 0 - 1100	ns (p = .13)	$\eta^2 = .029$
CRP, mg/L <sup>h</sup>	0.14 [IQR: 0.07 - 0.3] range: 0 - 7.4	0.11 [IQR: 0.06 - 0.28] range: 0 - 0.53	0.14 [IQR: 0.07 - 0.3] range: 0 - 2.4	0.16 [IQR: 0.035 - 0.31] range: 0 - 7.4	ns (p = .80)	$\eta^2 = .022$
HbA1c, % <sup>i</sup>	5.7 [IQR: 5.5 - 6] range: 4.9 - 8.8	5.5 [IQR: 5.4 - 5.8] range: 4.9 - 7.5	5.8 [IQR: 5.6 - 6.1] range: 4.9 - 7.8	5.7 [IQR: 5.3 - 5.9] range: 5 - 8.8	ns (p = .11)	$\eta^2 = .035$

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic

<sup>b</sup>Females: hemoglobin (Hb) < 120 g/dL, males: Hb < 140 g/dL

<sup>c</sup>Blood hemoglobin

<sup>d</sup>Serum ferritin

<sup>e</sup>Serum transferrin saturation

<sup>f</sup>Serum soluble transferrin receptor

<sup>g</sup>Serum N-terminal pro-brain natriuretic peptide

<sup>h</sup>Serum C-reactive protein

<sup>i</sup>Blood glycated hemoglobin

*Supplementary Table S6: Significant and near-significant ( $p < .1$ ) factors differentiating between female and male participants and illness perception in the genders (Brief Illness Perception Questionnaire, BIPQ). Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are shown as percentages and counts within the strata.*

Variable	Females	Males	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	26	48		
Smoking history <sup>b</sup>	23% (n = 6)	46% (n = 22)	ns (p = .09)	V = .22
Weight class <sup>c</sup>	normal: 58% (n = 15) overweight: 31% (n = 8) obesity: 12% (n = 3)	normal: 27% (n = 13) overweight: 50% (n = 24) obesity: 23% (n = 11)	p = .03	V = .3
Cardiovascular comorbidity	27% (n = 7)	52% (n = 25)	ns (p = .07)	V = .24
WHO COVID-19 severity <sup>d</sup>	3 [IQR: 2 - 3.8] range: 2 - 7	4 [IQR: 3 - 5] range: 2 - 7	p = .002	r = .37
COVID-19 severity	A: 38% (n = 10) HM: 46% (n = 12) HS: 15% (n = 4)	A: 10% (n = 5) HM: 58% (n = 28) HS: 31% (n = 15)	p = .01	V = .34
PSS present <sup>e</sup>	88% (n = 23)	62% (n = 30)	p = .04	V = .27
Number of PSS <sup>e</sup>	2 [IQR: 1 - 4] range: 0 - 6	1 [IQR: 0 - 2] range: 0 - 6	p = .02	r = .27
Dyspnea (mMRC $\geq$ 1) <sup>f</sup>	35% (n = 9)	15% (n = 7)	ns (p = .09)	V = .23
CT abnormality (CT score $\geq$ 1)	19% (n = 5)	73% (n = 35)	p < .001	V = .51
CT severity score <sup>g</sup>	0 [IQR: 0 - 0] range: 0 - 10	2 [IQR: 0 - 5] range: 0 - 14	p < .001	r = .49
Hb, g/dL <sup>h</sup>	140 [IQR: 130 - 150] range: 120 - 160	160 [IQR: 150 - 160] range: 110 - 180	p < .001	r = .51
FT, $\mu$ g/L <sup>i</sup>	150 [IQR: 59 - 180] range: 13 - 310	200 [IQR: 120 - 290] range: 13 - 1300	p = .02	r = .26
Total IP score (BIPQ sum)	24 [IQR: 18 - 38] range: 2 - 59	22 [IQR: 15 - 31] range: 0 - 50	ns (p = .26)	r = .13
Emotion/concern/consequences (BIPQ 1/2/5/6/8)	6 [IQR: 4 - 25] range: 0 - 43	9.5 [IQR: 3.8 - 18] range: 0 - 47	ns (p > .99)	r = .0013
Lacking control/coherence (BIPQ 3/4/7)	12 [IQR: 5.2 - 18] range: 0 - 30	10 [IQR: 4 - 15] range: 0 - 30	ns (p = .26)	r = .13
Consequences (BIPQ Q1)	1 [IQR: 0.25 - 4.5] range: 0 - 9	1 [IQR: 0 - 3.2] range: 0 - 10	ns (p = .68)	r = .049

<b>Variable</b>	<b>Females</b>	<b>Males</b>	<b>Significance<sup>a</sup></b>	<b>Effect size<sup>a</sup></b>
Timeline (BIPQ Q2)	1 [IQR: 0.25 - 5] range: 0 - 8	1.5 [IQR: 0 - 4] range: 0 - 10	ns (p = .47)	r = .085
Lacking personal control (BIPQ Q3)	5 [IQR: 2 - 8.8] range: 0 - 10	3.5 [IQR: 1 - 5.5] range: 0 - 10	ns (p = .26)	r = .13
Lacking treatment control (BIPQ Q4)	2 [IQR: 0 - 8.8] range: 0 - 10	2 [IQR: 1 - 5] range: 0 - 10	ns (p = .77)	r = .035
Identity (BIPQ Q5)	1 [IQR: 1 - 5] range: 0 - 9	2 [IQR: 0 - 3] range: 0 - 10	ns (p = .72)	r = .042
Concern (BIPQ Q6)	1 [IQR: 0 - 5] range: 0 - 9	2 [IQR: 0 - 3] range: 0 - 9	ns (p = .66)	r = .053
Lacking coherence (BIPQ Q7)	5 [IQR: 0.25 - 5] range: 0 - 10	2 [IQR: 0 - 5] range: 0 - 10	ns (p = .19)	r = .15
Emotional representation (BIPQ Q8)	1 [IQR: 0 - 4.5] range: 0 - 10	2 [IQR: 0 - 4.2] range: 0 - 10	ns (p = .78)	r = .034

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Mann-Whitney test with r effect size statistic

<sup>b</sup>Former or active smoker

<sup>c</sup>Overweight: body mass index 25 - 30 kg/m<sup>2</sup>, obesity: > 30 kg/m<sup>2</sup>

<sup>d</sup>WHO ordinal scale for clinical improvement

<sup>e</sup>PSS: persistent somatic symptoms

<sup>f</sup>mMRC: Modified Medical Research Council dyspnea score

<sup>g</sup>Chest computer tomography severity score

<sup>h</sup>Blood hemoglobin

<sup>i</sup>Serum ferritin

**Supplementary Table S7: Significant and near-significant ( $p < .1$ ) factors differentiating between participants with and without respiratory comorbidity. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are shown as percentages and counts within the strata.**

Variable	No respiratory comorbidity	Respiratory comorbidity	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	56	18		
Comorbidity present	66% (n = 37)	100% (n = 18)	$p = .01$	$V = .33$
Number of comorbidities	1.5 [IQR: 0 - 3.2] range: 0 - 7	3 [IQR: 3 - 4.8] range: 1 - 9	$p = .002$	$r = .37$
Gastrointestinal comorbidity	5.4% (n = 3)	39% (n = 7)	$p = .001$	$V = .42$
Fatigue score (likert CFS) <sup>b</sup>	12 [IQR: 11 - 15] range: 1 - 32	16 [IQR: 13 - 22] range: 4 - 25	$p = .02$	$r = .27$
Fatigue (bimodal CFS $\geq 4$ ) <sup>b</sup>	34% (n = 19)	61% (n = 11)	ns ( $p = .08$ )	$V = .24$
LFT abnormality <sup>c</sup>	25% (n = 14)	56% (n = 10)	$p = .03$	$V = .28$
Total IP score (BIPQ sum) <sup>d</sup>	20 [IQR: 12 - 30] range: 0 - 59	32 [IQR: 22 - 46] range: 4 - 53	$p = .005$	$r = .33$
Emotion/concern/consequences (BIPQ 1/2/5/6/8) <sup>d</sup>	6 [IQR: 2.8 - 15] range: 0 - 43	14 [IQR: 10 - 34] range: 0 - 47	$p = .03$	$r = .26$
Consequences (BIPQ Q1) <sup>d</sup>	1 [IQR: 0 - 3] range: 0 - 9	2 [IQR: 1 - 6.5] range: 0 - 10	$p = .03$	$r = .25$
Timeline (BIPQ Q2) <sup>d</sup>	1 [IQR: 0 - 3.2] range: 0 - 10	4.5 [IQR: 1.2 - 8] range: 0 - 10	$p = .006$	$r = .32$
Emotional representation (BIPQ Q8) <sup>d</sup>	1 [IQR: 0 - 3] range: 0 - 10	4 [IQR: 2 - 7] range: 0 - 10	ns ( $p = .05$ )	$r = .23$

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Mann-Whitney test with r effect size statistic

<sup>b</sup>CFS: Chalder's fatigue score

<sup>c</sup>Lung function testing abnormality: FVC < 80% or FEV1 < 80% or TLC < 80% or DLCO < 80% predicted or FEV1:FVC < 70% predicted reference value

<sup>d</sup>BIPQ: brief illness perception questionnaire

*Supplementary Table S8: Significant and near-significant ( $p < .1$ ) differences between the illness perception clusters. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are shown as percentages and counts within the strata.*

Variable	Cluster #1	Cluster #2	Cluster #3	Significance <sup>a</sup>	Effect size <sup>a</sup>
n, participants	38	20	16		
Smoking history <sup>b</sup>	26% (n = 10)	55% (n = 11)	44% (n = 7)	ns (p = .09)	V = .26
Metabolic comorbidity	26% (n = 10)	45% (n = 9)	56% (n = 9)	ns (p = .09)	V = .26
Respiratory comorbidity	16% (n = 6)	25% (n = 5)	44% (n = 7)	ns (p = .09)	V = .25
WHO COVID-19 severity <sup>c</sup>	4 [IQR: 3 - 5] range: 2 - 7	3 [IQR: 2.8 - 4] range: 2 - 6	4 [IQR: 3 - 6] range: 2 - 7	p = .048	$\eta^2 = .058$
PSS present <sup>d</sup>	63% (n = 24)	65% (n = 13)	100% (n = 16)	p = .02	V = .33
Number of PSS <sup>d</sup>	1 [IQR: 0 - 2] range: 0 - 5	1 [IQR: 0 - 2.2] range: 0 - 6	4 [IQR: 2.8 - 5] range: 2 - 6	p < .001	$\eta^2 = .31$
Reduced performance (ECOG $\geq 1$ ) <sup>e</sup>	24% (n = 9)	20% (n = 4)	81% (n = 13)	p < .001	V = .51
Dyspnea (mMRC $\geq 1$ ) <sup>f</sup>	11% (n = 4)	20% (n = 4)	50% (n = 8)	p = .006	V = .37
Sleep problems	29% (n = 11)	15% (n = 3)	62% (n = 10)	p = .008	V = .36
Gastrointestinal symptoms	0% (n = 0)	0% (n = 0)	12% (n = 2)	p = .02	V = .32
Hair loss	0% (n = 0)	0% (n = 0)	38% (n = 6)	p < .001	V = .57
Fatigue score (likert CFS) <sup>g</sup>	11 [IQR: 11 - 13] range: 1 - 29	12 [IQR: 11 - 15] range: 4 - 24	22 [IQR: 16 - 25] range: 12 - 32	p < .001	$\eta^2 = .34$
Fatigue (bimodal CFS $\geq 4$ ) <sup>g</sup>	24% (n = 9)	30% (n = 6)	94% (n = 15)	p < .001	V = .57
LFT abnormality <sup>h</sup>	21% (n = 8)	45% (n = 9)	44% (n = 7)	ns (p = .10)	V = .25
CT severity score	1 [IQR: 0 - 5] range: 0 - 14	0 [IQR: 0 - 1.2] range: 0 - 5	4.5 [IQR: 0 - 10] range: 0 - 13	p = .02	$\eta^2 = .09$
Rehabilitation	32% (n = 12)	10% (n = 2)	56% (n = 9)	p = .01	V = .35

