Who is at risk of poor mental health following COVID-19 outpatient management?

Supplementary Material

Health after COVID-19 in Tyrol study team

2021-09-13

Contents

Supplementary Methods

Study design and participants

The multi-center international online survey study 'Health after COVID-19 in Tyrol' (ClinicalTrials.gov: NCT04661462) was conducted between the 30th September 2020 and 11th July 20221 in two independently recruited cohorts in the neighboring central European regions, Tyrol/Austria and South Tyrol/Italy.[1] The study populations encompassed residents of the study regions aged $\geq$ 16 (Austria/Tyrol) or $\geq$ 18 years (Italy/South Tyrol) who experienced a SARS-CoV2 infection corroborated by a laboratory test (PCR or seropositivity). The analysis inclusion criterion was a minimum observation time between the diagnosis of SARS-CoV2 infection and the survey completion $\geq$ 28 days. The analysis exclusion criterion was hospitalization because of COVID-19. The scheme of study and analysis enrollment is depicted in **Figure 1**. The participants were invited by a public media call (both cohorts) or by their general practitioners (Italy/South Tyrol).

The study was conducted in accordance with the Declaration of Helsinki as well as the national and European data policies. Each participant gave a digitally signed informed consent prior to the survey start. The study protocol was approved by the institutional review boards of the Medical University of Innsbruck (Tyrol/Austria, approval number: 1257/2020) and of the South Tyrol province (South Tyrol/Italy, 0150701).

Measures, definitions and data transformation

The detailed description of the questionnaire is provided by Sahanic et al.[1] In brief, dates of the study completion and SARS-CoV2 infection diagnosis, data on biometry (weight, height), demographics (age, sex), pre-existing co-morbidities, socioeconomic status (residence region, mother tongue, employment status and profession), smoking history and COVID-19 relevant medication, symptom duration (44 items), symptomatic therapy and course of SARS-CoV2 infection, recovery duration and status as well as mental health and psychosocial stress following the disease were queried. The complete list of features analyzed in the current report is presented in **Supplementary Table S1** and the baseline characteristic of the study populations is shown in **Table 1** and **Supplementary Table S2**.

Self-reported COVID-19 symptoms were retrospectively assigned to the following duration classes: absent, present for 1 – 3 days, $\leq$ 1 week, $\leq$ 2 weeks, $\leq$ 4 weeks, $\leq$ 3 months, $\leq$ 6 months and $>$ 6 months. Acute symptoms were defined as complaints when present in the first 2 weeks of the disease, sub-acute when present in 2 - 4 weeks after clinical onset, persistent when present for $\geq$ 4 weeks.[1] Confusion, impaired concentration and forgetfulness were classified as 'neurocognitive symptoms'. For modeling, the overall number of overall acute symptoms was stratified by quartiles (Q1 - Q4) and the overall persistent symptom count by median and 75th percentile . The number of acute neuro-cognitive symptoms was stratified by median and 75th percentile and for persistent neuro-cognitive manifestations coded as an index present/absent variable (**Supplementary Table S1** and **Table 2**).

Depression/anxiety before SARS-CoV2 infection, pre-existing sleep disorders, acute COVID-19 perception (common cold-, influenza-, gastroenteritis-like or unique/not experienced before), symptom relapse, complete convalescence, rehabilitation need and percent physical performance loss following COVID-19 were surveyed as single question items each. For modeling, the physical performance loss was stratified as 0 - 25%, 26 - 50%, 51% - 75% and 76 - 100%.

Self-perceived overall mental health (OMH) and quality of life (QoL) were assessed as single questions ('excellent', 'good', 'fair', 'poor', scored: 0, 1, 2, 3). Anxiety/depression following COVID-19 at time of study completion were investigated using PHQ-4 module (two questions each, possible answers: 'never', 'some days', 'over 50% of days', 'almost every day', scoring: 0, 1, 2, 3 points). For separate positivity screening for depression (DPR) or anxiety (ANX), cutoffs of $\geq$ 3 point sum were applied to the respective scales. PHQ-4 questions were also analyzed as a combined depression/anxiety score with a cutoff of $\geq$ 6 point sum.[2] Psychosocial stress factors were measured with an adapted version of the PHQ stress module.[3,4] The items on weight, sexuality and past traumatic/serious events were were removed, the item on worries and

dreams was adapted to COVID-19, resulting in seven questions with possible answers: 'no', 'little', 'some', 'a lot', scoring: 0, 1, 2, 3). For modeling, the stress scoring was re-coded as quartile strata encompassing 0 - 2, 3 - 4, 5 - 6, 7 - 21 points.

