# **Distinct smell and taste disorder phenotype of post-acute COVID-19 sequelae**

# **Supplementary Material**

HACT and CovILD study teams

# **Supplementary Methods**

# **Study procedures and variables**

The complete list of study variables and stratification scheme is provided in **Supplementary Table S1** for the HACT study and in **Supplementary Table S2** for the CovILD study.

## **COVID-19 symptoms**

A total of 42 was surveyed in the HACT study cohorts (**Supplementary Figure S1** and **Supplementary Table S1**). The symptom duration was coded as follows: absent: 0 days, 1 - 3 days: 3 days, up to 1 week: 7 days, up to 2 weeks: 14 days, up to 4 weeks: 28 days, up to 3 months: 90 days, up to 6 months: 90 days, over 6 months: 90 days. Acute symptoms were defined as complaints present during the first 14 days after clinical onset of COVID-19. Long COVID symptoms were defined as complaints present for *≥* 28 days. PASC symptoms were defined as manifestations present for *≥* 90 days (1,2).

In the CovILD study, a total of 8 symptoms (reduced physical performance, hyposmia/anosmia, dyspnea, sleep problems, cough, fever, night sweating, gastrointestinal symptoms) were prospectively recorded with a standardized questionnaire at each of 60-, 100-, 180- and 360-day post COVID-19 follow-up (**Supplementary Figure S3** and **Supplementary Table S2**). Acute COVID-19 symptoms were assessed retrospectively (3,4).

## **Rating of physical recovery, mental health and quality of life in the HACT study**

Self-perceived complete recovery, rehabilitation need and new medication since COVID-19 at the time of study participation were surveyed as single yes/no items. Percentage of physical performance loss as compared with the time before COVID-19 was rated with a 0 - 100% scale (5,6). Quality of life impairment (QoL) and overall mental health impairment (OMH) were rated with a 4 item Likert scale each (possible answers: "excellent," "good," "fair," "poor," scored: 0, 1, 2, and 3) (5,6). Anxiety (ANX) and depression (DPR) were assessed with the Patient Health Questionnaire (PHQ-4) (5–7). Psychosocial stress was scored with a modified 7 item PHQ stress module as described before (5,6,8).

## **Rating of hyposmia with sniffing stick test**

Objective hyposmia at the 100-day and 360-day follow-up in the CovILD study participants was investigated with the 16-item sniffing stick test as described (9). Clinically relevant hyposmia was defined as *≤* 12 correct answers (9,10). In the analysis, participants with the complete answers concerning self-reported hyposmia and complete test results were included.

# **Statistical analysis**

## **Data transformation, descriptive statistic**

Data transformation and statistical analysis was done with R version 4.0.5 with *tidyverse* data science environment (11). Analysis results were visualized with *ggplot2* (12), *cowplot* (13) as well as in-house developed *ExDA* [\(https://github.com/PiotrTymoszuk/ExDA](https://github.com/PiotrTymoszuk/ExDA)) and *figur* (<https://github.com/PiotrTymoszuk/figur>) packages.

Descriptive statistics including median with interquartile ranges and frequency of complete answers for numeric and categorical variables were calculated with base R functions and *ExDA* package.

## **Statistical hypothesis testing**

Since multiple study variables were non-normally distributed as assessed by Shapiro-Wilk test and visual assessment of their distribution (quantile-quantile plots), statistical significance for differences in outcome numeric variables were assessed with Mann-Whitney U test by Wilcoxon r effect size statistic (two groups) or Kruskal-Wallis test with *η* 2 effect size statistic. Differences in frequency distribution for categorical outcome variables were assessed by  $\chi^2$  test with Cramer V effect size statistic. Interrater assessment of self-reported and sniffing test hyposmia was accomplished with Cohen's *κ* statistic (14). *κ* significance (*κ ≠*0) was estimated with Wald Z test. P values were adjusted for multiple testing with Benjamini-Hochberg method (15) separately for each analysis task and cohort. Packages *rstatix* (16), *vcd* (17) and *Exda* [\(https://github.com/PiotrTymoszuk/ExDA\)](https://github.com/PiotrTymoszuk/ExDA) were used for statistical hypothesis testing.