<sup>a</sup>Categorical variables:  $\chi^2$  test with Cramer V effect size statistic, numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic

<sup>b</sup>Former or active smoker

<sup>c</sup>WHO ordinal scale for clinical improvement

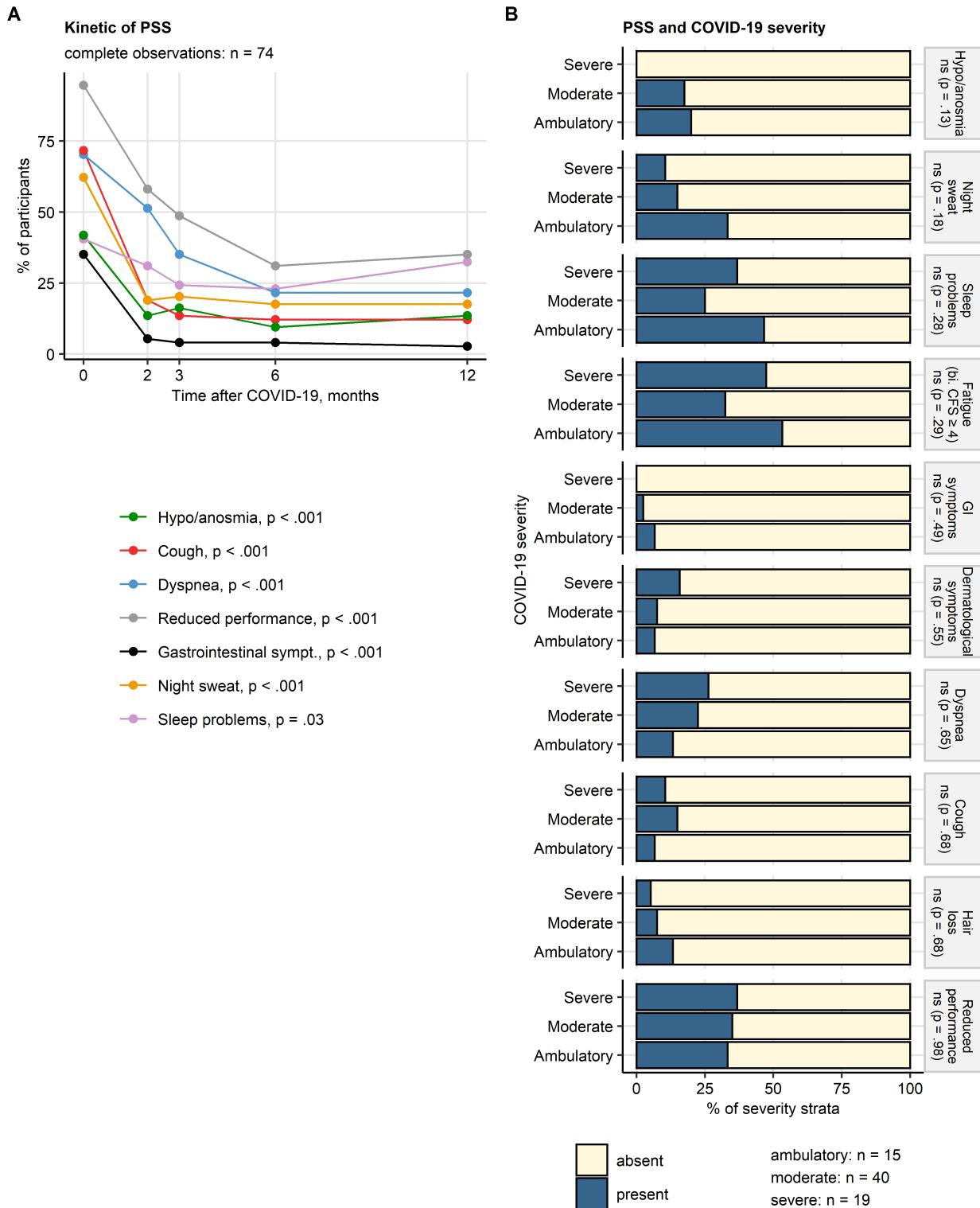
<sup>d</sup>PSS: persistent somatic symptoms

<sup>e</sup>ECOG: Eastern Cooperative Oncology Group performance score

<sup>f</sup>mMRC: Modified Medical Research Council dyspnea score

Variable	Cluster #1	Cluster #2	Cluster #3	Significance <sup>a</sup>	Effect size <sup>a</sup>
<sup>g</sup> Chalder's Fatigue Score					
<sup>h</sup> Lung function testing abnormality: FVC < 80% or FEV1 < 80% or TLC < 80% or DLCO < 80% predicted or FEV1:FVC < 70% predicted reference value					

## Supplementary Figures

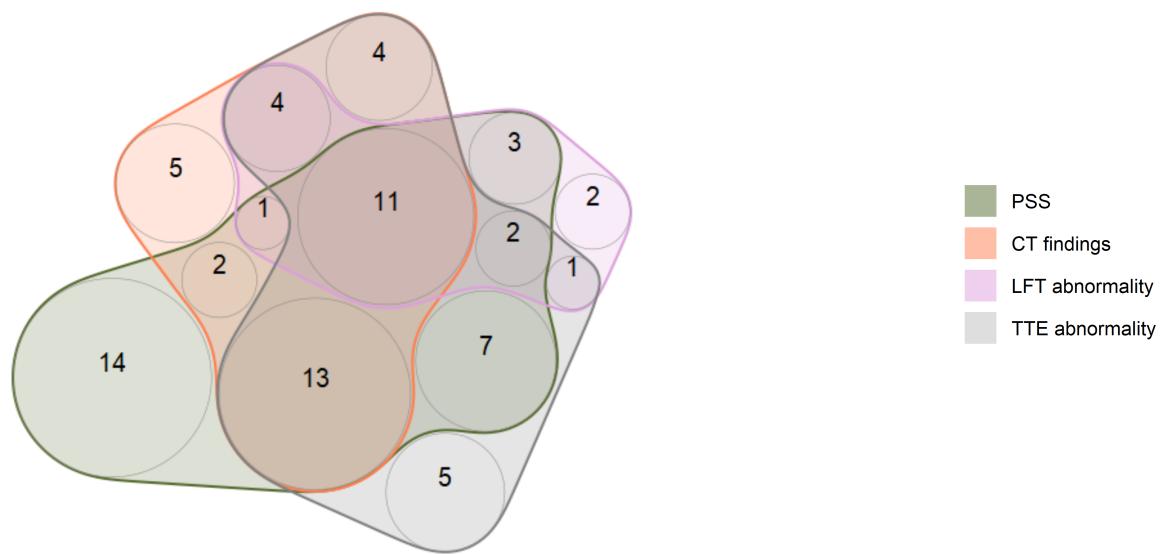


**Supplementary Figure S1. Longitudinal changes of persistent somatic symptom rates, frequency of persistent somatic symptoms in the COVID-19 severity strata at the one-year follow-up.**

(A) Frequencies of persistent somatic symptoms (PSS) in acute COVID-19 (time 0) and at the consecutive follow-ups after COVID-19 diagnosis expressed as percentages of the analysis collective. Only PSS recorded at every follow-up were analyzed. Statistical significance of changes in frequencies in time was investigated by Cochran Q test. Number of complete observations is indicated in the plot caption. P values for specific PSS are presented in the plot legend.

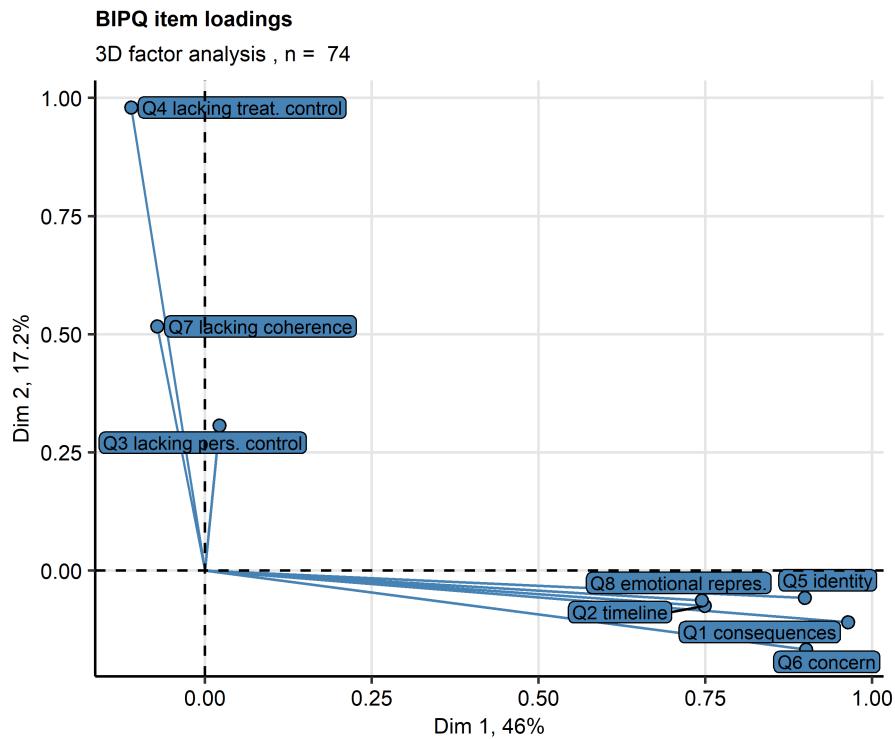
(B) Frequencies of PSS at the one-year follow up were compared between survivors of ambulatory, moderate and severe COVID-19 by  $\chi^2$  test. Percentages of individuals with and without the specific PSS within the severity strata are presented in a stack plot. Significance is indicated in the plot facets. Numbers of complete observations are displayed next to the plot.  
bi. CFS: bimodal Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ .

Overlap between COVID-19 sequelae at the one-year follow-up



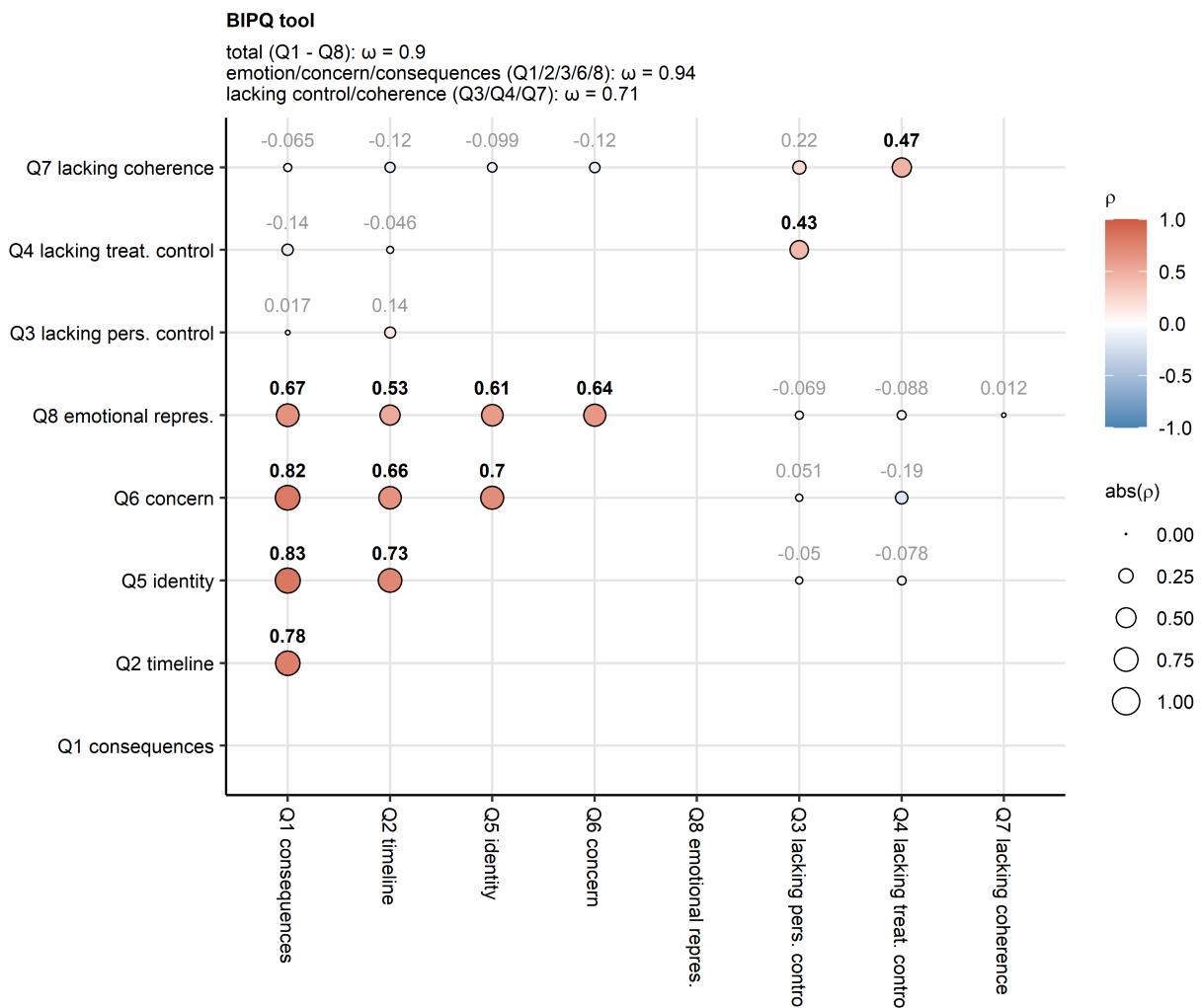
**Supplementary Figure S2. Overlap between persistent somatic symptoms and cardiopulmonary findings at the one-year follow-up.**

Numbers of the analyzed CovILD study participants (total:  $n = 74$ ) with at least one persistent somatic symptom (PSS), any abnormality in chest computed tomography (CT findings) or lung function testing (LFT abnormality) or diastolic dysfunction in trans-thoracic echocardiography (TTE abnormality) at the one-year follow-up presented in a quasi-proportional Venn plot.



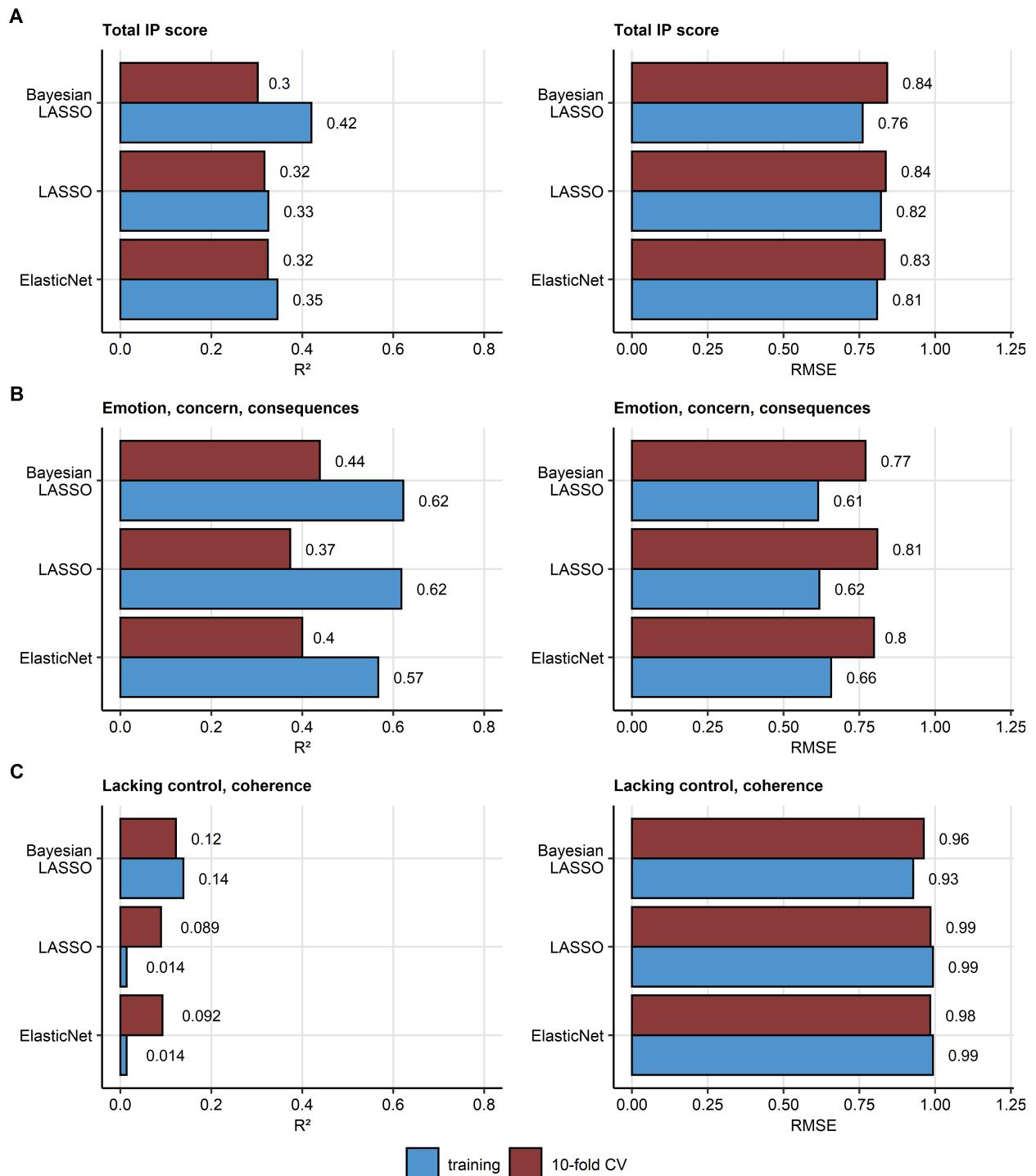
**Supplementary Figure S3. Factor analysis of the BIPQ tool items.**

*Brief illness perception questionnaire (BIPQ) items were subjected to three-dimensional factor analysis. Loadings of the BIPQ items for the first two factors are presented as points. Variances associated with the factors are shown in the plot axes. Numbers of complete observations are indicated in the plot caption.*



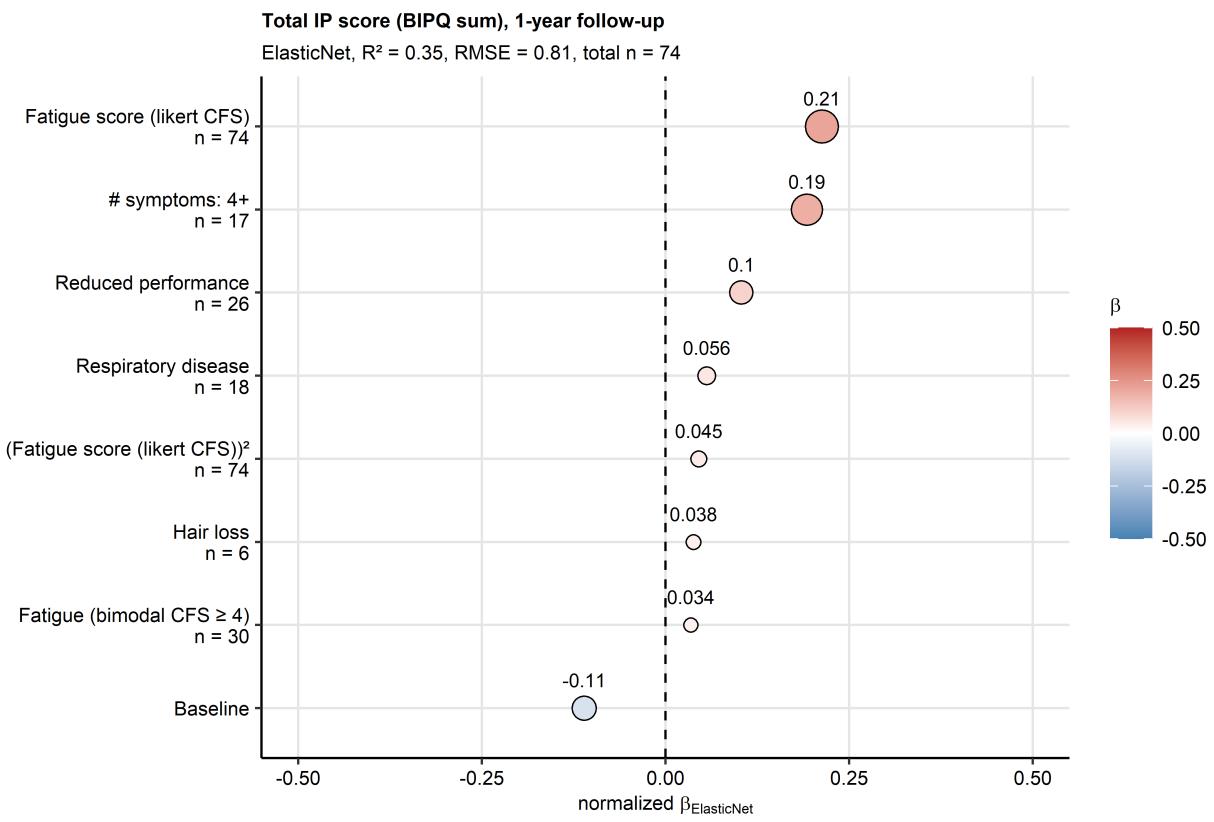
**Supplementary Figure S4. Coherence of the BIPQ tool.**

Pair-wise correlation between the Brief Illness Perception Questionnaire (BIPQ) items was assessed by Spearman test (observations:  $n = 74$ ) and the results presented in a bubble plot. Point size corresponds to absolute values of the correlation coefficient  $p$ . Point color codes for the  $p$  value. The points are labeled with  $p$  values, significant effects are highlighted in bold. Internal consistency of the BIPQ tool was measured by McDonald's  $\omega$ .  $\omega$  values for the entire BIPQ tool, the emotion/concern/consequences (Q1/2/3/6/8) and the lacking control/coherence (Q3/Q4/Q7) factors are indicated in the plot caption.