Statistical analysis

Data transformation, visualization, descriptive statistic and hypothesis testing

The study variables were transformed, analyzed and visualized with R version 4.0.5 with *tidyverse*,[5,6] *cowplot*[7] and *ggvenn* packages.

For categorical variables, numbers and percents of complete answers are presented. As most of the analyzed numeric features had a discrete or non-normal distribution as checked by Shapiro-Wilk test, medians, interquartile ranges (IQR) and feature ranges are presented. To compare differences in distribution of categorical features, χ^2 test was applied. To assess significance of differences in numeric variables between groups, U or Kruskal-Wallis test was used, as appropriate. P values were corrected for multiple comparisons with Benjamini-Hochberg method.[8] The set of tools used for descriptive statistics and hypothesis testing is available from https://github.com/PiotrTymoszuk/counting-tools.

Random forest modeling of mental health scoring

Multi-parameter random forest regression models[9] describing the scoring of OMH, QoL, ANX and DPR separately in the Austria/Tyrol and Italy/South Tyrol cohorts as functions of 144 independent parameters (**Supplementary Table S1**) were constructed and verified in by 10-fold cross validation using *caret* package (function *train()*, *mtry* argument specifying the number of random tree models was set to 500).[10] The model fits to the training data set are presented in **Supplementary Figures S3 - S2** and were assessed by Spearman regression and mean absolute error (MAE) values for the training and cross-validation data sets.

To discern the features with the largest effect on the OMH, QoL, ANX and DPR scoring each, differences in mean squared error (ΔMSE) associated with the model components were extracted using *importance()* function on the final models developed by *caret* (**Supplementary Figures S3 - S2**).[11] To identify common factors with the greatest impact on the combined mental health scoring, the normalized values of ΔMSE of the OMH, QoL, ANX and DPR scoring for each model component were subjected to centered principal component analysis (PCA) using *PCAproj()* function from *pcaPP* package.[12] The features with the 10 largest Euclidean distances from the PCA center were extracted.

Uni-variate modeling

Correlation of the OMH, QoL, ANX and DPR scoring and the most influential factors identified by the random forest technique was assessed by age- and sex-weighted Poisson regression (generalized linear modeling, log link function). The frequency weights for the Austria/Tyrol and Italy/South Tyrol cohort were based on the age and sex distribution of COVID cases in Tyrol[13] and Italy,[14] respectively[1]. Significance of model estimates and their 95% confidence intervals were determined with Wald Z test. P values were corrected for multiple comparisons by Benjamini-Hochberg method.[8] The correlation was deemed significant when significant association was present in both study cohorts. Model estimate extraction and visual quality control was accomplished with home-developed R tools (https://github.com/PiotrTymoszuk/lm_qc_tools). For complete uni-variate modeling results, see: **Supplementary Table S3**.

Pooled Austria/Italy β estimates referenced to in the text and presented in **Supplementary Table S4** were calculated using inversed variance method and *meta* package.[15,16]

Definition of the mental disorder risk clusters

The study participants, separately in the Austian and Italian cohort, were clustered using a two-step combined self-organizing map (SOM) and hierarchical clustering algorithm[17,18] in respect to the most influential factors affecting the net mental health scoring identified by multi-parameter random forest modeling and PCA. In the first step, participants were assigned to the nodes of 11 × 11 unit hexagonal-topology grid with the Jaccard distance measure between the participants. The grid size was estimated with the $5 \times \sqrt{N}$ formula, where N is the number of observations in the smaller Italian data set, as proposed by Vesanto et al.[19] SOM assignment was accomplished with the tools provided by *kohonen* package and home-developed wrappers (https://github.com/PiotrTymoszuk/SOM_tools). The SOM training process is visualized in **Figure S6A**. In the second step, SOM nodes were subjected to hierarchical clustering with Ward D2 method and Euclidean distance measure. The optimal cluster number (k = 3) was determined by the bend of the within sum-of-squares and visual analysis of the dendrograms (**Supplementary Figure S6BC**). The hierarchical clustering was done with the base *hclust()* function and home-developed wrappers for clustering quality control and visualization (https://github.com/PiotrTymoszuk/cluster_tools).

Data availability

As this study is still ongoing, the complete data will be made available on a serious request to the corresponding author and made publicly available after the completion. Analysis of the psychosocial features is available as an online *R shiny* dashboard at Mental Health after COVID-19 in Tyrol (https://im2-ibk.shinyapps.io/mental_health_dashboard/). The R analysis pipeline is available at https://github.com/PiotrTymoszuk/mental-health-after-COVID-19.[20]

Supplementary Tables

Table S1: **Survey variables used for construction of random forest models.**. The table is available online.