## **Modeling of symptom recovery kinetic**

To model recovery kinetics for binary symptom variables (0: absent, 1: present), secondorder mixed-effect logistic (categorical features) modeling was applied (packages: *lme4*, *lmerTest* and development package *kinet* [\[https://github.com/PiotrTymoszuk/kinet\]](https://github.com/PiotrTymoszuk/kinet)) (18– 20). Each model followed the general formula:

```
Re s p on s e∼ t im e+t im e
2
+(1∨i n di v i du al)
```
where (1∨*i nd i v id ua l*) indicates the random effect of the individual and *t i me* and *t i me* 2 indicate the first- and second-order time effect terms. The first-order term estimate was interpreted as a measure of the recovery speed and the second-order term estimate was used to assess the plateau/rebound effect. Significance of the accuracy gain of the full second-order model compared with the nested null model was determined by likelihood ratio test (LRT) versus the nested first-order and null models, respectively. Likelihood ratio *λ* statistic (full versus null model) was used as an effect size measure. Individuals from the HACT or CovILD studies with the complete longitudinal symptom record were included in

the kinetic modeling tasks. Results of the kinetic modeling were adjusted for multiple comparisons with Benjamini-Hochberg method (15).

# **Symptom-symptom distances and multi-dimensional scaling**

To assess co-occurrence or exclusivity of symptoms, simple matching distances between manifestations of each acute COVID-19, long COVID and PASC in the HACT study cohorts were calculated (package *scrime* and development package *clustTools* [\[https://github.com/PiotrTymoszuk/clustTools](https://github.com/PiotrTymoszuk/clustTools)]) (21,22). Subsequently, the distance matrix was subjected to multi-dimensional scaling (MDS, k = 2 dimensions, package *stats*, function *cmdscale()*). Association of specific symptoms was assessed by visual analysis of MDS coordinate plots.

## **Apriori analysis of acute COVID-19, long COVID and PASC symptoms in the HACT study**

Frequent combinations of symptoms of long COVID and PASC in the HACT study cohorts were identified with the apriori algorithm (package *arules*) (23,24) with the minimal support cutoff of 0.1, 2 - 10 item transaction length, confidence > 0.8 and lift > 2. The support statistic were used to estimate the symptom combination frequency. The confidence value was treated as an estimate of conditional probability of the symptom cooccurrence. The lift statistic was interpreted as a measure of the symptom dependence (lift = 1, symptoms are independent). The complete analysis results are presented in **Supplementary Tables S3** and **S4**.

## **Clustering analysis**

COVID-19 recovery clusters of the training Autria (AT) HACT cohort participants in respect to symptom-specific recovery times (**Figure 1A**) were defined with the PAM (partitioning around medoids) algorithm and Euclidean distance statistic (packages *cluster*, *philentropy* and development package *clustTools* [\[https://github.com/PiotrTymoszuk/clustTools](https://github.com/PiotrTymoszuk/clustTools)]) (25,26). The set of participants with the complete clustering variable set was included in the analysis. The symptom recovery times were not subjected to any type of preprocessing. 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, package *clustTools*) (27,28) for several clustering algorithms as presented in **Supplementary Figure S8A**. The optimal number of clusters was determined by the bend of the total within-cluster sum of squares curve (**Supplementary Figure S8B**, package *factoextra*) (29). Permutation importance of specific clustering variables was investigated by calculating difference in clustering variance (ratio of total between-cluster sum of squares to total sum of squares) between the initial clustering object and the clustering object with the given variable reshuffled at random (package *clustTools*). Assignment of the Italy HACT cohort participants to the recovery clusters was accomplished with k-nearest neighbors label propagation algorithm  $(k = 5)$  (28). The clustering efficacy in the training AT cohort and the test IT cohort measured by clustering variance statistic defined above was similar (AT: 0.59, IT: 0.57).