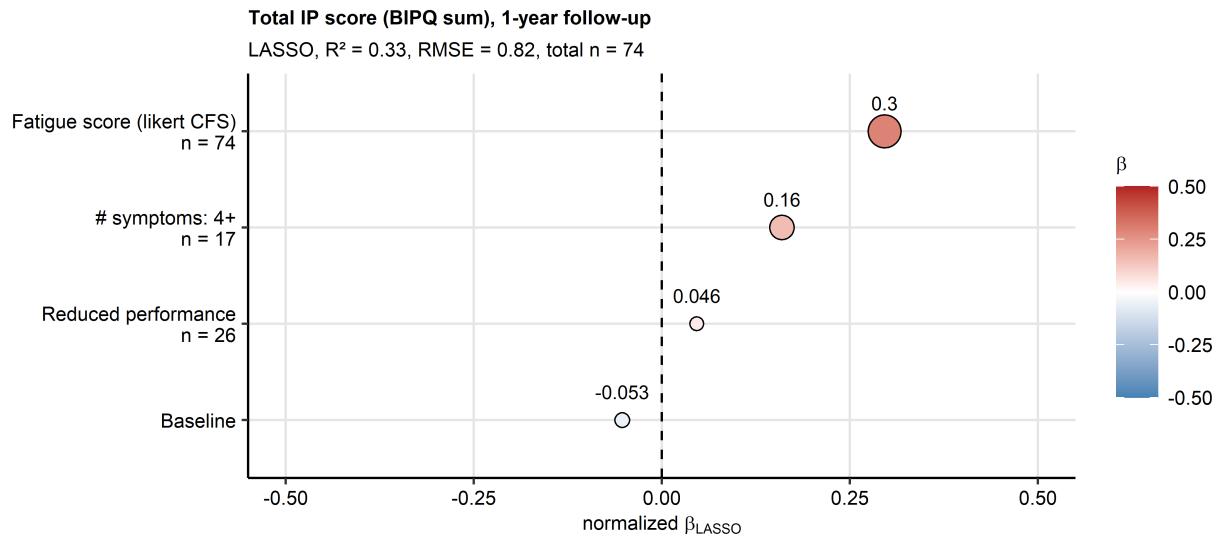
**Supplementary Figure S5. Performance and reproducibility of regularized multi-parameter models.**

*The total illness perception (IP) score (A), the emotion/concern/consequences (B) and the lacking control/coherence (C) IP component scores at the one-year follow-up were modeled by the Elastic Net, LASSO and Bayesian LASSO. Fraction of the modeling response variance explained by the model in the training data set and 10-repeats 10-fold cross-validation (CV) was estimated by the  $R^2$  statistic. Model error was expressed as root mean squared error (RMSE) in the training and CV data sets.*



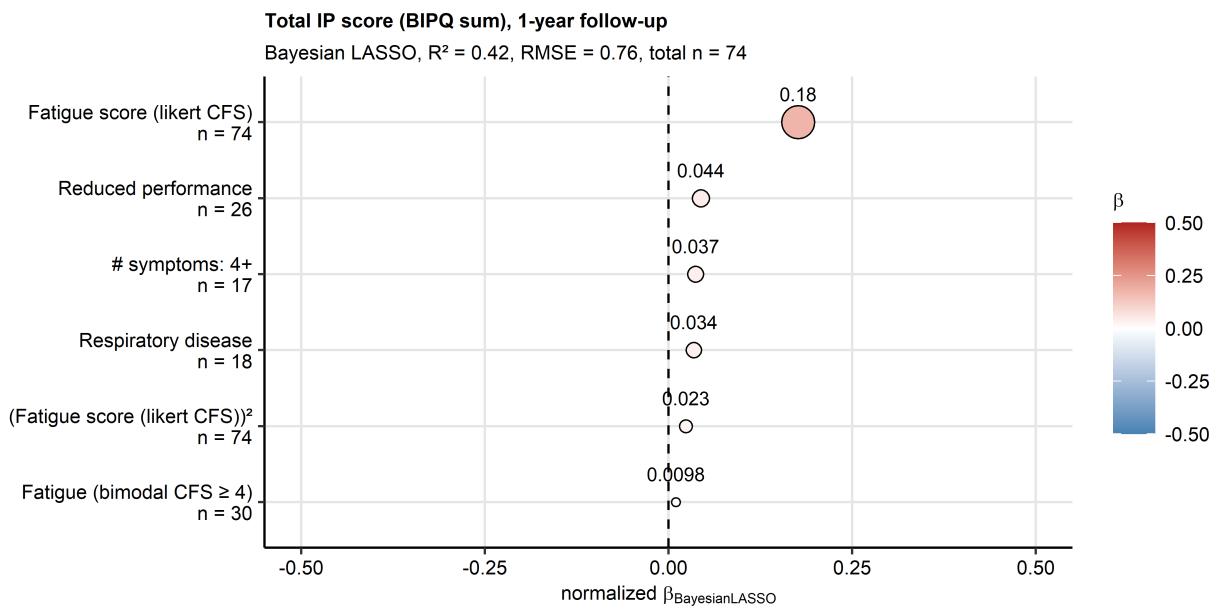
**Supplementary Figure S6. Non-zero coefficient estimates of the Elastic Net multi-parameter model of the total illness perception (IP) score at the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up.



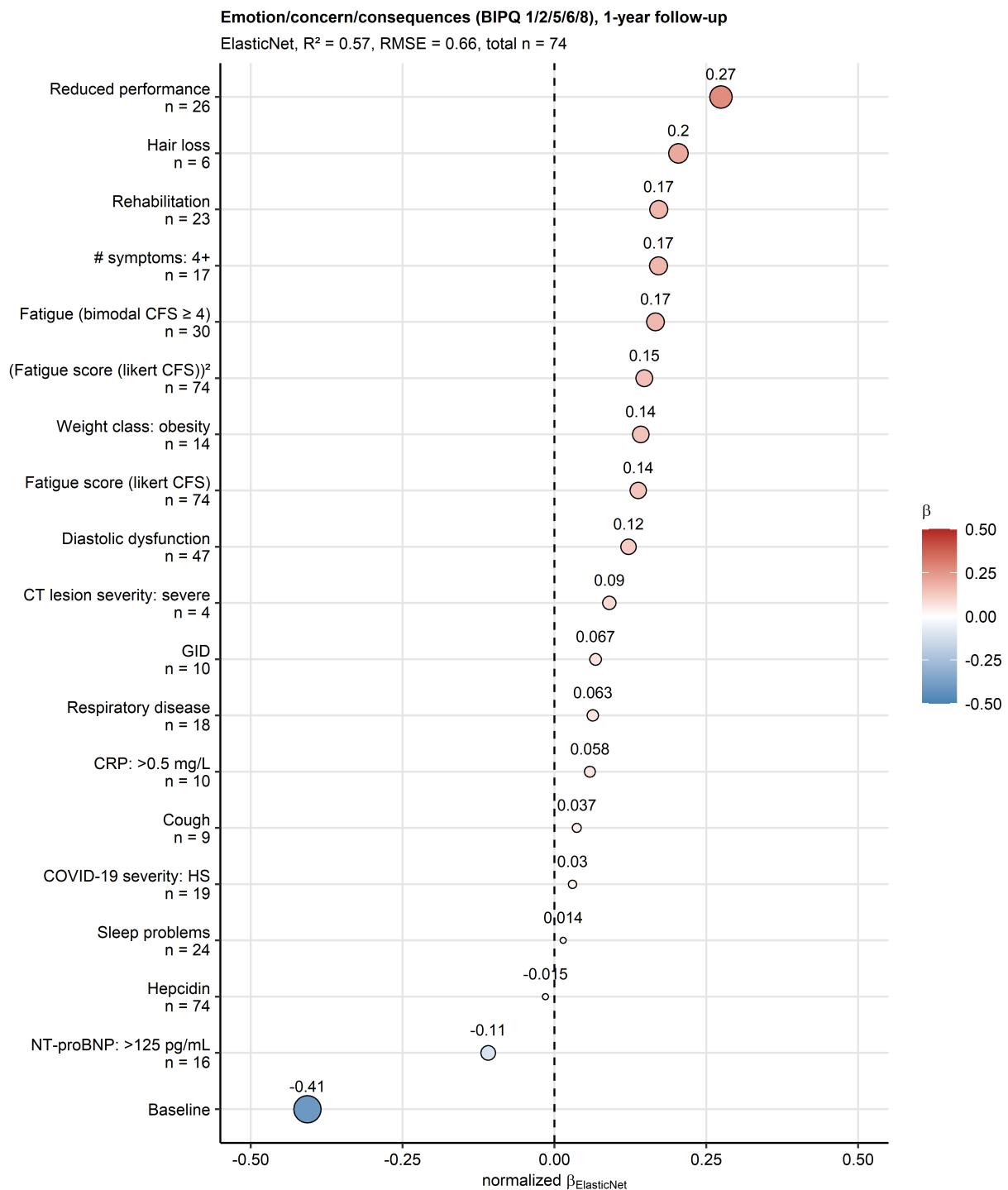
**Supplementary Figure S7. Non-zero coefficient estimates of the LASSO multi-parameter model of the total illness perception (IP) score at the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up.