Table S2: **Supplementary characteristic of the study cohorts.**
AT: Austria/Tyrol cohort, IT: Italy/South Tyrol cohort, Test: statistical test used for the AT vs IT comparison, Significance: test p value corrected for multiple comparisons with Benjamini-Hochberg method.

Variable	AT	IT	Test	Significance
Observation time	median(IQR) = 79 (40 - 175) range = 28 - 400 n = 1157	median(IQR) = 96 (60 - 138) range = 28 - 387 n = 893	U	p = 1.2e-07
	up to 60 days: 42.5% (492) 61 - 120 days: 20.2% (234) 121 - 180 days: 14.2% (164) more than 180 days: 23.1% (267) n = 1157	up to 60 days: 25.8% (230) 61 - 120 days: 35.4% (316) 121 - 180 days: 21.6% (193) more than 180 days: 17.2% (154) n = 893	χ^2	p = 1.3e-22
Survey completion	fall 2020: 63.4% (734) winter/spring 2021: 36.6% (423) n = 1157	fall 2020: 4.37% (39) winter/spring 2021: 95.6% (854) n = 893	χ^2	p = 4.2e-163
Region	capital: 19.9% (230) non-capital: 80.1% (927) n = 1157	capital: 56.6% (505) non-capital: 43.4% (388) n = 893	χ^2	p = 5.6e-65
Native language	German: 100% (1157) n = 1157	German: 55.3% (493) Italian: 36.7% (327) Ladin: 6.5% (58) Other: 1.57% (14) n = 892	χ^2	p = 4.5e-138
Employment sector	other: 18.7% (214) gastronomy/tourism: 8.82% (101) health services: 25.9% (296) food trade: 2.18% (25) public transportation: 0.786% (9) emergency services: 2.1% (24) construction: 2.97% (34) administration/office: 19.4% (222) industry: 5.68% (65) agriculture: 0.961% (11) education: 12.6% (144) n = 1145	other: 18.1% (157) gastronomy/tourism: 8.29% (72) health services: 20.1% (175) food trade: 1.84% (16) public transportation: 0.575% (5) construction: 3.11% (27) administration/office: 28.2% (245) industry: 4.95% (43) agriculture: 1.5% (13) education: 13.3% (116) n = 869	χ^2	p = 7.3e-06
Diabetes	no: 98.4% (1139) yes: 1.56% (18) n = 1157	no: 99.2% (886) yes: 0.784% (7) n = 893	χ^2	ns

Table S2: **Supplementary characteristic of the study cohorts.**
AT: Austria/Tyrol cohort, IT: Italy/South Tyrol cohort, Test: statistical test used for the AT vs IT comparison, Significance: test p value corrected for multiple comparisons with Benjamini-Hochberg method. *(continued)*

Variable	AT	IT	Test	Significance
Gastrointestinal disease	no: 98.4% (1139) yes: 1.56% (18) n = 1157	no: 99% (884) yes: 1.01% (9) n = 893	χ^2	ns
Malignancy	no: 97.6% (1129) yes: 2.42% (28) n = 1157	no: 97.1% (867) yes: 2.91% (26) n = 893	χ^2	ns
Freq. resp. infections	no: 95.6% (1106) yes: 4.41% (51) n = 1157	no: 97.1% (867) yes: 2.91% (26) n = 893	χ^2	ns
Freq. bact. Infections	no: 96.1% (1112) yes: 3.89% (45) n = 1157	no: 98.7% (881) yes: 1.34% (12) n = 893	χ^2	p = 0.0017
Hair loss	no: 86.3% (999) yes: 13.7% (158) n = 1157	no: 85.2% (761) yes: 14.8% (132) n = 893	χ^2	ns
Weight loss	none: 52.5% (604) mild: 16.3% (188) moderate: 26% (299) severe: 5.21% (60) n = 1151	none: 60.4% (538) mild: 15.6% (139) moderate: 19.6% (174) severe: 4.38% (39) n = 890	χ^2	p = 0.0025
Complete convalescence	no: 46% (531) yes: 54% (624) n = 1155	no: 36.7% (326) yes: 63.3% (563) n = 889	χ^2	p = 6.6e-05
Physical performance loss	0 - 25%: 74.7% (860) 26 - 50%: 17.5% (202) 51% - 75%: 6.26% (72) 76 - 100%: 1.48% (17) n = 1151	0 - 25%: 76.2% (674) 26 - 50%: 16.4% (145) 51% - 75%: 6.11% (54) 76 - 100%: 1.24% (11) n = 884	χ^2	ns
Subjective need for rehabilitation	no: 83% (957) yes: 17% (196) n = 1153	no: 86.8% (771) yes: 13.2% (117) n = 888	χ^2	p = 0.033

Table S3: **Results of univariate modeling for the most influential mental health scoring factors.** Top 10 factors with the largest impact on the net mental health scoring were determined by random forest prediction and principal component analysis as presented in Figure 2. Their correlation with overall mental health (OMH), quality of life (QoL), anxiety (ANX) and depression (DPR) scoring was investigated by univariate sex- and age-weighted Poisson regression. P values were corrected for multiple testing with Benjamini-Hochberg method. The table is available online.