## **Data and source code availability**

The raw data files will be made available upon request. The entire analysis pipeline was published at [https://github.com/PiotrTymoszuk/hyposmia\\_analsis\\_pipeline](https://github.com/PiotrTymoszuk/hyposmia_analsis_pipeline).

# **Supplementary Tables**

# **Supplementary Table S1:** HACT study variables.











# **Supplementary Table S2:** CovILD study variables.

**Supplementary Table S3:** Results of apriori analysis for long COVID-19 symptoms in the HACT study . Top 10 most frequent symptom combinations (highest support value) in the AT (Austria) and IT (Italy) cohort are presented. The full results are available as a supplementary Excel table.



**Supplementary Table S4:** Results of apriori analysis for PASC symptoms in the HACT study. Top 10 most frequent symptom combinations (highest support value) in the AT (Austria) and IT (Italy) cohort are presented. The full results are available as a supplementary Excel table.



**Supplementary Table S5:** Demographic and baseline clinical characteristic at the COVID-19 onset of the HACT study participants assigned to the recovery clusters, Austria (AT) cohort.

<b>Variable</b>	Cluster #1	Cluster #2	Cluster #3	Significance <sup>a</sup>	Effect size <sup>a</sup>
Sex	female: 79% (n = 77) male: $21\%$ (n = 21) complete: $n = 98$	female: $57\%$ (n = 140) male: $43\%$ (n = 106) complete: $n = 246$	female: $76\%$ (n = 103) male: $24\%$ (n = 32) complete: $n = 135$	p < 0.001	$V = 0.22$
Age, years	median: 42 [IQR: 30 - 50] range: 21 - 80 complete: $n = 98$	median: 43 [IQR: 29 - 53] range: 18 - 77 complete: $n = 246$	median: 48 [IQR: 38 - 53] range: 21 - 70 complete: $n = 135$	$p = 0.045$	$\eta^2 = 0.012$
<b>BMI</b> before COVID- 19 <sup>b</sup>	normal: $62\%$ (n = 60) overweight: $24\%$ (n = 23) obesity: $14\%$ (n = 14) complete: $n = 97$	normal: $55\%$ (n = 133) overweight: $29\%$ (n = 70) obesity: $17\%$ (n = 41) complete: $n = 244$	normal: $47\%$ (n = 64) overweight: $31\%$ (n = 42) obesity: $21\%$ (n = 29) complete: $n = 135$	$ns (p = 0.39)$	$V = 0.073$
Education	non-tertiary: 64% (n $= 62$ tertiary: $36\%$ (n = 35) complete: $n = 97$	non-tertiary: 63% (n $= 154$ tertiary: $37\%$ (n = 92) complete: $n = 246$	non-tertiary: 64% (n $= 86$ tertiary: $36\%$ (n = 49) complete: $n = 135$	$ns (p = 0.99)$	$V = 0.012$
<b>Employment status</b>	employed: $87\%$ (n = 85) unemployed: 7.1% (n $= 7$ leave: $3.1\%$ (n = 3) retired: $3.1\%$ (n = 3) complete: $n = 98$	employed: $80\%$ (n = 198) unemployed: 9.3% (n $= 23$ leave: $1.6\%$ (n = 4) retired: $8.5\%$ (n = 21) complete: $n = 246$	employed: $85\%$ (n = 115) unemployed: 7.4% (n $= 10$ leave: $0.74\%$ (n = 1) retired: $6.7\%$ (n = 9) complete: $n = 135$	$ns (p = 0.56)$	$V = 0.079$
Observation time	median: 180 [IQR: $130 - 210$ ] range: 93 - 400 complete: $n = 98$	median: 190 [IQR: $130 - 220$ ] range: 90 - 400 complete: $n = 246$	median: 180 [IQR: $140 - 220$ ] range: 90 - 380 complete: $n = 135$	$ns (p = 0.85)$	$\eta^2$ = - 0.0029
Comorbidity	$46\%$ (n = 45) complete: $n = 98$	$44\%$ (n = 109) complete: $n = 246$	$61\%$ (n = 83) complete: $n = 135$	$p = 0.0095$	$V = 0.15$
Hypertension	$9.2\%$ (n = 9) complete: $n = 98$	$10\% (n = 25)$ complete: $n = 246$	$13\% (n = 17)$ complete: $n = 135$	$ns (p = 0.82)$	$V = 0.041$
Cardiovascular disease	$0\%$ (n = 0) complete: $n = 98$	$2.8\%$ (n = 7) complete: $n = 246$	$2.2\%$ (n = 3) complete: $n = 135$	$ns (p = 0.36)$	$V = 0.076$