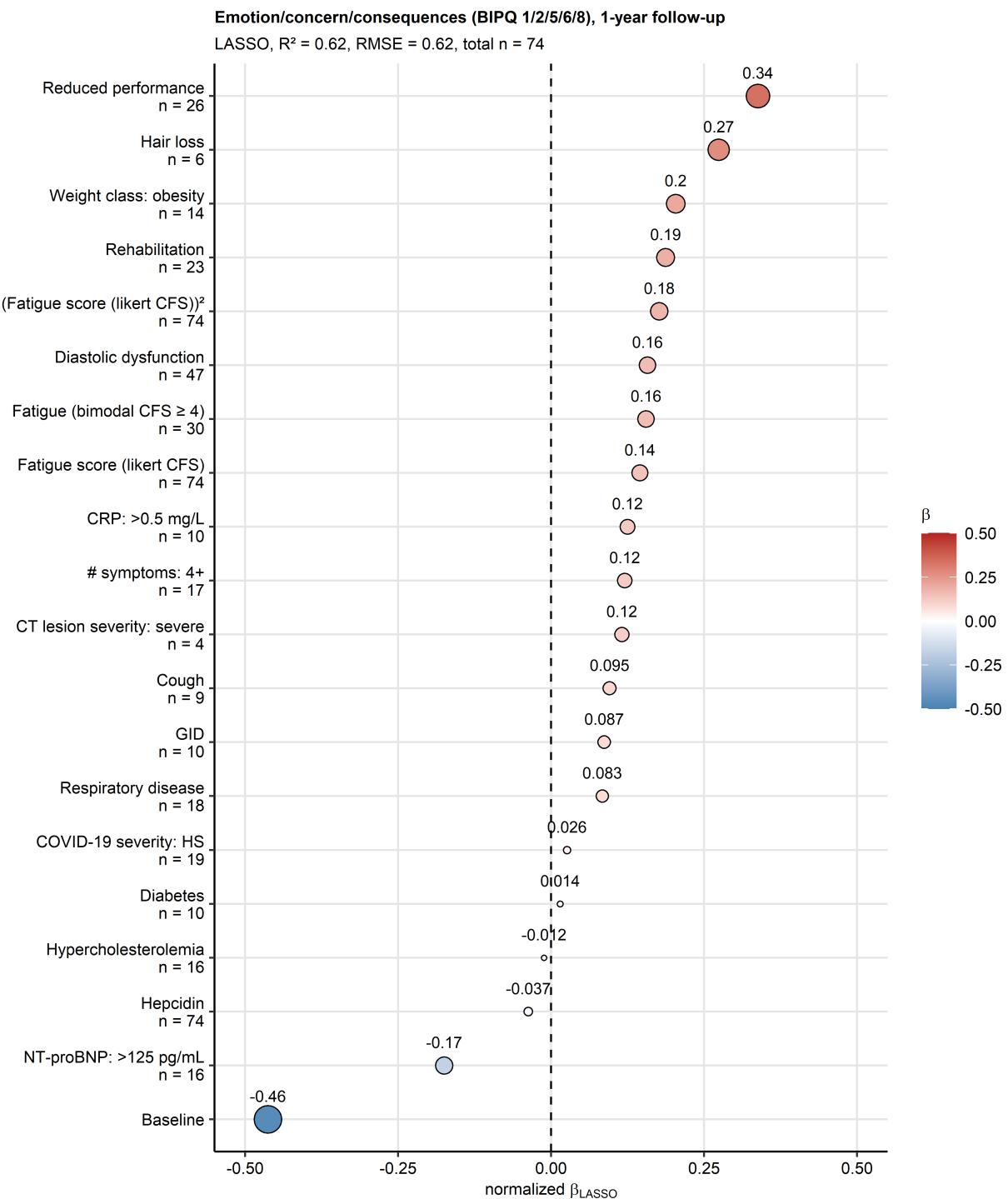
**Supplementary Figure S8. Non-zero coefficient estimates of the Bayesian LASSO multi-parameter model of the total illness perception (IP) score at the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up.



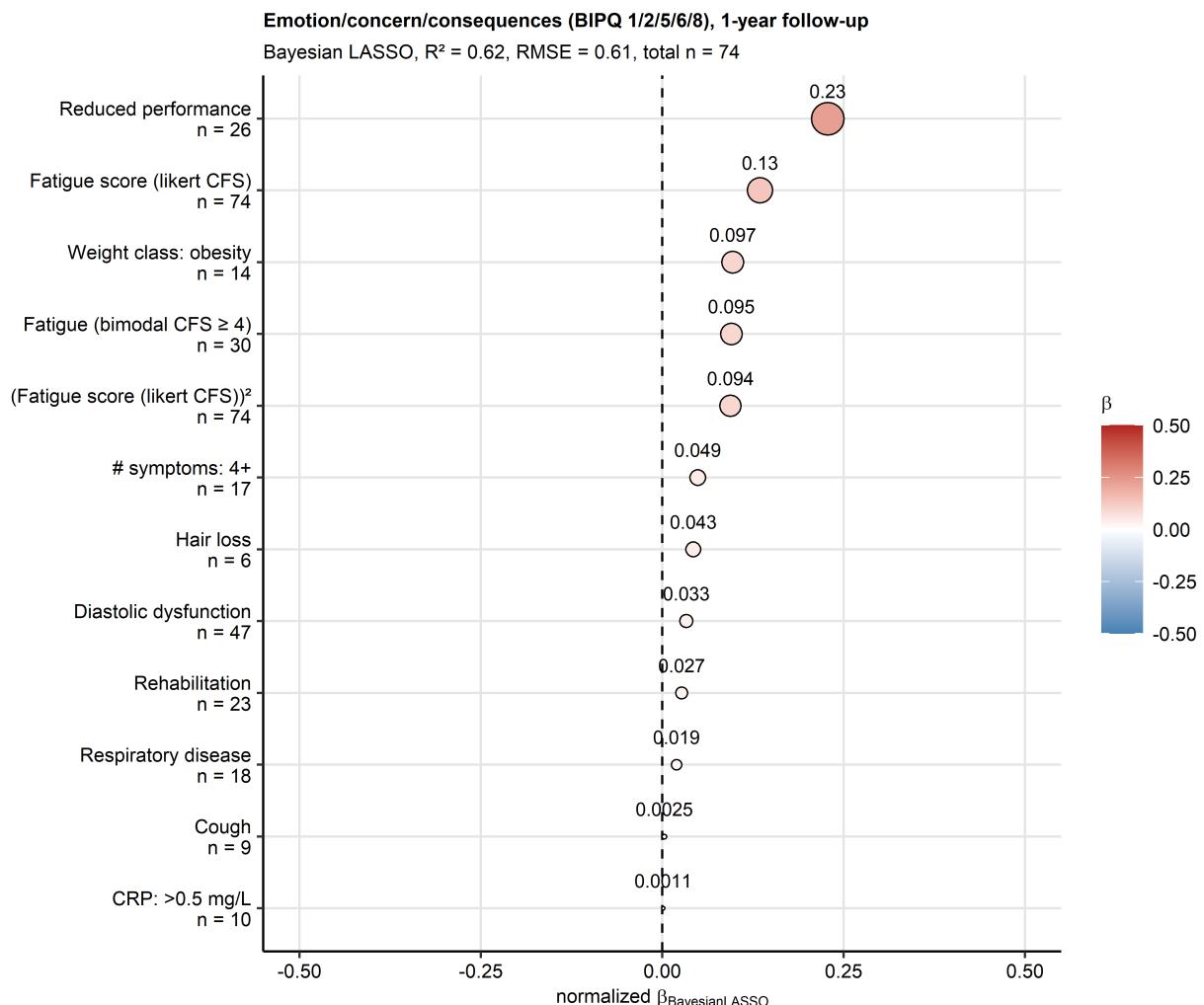
**Supplementary Figure S9. Non-zero coefficient estimates of the Elastic Net multi-parameter model of the emotion/concern/consequences illness perception component the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up; CT lesion severity: chest computed tomography abnormality severity at the one-year follow-up; GID: gastrointestinal disease; CRP: serum C-reactive protein; COVID-19 severity: HS: hospitalized/severe; NT-proBNP: N-terminal pro-brain natriuretic peptide.



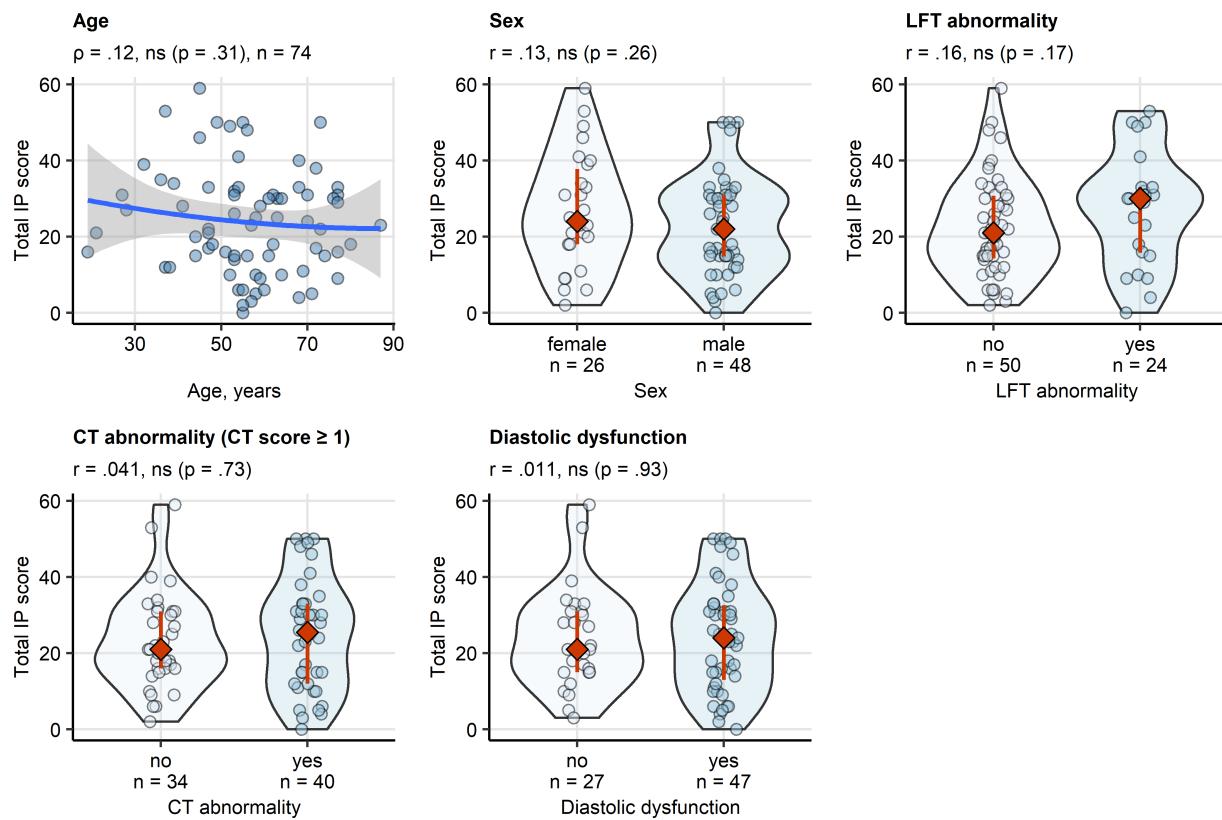
**Supplementary Figure S10. Non-zero coefficient estimates of the LASSO multi-parameter model of the emotion/concern/consequences illness perception component at the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up; CT lesion severity: chest computed tomography abnormality severity at the one-year follow-up; GID: gastrointestinal disease; CRP: serum C-reactive protein; COVID-19 severity: HS: hospitalized/severe; NT-proBNP: N-terminal pro-brain natriuretic peptide.