 $^{\rm a}$ Categorical variables:  $\chi^2$  test with Cramer V effect size statistic. Numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic. P values corrected form multiple testing with Benjamini-Hochberg method.

 $b$ BMI: body mass index, overweight > 25 kg/m<sup>2</sup>, obesity > 30 kg/m<sup>2</sup>,

<sup>c</sup>Frequent respiratory infections, > 2 per year.

<sup>d</sup>Frequent bacterial infections with antibiotic therapy, > 2 per year.



**Supplementary Table S6:** Demographic and baseline clinical characteristic at the COVID-19 onset of the HACT study participants assigned to the recovery clusters, Italy (IT) cohort.



 $^{\rm a}$ Categorical variables:  $\chi^2$  test with Cramer V effect size statistic. Numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic. P values corrected form multiple testing with Benjamini-Hochberg method.

 $b$ BMI: body mass index, overweight > 25 kg/m<sup>2</sup>, obesity > 30 kg/m<sup>2</sup>,

<sup>c</sup>Frequent respiratory infections, > 2 per year.

<sup>d</sup>Frequent bacterial infections with antibiotic therapy, > 2 per year.



**Supplementary Table S7:** COVID-19 course and recovery in the HACT study participants assigned to the recovery clusters, Austria (AT) cohort.



 $^{\rm a}$ Categorical variables:  $\chi^2$  test with Cramer V effect size statistic. Numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic. P values corrected form multiple testing with Benjamini-Hochberg method.

**bNumber of acute COVID-19 symptoms.** 

<sup>c</sup>Number of long COVID symptoms.

d Self-reported incomplete recovery.

<sup>e</sup>ANX: anxiety.

<sup>f</sup>DPR: depression.

<sup>g</sup>OMH score: overall mental health impairment score

hQoL score: quality of life impairment score



#### **Supplementary Table S8:** COVID-19 course and recovery in the HACT study participants assigned to the recovery clusters, Italy (AT) cohort.



 $^{\rm a}$ Categorical variables:  $\chi^2$  test with Cramer V effect size statistic. Numeric variables: Kruskal-Wallis test with  $\eta^2$  effect size statistic. P values corrected form multiple testing with Benjamini-Hochberg method.

**bNumber of acute COVID-19 symptoms.** 

<sup>c</sup>Number of long COVID symptoms.

d Self-reported incomplete recovery.

<sup>e</sup>ANX: anxiety.

<sup>f</sup>DPR: depression.

<sup>g</sup>OMH score: overall mental health impairment score

hQoL score: quality of life impairment score

### **Supplementary Figures**  $\mathbf{A}$

#### **Symptom frequency**

#### AT, HACT study, % cohort

#### **Symptom frequency**

#### IT, HACT study, % cohort



 $\mathbf B$ 

 $n = 479$ 

 $n = 427$ 

**Supplementary Figure S1. Frequency of COVID-19 symptoms in acute COVID-19, long COVID and PASC in the HACT study.**

Frequency of symptoms of acute COVID-19 (first 14 days post clinical onset), long COVID (*≥* 28 days) and PASC (post-acute sequelae of COVID-19, *≥* 90 days) in the AT (Austria, **A**) and IT (Italy, **B**) HACT study cohorts expressed as percentages of the cohort. Point size and color represents the percentage. Numbers of complete observations are indicated under the plot.