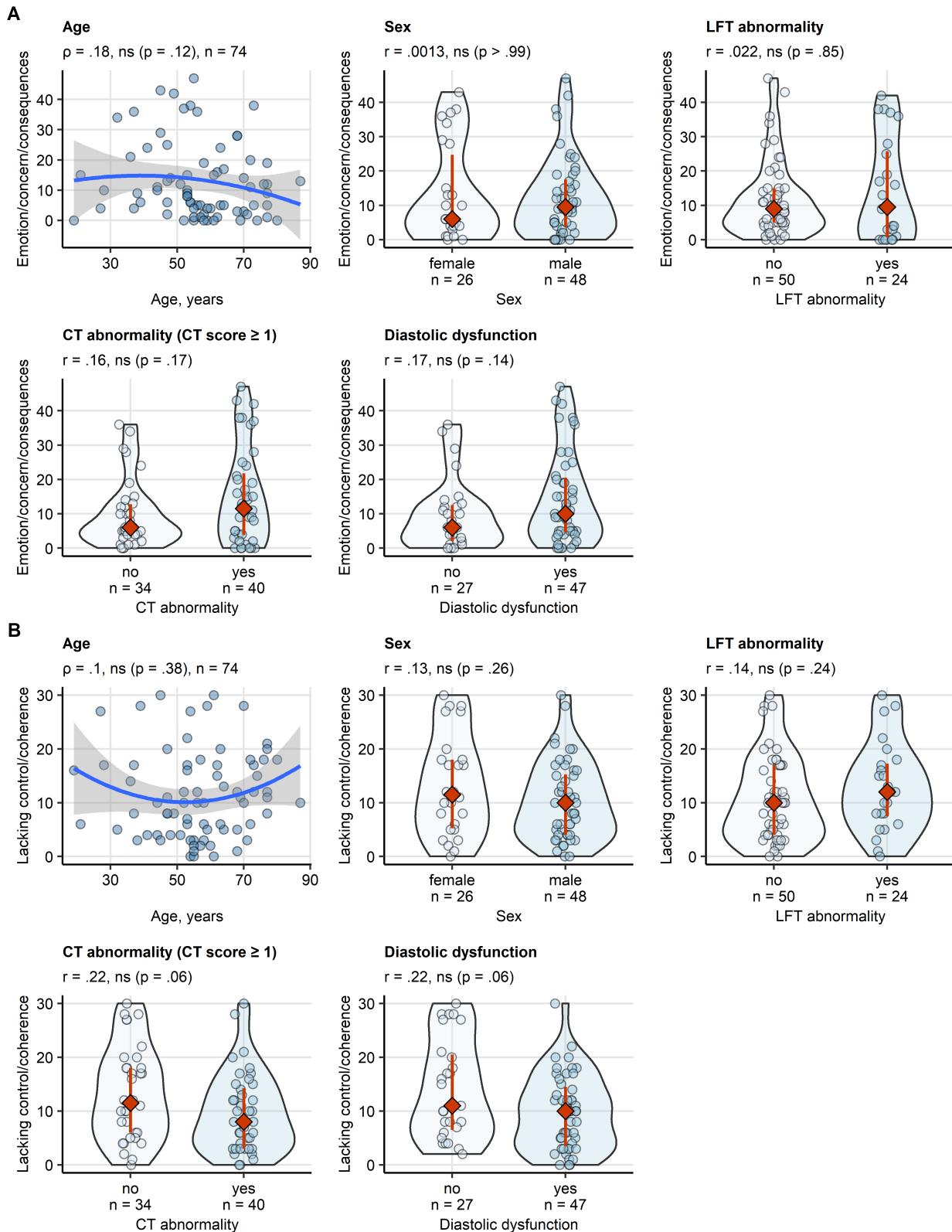
**Supplementary Figure S11. Non-zero coefficient estimates of the Bayesian LASSO multi-parameter model of the emotion/concern/consequences illness perception component at the one-year follow-up.**

Z-score normalized estimates ( $\beta$ ) of the non-zero coefficients of the model are presented in the plot as points. Point color codes  $\beta$  value, point size corresponds to the absolute value of  $\beta$ . Model's performance measures in the training data set ( $R^2$  and root mean squared error [RMSE]) and numbers of complete observations are indicated in the plot caption. The data points are labeled with their  $\beta$  values. CFS: Chalder's fatigue score; reduced performance: ECOG (Eastern Cooperative Oncology Group)  $\geq 1$ ; # PSS: number of persistent somatic symptoms at the one-year follow-up; CRP: serum C-reactive protein.



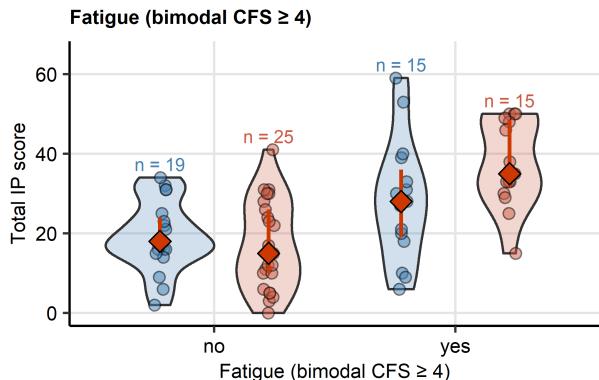
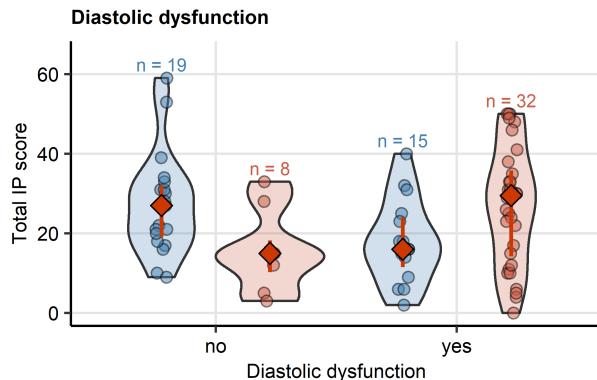
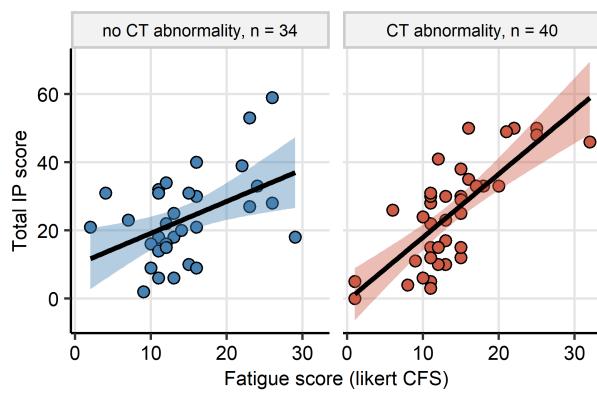
**Supplementary Figure S12. Age, sex, cardiopulmonary abnormalities and persistent somatic symptoms at the one-year follow-up and the total illness perception score.**

*Association of the total illness perception (IP) score with age at COVID-19 diagnosis, sex, and lung function (LFT) testing abnormality, chest computed tomography (CT) abnormality and heart diastolic dysfunction and presence of any persistent somatic symptoms (PSS) at the one-year follow-up. Statistical significance was assessed by Spearman's correlation (age) or Mann-Whitney test with r effect size statistic (remaining independent variables). The correlation with age is presented in a point plot; the blue line represents the fitted second-order trend and the gray ribbon depicts the 95% confidence intervals. For the remaining independent variables, the score values are presented in violin plots with single observations depicted as points, and red diamonds and whiskers representing medians with interquartile ranges. Effect size statistic and p values are indicated in the plot captions. Numbers of complete observations are displayed in the plot captions or in the plot X axes.*



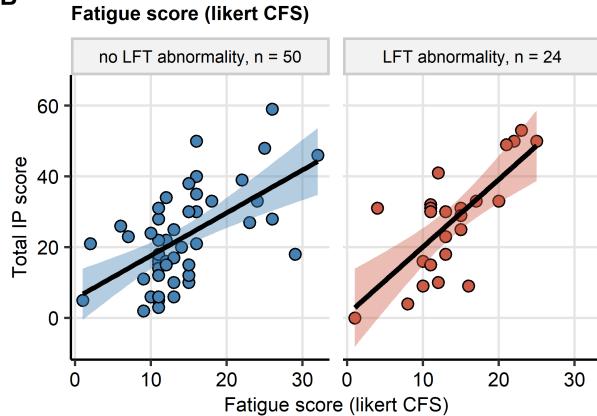
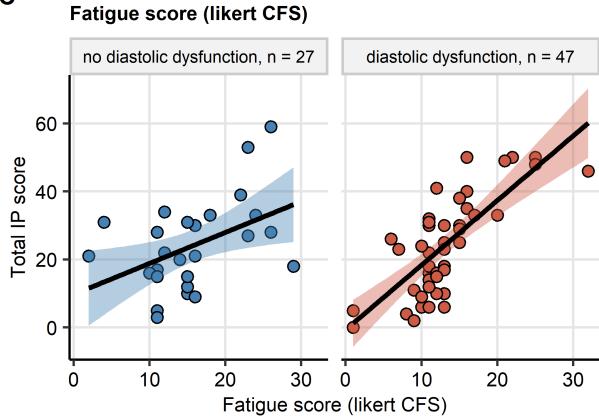
**Supplementary Figure S13. Age, sex, cardiopulmonary abnormalities and persistent somatic symptoms at the one-year follow-up and the illness perception components.**

*Association of the emotion/concern/consequences (A) and the lacking control/coherence (B) IP component scores with age at COVID-19 diagnosis, sex, and lung function (LFT) testing abnormality, chest computed tomography (CT) abnormality and heart diastolic dysfunction and presence of any persistent somatic symptoms (PSS) at the one-year follow-up. Statistical significance was assessed by Spearman's correlation (age) or Mann-Whitney test with r effect size statistic (remaining independent variables). The correlation with age is presented in a point plot; the blue line represents the fitted second-order trend and the gray ribbon depicts the 95% confidence intervals. For the remaining independent variables, the score values are presented in violin plots with single observations depicted as points, and red diamonds and whiskers representing medians with interquartile ranges. Effect size statistic and p values are indicated in the plot captions. Numbers of complete observations are displayed in the plot captions or in the plot X axes.*

**A****Fatigue score (likert CFS)**

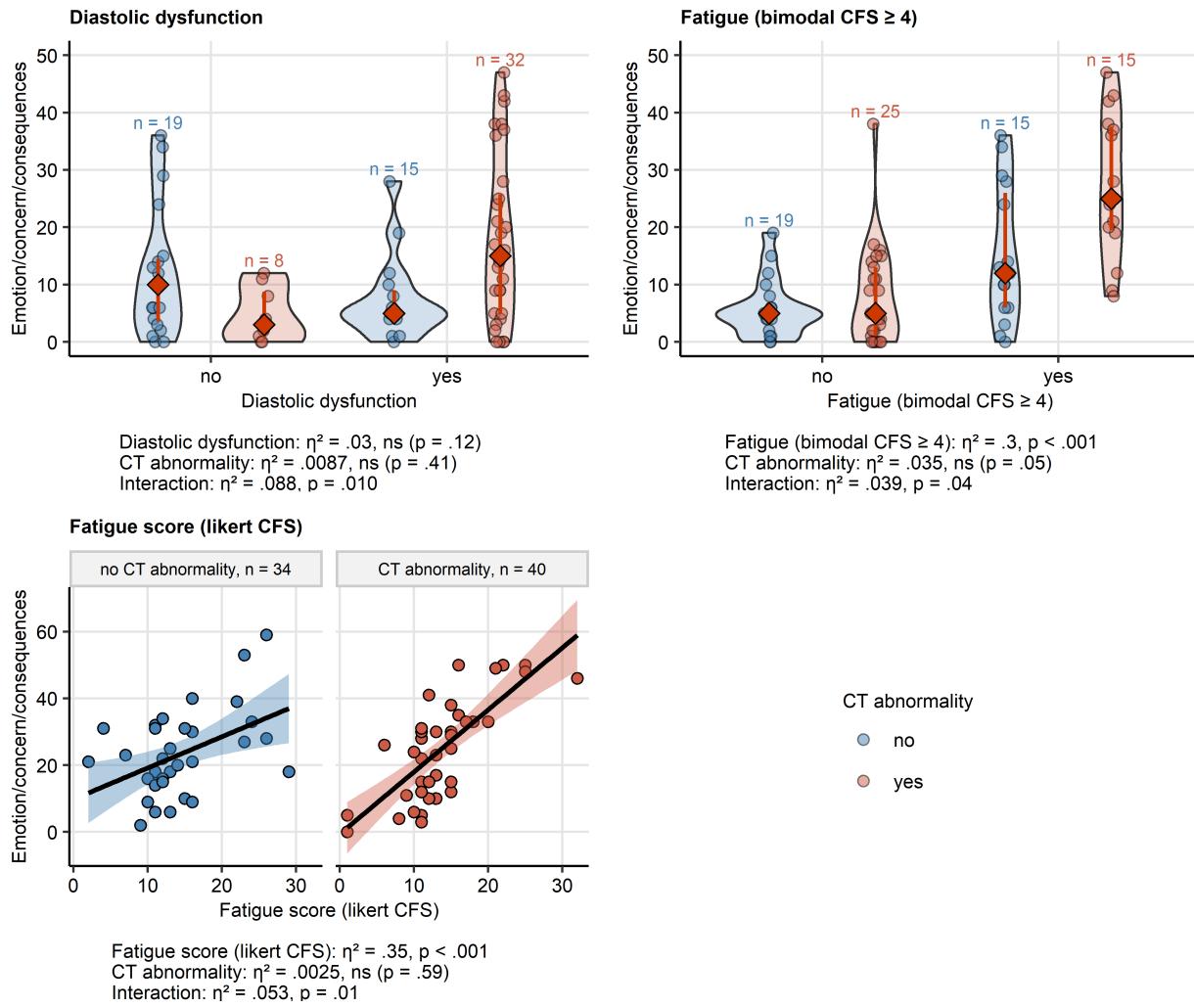
CT abnormality

- no
- yes

**B****C**

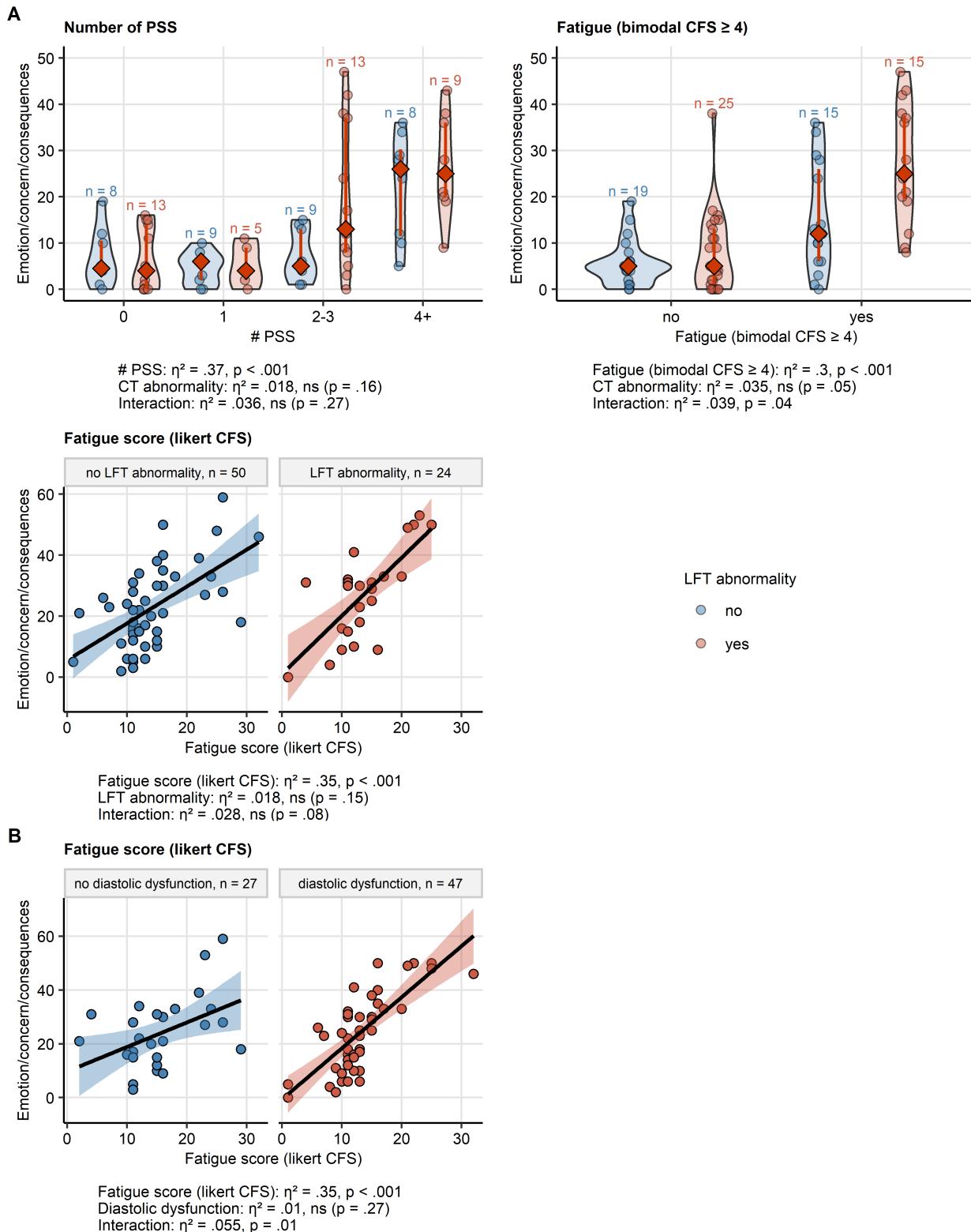
**Supplementary Figure S14. Interactions of diastolic dysfunction, fatigue and fatigue intensity with cardiopulmonary findings follow-up affecting the total illness perception at the one-year follow-up.**

*Interaction effects between the key factors affecting the total illness perception (IP) score and cardiopulmonary abnormalities (A: chest computed tomography [CT] abnormality, B: lung function testing [LFT] abnormality, C: diastolic dysfunction) at the one-year follow-up were investigated by two-way ANOVA with  $\eta^2$  effect size statistic. The key factors with significant and near-significant interaction effects ( $p < .1$ ) are presented. The associations with fatigue scoring are presented in a point plots; the black line represents the fitted linear trend and the tinted ribbon depicts the 95% confidence intervals. The total score values for the remaining explanatory variables are presented in violin plots with single observations depicted as points, and red diamonds and whiskers representing medians with interquartile ranges. Numbers of complete observations in the strata are displayed in the plot or in the plot facets. Effect size statistic and p values are shown under the plots. CFS: Chalder's fatigue score.*



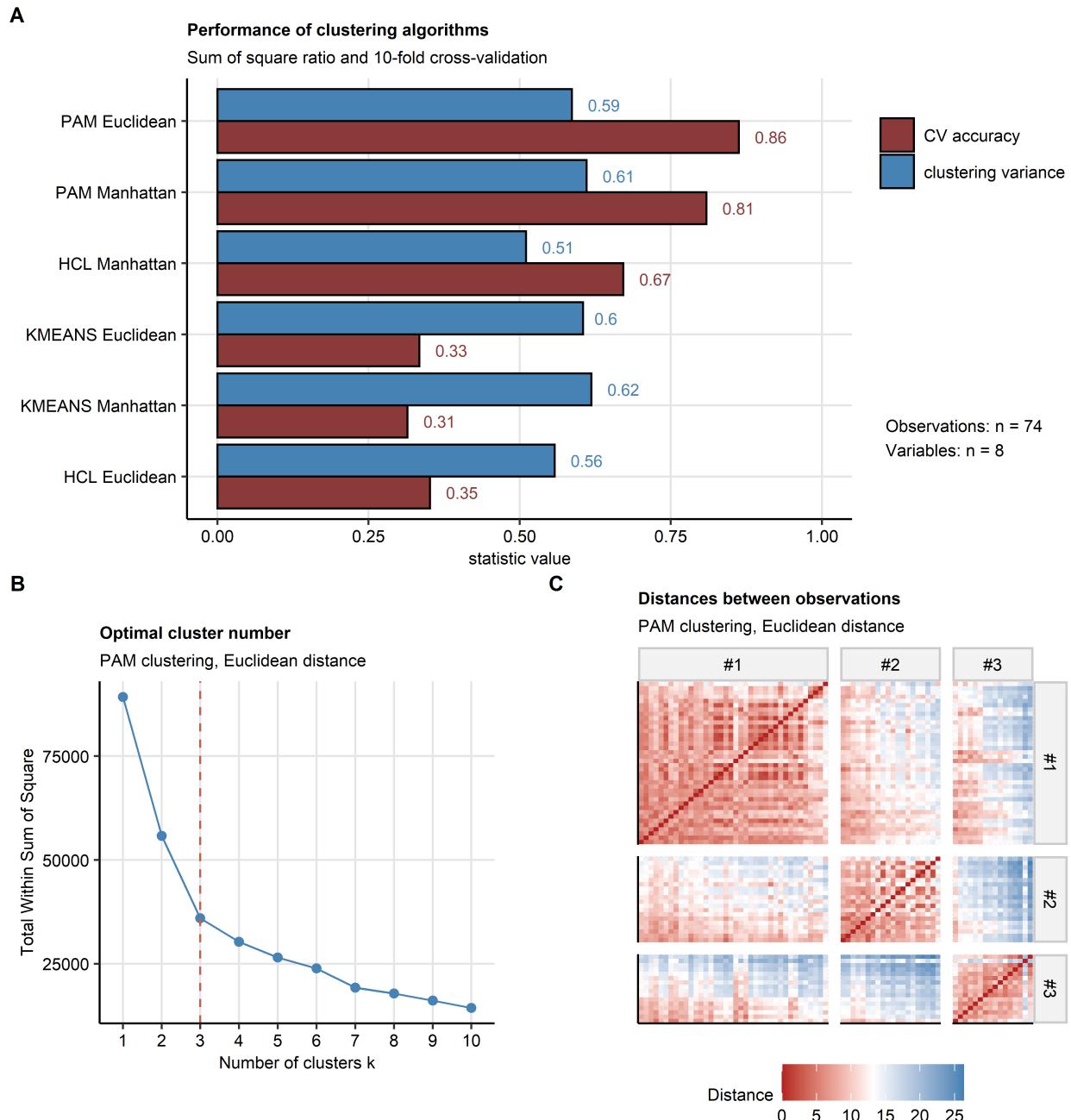
**Supplementary Figure S15. Interactions of diastolic dysfunction, fatigue and fatigue scoring with chest CT abnormality affecting the emotion/concern/consequences component of illness perception.**

*Interaction effects between the key factors affecting the emotion/concern/consequences IP component score and chest computed tomography (CT) abnormalities at the one-year follow-up were investigated by two-way ANOVA with  $\eta^2$  effect size statistic. The key factors with significant and near-significant interaction effects ( $p < .1$ ) are presented. The associations with fatigue scoring are presented in a point plots; the black line represents the fitted linear trend and the tinted ribbon depicts the 95% confidence intervals. The total score values for the remaining explanatory variables are presented in violin plots with single observations depicted as points, and red diamonds and whiskers representing medians with interquartile ranges. Numbers of complete observations in the strata are displayed in the plot or in the plot facets. Effect size statistic and p values are shown under the plots. CFS: Chalder's fatigue score.*