Dim. appetite: diminished appetite, Imp. concentration: impaired concentration, Imp. walk: impaired walk, Imp. FMS: impaired fine motor skills.



#### **Supplementary Figure S2. Kinetic of recovery from leading acute COVID-19 symptoms in the HACT study.**

Percentages of individuals with fever (**A**), diminished appetite (**B**), joint pain (**C**), muscle pain (**D**), fatigue (**E**) and tachypnea (**F**) in the AT (Austria) and IT (Italy) HACT study cohorts at particular time points after clinical onset. Numbers of complete observations are indicated under the plots.

















 $\overline{\mathsf{A}}$ 

 $\mathbf{C}$ 

## **Supplementary Figure S3. Symptom frequency in ambulatory, moderate and severe COVID-19 subsets of the CovILD study.**

Frequency of symptoms during acute COVID-19 and at the 60-, 100-, 180- and 360-day follow-ups in ambulatory (**A**), moderate (**B**) and severe COVID-19 (**C**) participants of the CovILD study expressed as percentages of individuals with the complete longitudinal data set. Point size and color represents the percentage. Numbers of complete observations are indicated under the plots.



#### **Supplementary Figure S4. Kinetic of recovery from smell disorders, reduced performance and dyspnea in ambulatory, moderate and severe COVID-19 subsets of the CovILD study.**

Percentages of individuals with the complete longitudinal data set suffering from smell disorders (**A**), reduced physical performance (**B**) and dyspnea (**C**) in the ambulatory, moderate and severe COVID-19 subsets of the CovILD study during acute COVID-19 and at the 60-, 100-, 180- and 360-day follow-ups. Numbers of complete observations are indicated under the plots.



#### **Supplementary Figure S5. Rates of self-reported hyposmia and hyposmia in sniffing stick test at 3-month post COVID-19 follow-up in the ambulatory, moderate and severe COVID-19 subsets of the CovILD study.**

Association of self-reported and sniffing stick test hyposmia rates was investigated with Cohen's *κ* statistic. Statistical significance was assessed with Wald's Z test corrected for multiple testing with Benjamini-Hochberg method. Percentages of individuals with selfreported and test hyposmia within the entire cohort (**A**), the ambulatory (**B**), moderate (**C**) and severe (**D**) COVID-19 subsets of the CovILD-19 subsets are presented in bubble plots. Point size and color represents the percentage. *κ* values with 95% confidence intervals, p values and numbers of complete observations are indicated in the plot captions.



#### **Supplementary Figure S6. Rates of self-reported hyposmia and hyposmia in sniffing stick test at 1-year post COVID-19 follow-up in the ambulatory, moderate and severe COVID-19 subsets of the CovILD study.**

Association of self-reported and sniffing stick test hyposmia rates was investigated with Cohen's *κ* statistic. Statistical significance was assessed with Wald's Z test corrected for multiple testing with Benjamini-Hochberg method. Percentages of individuals with selfreported and test hyposmia within the entire cohort (**A**), the ambulatory (**B**), moderate (**C**) and severe (**D**) COVID-19 subsets of the CovILD-19 subsets are presented in bubble plots. Point size and color represents the percentage. *κ* values with 95% confidence intervals, p values and numbers of complete observations are indicated in the plot captions.



**Supplementary Figure S7. Multi-dimensional scaling analysis of acute COVID-19 symptoms in the HACT study.**

Binary symptom occurrence data for acute COVID-19 (first 14 days after clinical onset) in the HACT study Austria (AT) and Italy (IT) cohorts were subjected to two-dimensional multi-dimensional scaling (MDS) with simple matching distance (SMD) between the symptoms. MDS coordinates are presented in point plots. Selected data points are labeled with the symptom names. Percentages of the data set variance associated with the MDS dimensions are indicated in the plot axes. Numbers of complete observations are indicated under the plots.