**Supplementary Figure S16. Interactions of persistent somatic symptom number, fatigue and fatigue scoring with LFT abnormality and diastolic dysfunction affecting the emotion/concern/consequences component of illness perception.**

*Interaction effects between the key factors affecting the emotion/concern/consequences IP component score and cardiopulmonary abnormalities (A: lung function testing [LFT] abnormality, B: diastolic dysfunction) at the one-year follow-up were investigated by two-way ANOVA with  $\eta^2$  effect size statistic. The key factors with significant and near-significant interaction effects ( $p < .1$ ) are presented. The associations with fatigue scoring are presented in a point plots; the black line represents the fitted linear trend and the tinted ribbon depicts the 95% confidence intervals. The total score values for the remaining explanatory variables are presented in violin plots with single observations depicted as points, and red diamonds and whiskers representing medians with interquartile ranges. Numbers of complete observations in the strata are displayed in the plot or in the plot facets. Effect size statistic and p values are shown under the plots. CFS: Chalder's fatigue score.*



### Supplementary Figure S17. Choice of the optimal clustering algorithm and the cluster number.

Study participants were clustered in respect to the Brief Illness Perception Questionnaire items (Q1 - Q8) using the PAM (partitioning around medoids) clustering algorithm and Euclidean distance between the observations. Numbers of complete observations and clustering variables are indicated in (A)

*(A) Comparison of clustering algorithm performance (PAM, HCL: Ward D2 hierarchical clustering, KMEANS). Explained clustering variance was defined as a ratio of the between-cluster sum of squares to the total sum of squares, cluster assignment accuracy was investigated in 10-fold cross-validation (CV). Note the superior CV accuracy of the PAM/Euclidean distance.*

*(B, C) Choice of the optimal cluster number for the hierarchical clustering/Euclidean distance algorithm. The optimal cluster number was selected based on the bend of the curve of within-cluster sum of squares (B, dashed red line: the selected cluster number) and visual inspection of the distance heat map (C).*

## References

- [1] E. Broadbent, K.J. Petrie, J. Main, J. Weinman, The Brief Illness Perception Questionnaire, *Journal of Psychosomatic Research*. 60 (2006) 631–637.  
<https://doi.org/10.1016/j.jpsychores.2005.10.020>.
- [2] T. Sonnweber, S. Sahanic, A. Pizzini, A. Luger, C. Schwabl, B. Sonnweber, K. Kurz, S. Koppelstätter, D. Haschka, V. Petzer, A. Boehm, M. Aichner, P. Tymoszuk, D. Lener, M. Theurl, A. Lorsbach-Köhler, A. Tancevski, A. Schapfl, M. Schaber, R. Hilbe, M. Nairz, B. Puchner, D. Hüttenberger, C. Tschartschenthaler, M. Aßhoff, A. Peer, F. Hartig, R. Bellmann, M. Joannidis, C. Gollmann-Tepenköylü, J. Hofeld, G. Feuchtner, A. Egger, G. Hoermann, A. Schroll, G. Fritzsche, S. Wildner, R. Bellmann-Weiler, R. Kirchmair, R. Helbok, H. Prosch, D. Rieder, Z. Trajanoski, F. Kronenberg, E. Wöll, G. Weiss, G. Widmann, J. Löffler-Ragg, I. Tancevski, Cardiopulmonary recovery after COVID-19: An observational prospective multicentre trial, *European Respiratory Journal*. 57 (2021).  
<https://doi.org/10.1183/13993003.03481-2020>.
- [3] T. Sonnweber, P. Tymoszuk, S. Sahanic, A. Boehm, A. Pizzini, A. Luger, C. Schwabl, M. Nairz, P. Grubwieser, K. Kurz, S. Koppelstätter, M. Aichner, B. Puchner, A. Egger, G. Hoermann, E. Wöll, G. Weiss, G. Widmann, I. Tancevski, J. Löffler-Ragg, Investigating phenotypes of pulmonary COVID-19 recovery: A longitudinal observational prospective multicenter trial, *eLife*. 11 (2022). <https://doi.org/10.7554/ELIFE.72500>.
- [4] A.K. Luger, T. Sonnweber, L. Gruber, C. Schwabl, K. Cima, P. Tymoszuk, A.K. Gerstner, A. Pizzini, S. Sahanic, A. Boehm, M. Coen, C.J. Strolz, E. Wöll, G. Weiss, R. Kirchmair, G.M. Feuchtner, H. Prosch, I. Tancevski, J. Löffler-Ragg, G. Widmann, Chest CT of Lung Injury 1 Year after COVID-19 Pneumonia: The CovILD Study, *Radiology*. 304 (2022) 462–470.  
<https://doi.org/10.1148/radiol.211670>.
- [5] D.M. Hansell, A.A. Bankier, H. MacMahon, T.C. McLoud, N.L. Müller, J. Remy, Fleischner Society: Glossary of terms for thoracic imaging, 246 (2008) 697–722.  
<https://doi.org/10.1148/radiol.2462070712>.
- [6] T. Chalder, G. Berelowitz, T. Pawlikowska, L. Watts, S. Wessely, D. Wright, E.P. Wallace, Development of a fatigue scale, *Journal of Psychosomatic Research*. 37 (1993) 147–153. [https://doi.org/10.1016/0022-3999\(93\)90081-P](https://doi.org/10.1016/0022-3999(93)90081-P).
- [7] R.K. Morriss, A.J. Wearden, R. Mullis, Exploring the validity of the chalder fatigue scale in chronic fatigue syndrome, *Journal of Psychosomatic Research*. 45 (1998) 411–417.  
[https://doi.org/10.1016/S0022-3999\(98\)00022-1](https://doi.org/10.1016/S0022-3999(98)00022-1).

- [8] R.O. Crapo, R. Casaburi, A.L. Coates, P.L. Enright, N.R. MacIntyre, R.T. McKay, D. Johnson, J.S. Wanger, R.J. Zeballos, V. Bittner, C. Mottram, ATS statement: Guidelines for the six-minute walk test, 166 (2002) 111–117. <https://doi.org/10.1164/ajrccm.166.1.at1102>.
- [9] M.S. Bartlett, THE STATISTICAL CONCEPTION OF MENTAL FACTORS, British Journal of Psychology. General Section. 28 (1937) 97–104. <https://doi.org/10.1111/j.2044-8295.1937.tb00863.x>.
- [10] R.P. McDonald, Test theory: A unified treatment, 1st Editio, Psychology Press, New Yor, 1999. <https://doi.org/10.4324/9781410601087>.
- [11] H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the Tidyverse, Journal of Open Source Software. 4 (2019) 1686. <https://doi.org/10.21105/joss.01686>.
- [12] L. Henry, Hadley. Wickham, rlang: Functions for Base Types and Core R and 'Tidyverse' Features, (2022). <https://cran.r-project.org/web/packages/rlang/index.html>.
- [13] M. Gagolewski, B. Tartanus, Package 'stringi', (2021).  
<https://cran.r-project.org/web/packages/stringi/index.html>  
<http://cran.ism.ac.jp/web/packages/stringi/stringi.pdf>.
- [14] A. Kassambara, rstatix: Pipe-Friendly Framework for Basic Statistical Tests, (2021). <https://cran.r-project.org/package=rstatix>.
- [15] W. Revelle, Package 'psych' - Procedures for Psychological, Psychometric and Personality Research, R Package. (2015) 1–358.  
<https://cran.r-project.org/web/packages/psych/index.html> <http://personality-project.org/r/psych-manual.pdf>.
- [16] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software. 33 (2010) 1–22. <https://doi.org/10.18637/jss.v033.i01>.
- [17] R.B. Gramacy, monomvn: Estimation for MVN and Student-t Data with Monotone Missingness, (2022). <https://cran.r-project.org/web/packages/monomvn/index.html>.
- [18] M. Kuhn, Building predictive models in R using the caret package, Journal of Statistical Software. 28 (2008) 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- [19] E. Schubert, P.J. Rousseeuw, 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. [https://doi.org/10.1007/978-3-030-32047-8\\_16](https://doi.org/10.1007/978-3-030-32047-8_16).

- [20] H.-G. Drost, Philentropy: Information Theory and Distance Quantification with R, Journal of Open Source Software. 3 (2018) 765. <https://doi.org/10.21105/joss.00765>.
- [21] A. Kassambara, F. Mundt, factoextra: Extract and Visualize the Results of Multivariate Data Analyses, (2020).  
<https://cran.r-project.org/web/packages/factoextra/index.html>.
- [22] Hadley. Wickham, ggplot2: Elegant Graphics for Data Analysis, 1st ed., Springer-Verlag, New York, 2016. <https://ggplot2.tidyverse.org>.
- [23] J.G. Pérez-Silva, M. Araujo-Voces, V. Quesada, NVenn: Generalized, quasi-proportional Venn and Euler diagrams, in: Bioinformatics, Oxford Academic, 2018: pp. 2322–2324. <https://doi.org/10.1093/bioinformatics/bty109>.
- [24] C.O. Wilke, Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures, 1st ed., O'Reilly Media, Sebastopol, 2019.
- [25] D. Gohel, flextable: Functions for Tabular Reporting, (2022). <https://cran.r-project.org/web/packages/flextable/index.html>.
- [26] J. Allaire, Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, rmarkdown: Dynamic Documents for R, (2022).  
<https://cran.r-project.org/web/packages/rmarkdown/index.html>.
- [27] Y. Xie, knitr: A General-Purpose Package for Dynamic Report Generation in R, (2022). <https://cran.r-project.org/web/packages/knitr/index.html>.
- [28] Y. Xie, Bookdown: Authoring books and technical documents with R Markdown, 2016. <https://doi.org/10.1201/9781315204963>.
- [29] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society. Series B: Statistical Methodology. 67 (2005) 301–320.  
<https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [30] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society. Series B (Methodological). 58 (1996) 267–288.  
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- [31] T. Park, G. Casella, The Bayesian Lasso, Journal of the American Statistical Association. 103 (2008) 681–686. <https://doi.org/10.1198/016214508000000337>.
- [32] A. Kirpich, E.A. Ainsworth, J.M. Wedow, J.R.B. Newman, G. Michailidis, L.M. McIntyre, Variable selection in omics data: A practical evaluation of small sample sizes, PLoS ONE. 13 (2018). <https://doi.org/10.1371/journal.pone.0197910>.
- [33] B. Ripley, MASS: Support Functions and Datasets for Venables and Ripley's MASS, (2022). <https://cran.r-project.org/package=MASS>.

- [34] O. Morozova, O. Levina, A. Uusküla, R. Heimer, Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia, *BMC Medical Research Methodology*. 15 (2015) 1–17.  
<https://doi.org/10.1186/s12874-015-0066-2>.
- [35] T. Lange, V. Roth, M.L. Braun, J.M. Buhmann, Stability-based validation of clustering solutions, *Neural Computation*. 16 (2004) 1299–1323.  
<https://doi.org/10.1162/089976604773717621>.
- [36] M. Leng, J. Wang, J. Cheng, H. Zhou, X. Chen, Adaptive semi-supervised clustering algorithm with label propagation, *Journal of Software Engineering*. 8 (2014) 14–22.  
<https://doi.org/10.3923/jse.2014.14.22>.