#### **Supplementary Figure S8. Definition of the COVID-19 recovery clusters and clustering feature importance in the HACT study.**

Individuals of the training Austria (AT) cohort of the HACT study were subjected to clustering in respect to symptom-specific recovery times with the PAM (partitioning around medoids) algorithm and Euclidean distance measure.

**(A)** Comparison of performance of various algorithms (HCl: hierarchical clustering, SOM + HCl: combined self-organizing map and hierarchical clustering, k-means) and distance statistic in clustering of the training data set investigated by clustering variance (ratio of

total between-cluster sum of squares to total sum of squares) and cluster assignment accuracy in 10-fold cross-validation (CV).

**(B)** Determination of the optimal cluster number in the PAM clustering of the training cohort by the bend of the total within-cluster sum of squares curve.

**(C)** Permutation importance of the clustering features (symptoms) for clustering of the training cohort expressed as the difference in clustering variance (ratio of total betweencluster sum of squares to total sum of squares) between the initial clustering object and the clustering object with the given variable reshuffled at random.



#### **Supplementary Figure S9. Numbers of acute COVID-19 and long COVID symptoms in the HACT study recovery clusters.**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in numbers of acute COVID-19 (first 14 days after clinical onset, **A**) and long COVID symptoms (*≥* 28 days, **B**) between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. Symptom counts are presented in violin plots. Points represent single observations, orange diamonds with whiskers code for medians and interquartile ranges. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.



#### **Supplementary Figure S10. Recovery from smell and taste disorders in the HACT study recovery clusters**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in recovery times for smell (**A**) and taste disorders (**D**) between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. Symptom counts are presented in violin plots. Points represent single observations, orange diamonds with whiskers code for medians and interquartile ranges. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.



#### **Supplementary Figure S11. Recovery from fatigue, tiredness and tachypnea in the HACT study recovery clusters.**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in recovery times for fatigue (**A**), day tiredness (**B**) and tachypnea (**C**) between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. Symptom counts are presented in violin plots. Points represent single observations, orange diamonds with

whiskers code for medians and interquartile ranges. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.



#### **Supplementary Figure S12. Recovery from neurocognitive symptoms of COVID-19 in the HACT study recovery clusters.**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in recovery times for forgetfulness (**A**), concentration impairment (**B**) and confusion (**C**) between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. Symptom counts are presented in violin plots. Points represent single observations, orange

diamonds with whiskers code for medians and interquartile ranges. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.



**Supplementary Figure S13. Physical performance loss, impairment of quality of life and self-perceived complete convalescence at the survey completion in the HACT study recovery clusters.**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in self-reported physical performance loss (**A**) and scoring of impairment of quality of life (QoL, **B**) at the survey completion between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. The scoring data were presented in violin plots. Points represent single observations, orange diamonds with whiskers code for medians and interquartile ranges. Differences in frequency of self-perceived complete convalescence  $(C)$  between the clusters were investigated by  $\chi^2$  test and Cramer V effect size statistic. The frequencies are presented as bar plots. P values were corrected for multiple testing with Benjamini-Hochberg method. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.



### **Supplementary Figure S14. Rating of anxiety, depression, overall mental health impairment and stress at the survey completion in the HACT study recovery clusters.**

The recovery clusters were defined in the training Austria cohort (AT, **Supplementary Figure S8**) and the cluster assignment predicted in the test Italy cohort (IT, **Figure 5**). Differences in rating of anxiety (ANX, **A**), depression (DPR, **B**), stress (**C**) and self-reported overall mental health impairment (OMH, **D**) at the survey completion between the clusters were assessed by Kruskal-Wallis test and  $\eta^2$  effect size statistic. P values were corrected for multiple testing with Benjamini-Hochberg method. The scoring data were presented in violin plots. Points represent single observations, orange diamonds with whiskers code for medians and interquartile ranges. Effect size statistics and p values are indicated in the plot caption. Numbers of complete observations are displayed under the plots.

## **References**

1. Venkatesan P. NICE guideline on long COVID. *The Lancet Respiratory medicine* (2021) 9:129. doi: [10.1016/S2213-2600\(21\)00031-X](https://doi.org/10.1016/S2213-2600(21)00031-X)

2. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19: Summary of NICE, SIGN, and RCGP rapid guideline. *The BMJ* (2021) 372: doi: [10.1136/bmj.n136](https://doi.org/10.1136/bmj.n136)

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

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

5. Hüfner K, Tymoszuk P, Ausserhofer D, Sahanic S, Pizzini A, Rass V, Galffy M, Böhm A, Kurz K, Sonnweber T, et al. Who Is at Risk of Poor Mental Health Following Coronavirus Disease-19 Outpatient Management? *Frontiers in medicine* (2022) 9: doi: [10.3389/FMED.2022.792881](https://doi.org/10.3389/FMED.2022.792881)

6. Sahanic S, Tymoszuk P, Ausserhofer D, Rass V, Pizzini A, Nordmeyer G, Hüfner K, Kurz K, Weber PM, Sonnweber T, et al. Phenotyping of acute and persistent COVID-19 features in the outpatient setting: exploratory analysis of an international cross-sectional online survey. *Clinical Infectious Diseases* (2021) doi: [10.1093/CID/CIAB978](https://doi.org/10.1093/CID/CIAB978)

7. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, Schneider A, Brähler E. A 4 item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders* (2010) 122:86–95. doi: [10.1016/j.jad.2009.06.019](https://doi.org/10.1016/j.jad.2009.06.019)

8. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening psychischer störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)". Ergebnisse der Deutschen validierungsstudie. *Diagnostica* (2004) 50:171–181. doi: [10.1026/0012-1924.50.4.171](https://doi.org/10.1026/0012-1924.50.4.171)

9. Rass V, Beer R, Schiefecker AJ, Kofler M, Lindner A, Mahlknecht P, Heim B, Limmert V, Sahanic S, Pizzini A, et al. Neurological outcome and quality of life 3 months after COVID-19: A prospective observational cohort study. *European Journal of Neurology* (2021) doi: [10.1111/ene.14803](https://doi.org/10.1111/ene.14803)

10. Hummel T, Kobal G, Gudziol H, Mackay-Sim A. Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. *European archives of oto-rhinolaryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery* (2007) 264:237–243. doi: [10.1007/S00405-006-0173-0](https://doi.org/10.1007/S00405-006-0173-0)

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

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

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

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

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

16. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2021) <https://cran.r-project.org/package=rstatix>

17. Meyer D, Zeileis A, Hornik K. vcd: Visualizing Categorical Data. (2021) [https://cran.r](https://cran.r-project.org/web/packages/vcd/index.html)[project.org/web/packages/vcd/index.html](https://cran.r-project.org/web/packages/vcd/index.html)

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

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

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

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

22. Schwender H, Fritsch A. scrime: Analysis of High-Dimensional Categorical Data Such as SNP Data. (2018) <https://cran.r-project.org/web/packages/scrime/index.html>

23. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. *ACM SIGMOD Record* (1993) 22:207–216. doi: [10.1145/170036.170072](https://doi.org/10.1145/170036.170072)

24. Hahsler M, Grün B, Hornik K. Arules - A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* (2005) 14: doi: [10.18637/JSS.V014.I15](https://doi.org/10.18637/JSS.V014.I15)

25. Drost H-G. Philentropy: Information Theory and Distance Quantification with R. *Journal of Open Source Software* (2018) 3:765. doi: [10.21105/joss.00765](https://doi.org/10.21105/joss.00765)

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

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

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

29. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. (2020)<https://cran.r-project.org/web/packages/factoextra/index.html>