Table S4: **Pooled Austria/Italy cohort results of univariate modeling for the most influential mental health scoring factors.**
Pooled Austria/Italy β estimates for the factors with the largest impact on the net mental health scroring were calculated with inverse variance method. P values were corrected for multiple testing with Benjamini-Hochberg method. The table is available online.

Table S5: **Prevalence of depression and anxiety as a function of stress scoring, overall and neurocognitive symptom burden.**
Distribution of the depression, anxiety and the combined depression/anxiety positivity was compared between the variable strata by χ^2 test. P values were corrected for multiple comparisons with Benjamini-Hochberg method.

Variable	Strata	Cohort	Anx. prevalence	Depr. prevalence	Anx. significance	Depr. significance
Stress score	1Q	Austria	3.16% (412)	3.65% (411)	p = 0.022	p = 0.0014
	2Q	Austria	7.42% (283)	10.9% (284)		
	3Q	Austria	15.9% (207)	23% (209)		
	4Q	Austria	30.9% (246)	41.9% (246)		
Stress score	1Q	Italy	4.35% (322)	7.76% (322)	p = 0.0043	p = 0.022
	2Q	Italy	12.2% (188)	14.9% (188)		
	3Q	Italy	24.8% (222)	32.4% (222)		
	4Q	Italy	48.7% (158)	51.3% (158)		
# acute symptoms	1Q	Austria	6.76% (340)	6.76% (340)	ns	ns
	2Q	Austria	6.69% (254)	12.2% (255)		
	3Q	Austria	13.4% (292)	19.2% (292)		
	4Q	Austria	24.2% (264)	33.8% (266)		
# acute symptoms	1Q	Italy	4.82% (249)	7.63% (249)	p = 0.0054	ns
	2Q	Italy	14% (236)	13.6% (236)		
	3Q	Italy	26.1% (188)	26.1% (188)		
	4Q	Italy	35.6% (219)	49.1% (218)		
# persistent symptoms	0	Austria	8.49% (601)	11.4% (603)	ns	ns
	1 - 3	Austria	9.76% (287)	13.2% (288)		
	>3	Austria	24.4% (262)	35.5% (262)		
# persistent symptoms	0	Italy	11.1% (452)	14.2% (452)	ns	ns
	1 - 3	Italy	15.2% (231)	19.5% (231)		
	>3	Italy	41.6% (209)	47.1% (208)		
# acute NC	0	Austria	6.48% (571)	5.77% (572)	p = 0.0016	p = 2.4e-10
	1	Austria	14.1% (234)	21.7% (235)		
	2	Austria	14.8% (196)	24.9% (197)		
	3	Austria	29.3% (150)	44.7% (150)		
# acute NC	0	Italy	8.62% (464)	10.6% (464)	p = 0.0016	p = 0.014
	1	Italy	19.7% (127)	19.7% (127)		
	2	Italy	24.7% (146)	32.2% (146)		
	3	Italy	45.5% (156)	55.5% (155)		
	0	Austria	9.15% (940)	12.6% (943)		
	1	Austria	13.1% (84)	27.4% (84)		

Table S5: **Prevalence of depression and anxiety as a function of stress scoring, overall and neurocognitive symptom burden.**
Distribution of the depression, anxiety and the combined depression/anxiety positivity was compared between the variable strata by χ^2 test. P values were corrected for multiple comparisons with Benjamini-Hochberg method. *(continued)*

Variable	Strata	Cohort	Anx. prevalence	Depr. prevalence	Anx. significance	Depr. significance
# persist. NC	2	Austria	28.9% (90)	40% (90)	ns	p = 0.002
	3	Austria	54.1% (37)	59.5% (37)		
# persist. NC	0	Italy	13.3% (691)	16.5% (690)	p = 0.024	ns
	1	Italy	28% (50)	30% (50)		
	2	Italy	29.1% (86)	33.7% (86)		
	3	Italy	62.1% (66)	74.2% (66)		

Table S6: **Differences in frequency of the survey items not used for cluster definition between the mental disorder risk clusters.**
Cluster: LR - low risk, IR - intermediate risk, HR - high risk cluster, N: number of observations assigned to the cluster, Significance: p values obtained by χ^2 test corrected for multiple testing with Benjamini-hochber method. The table is available online.

Supplementary Figures

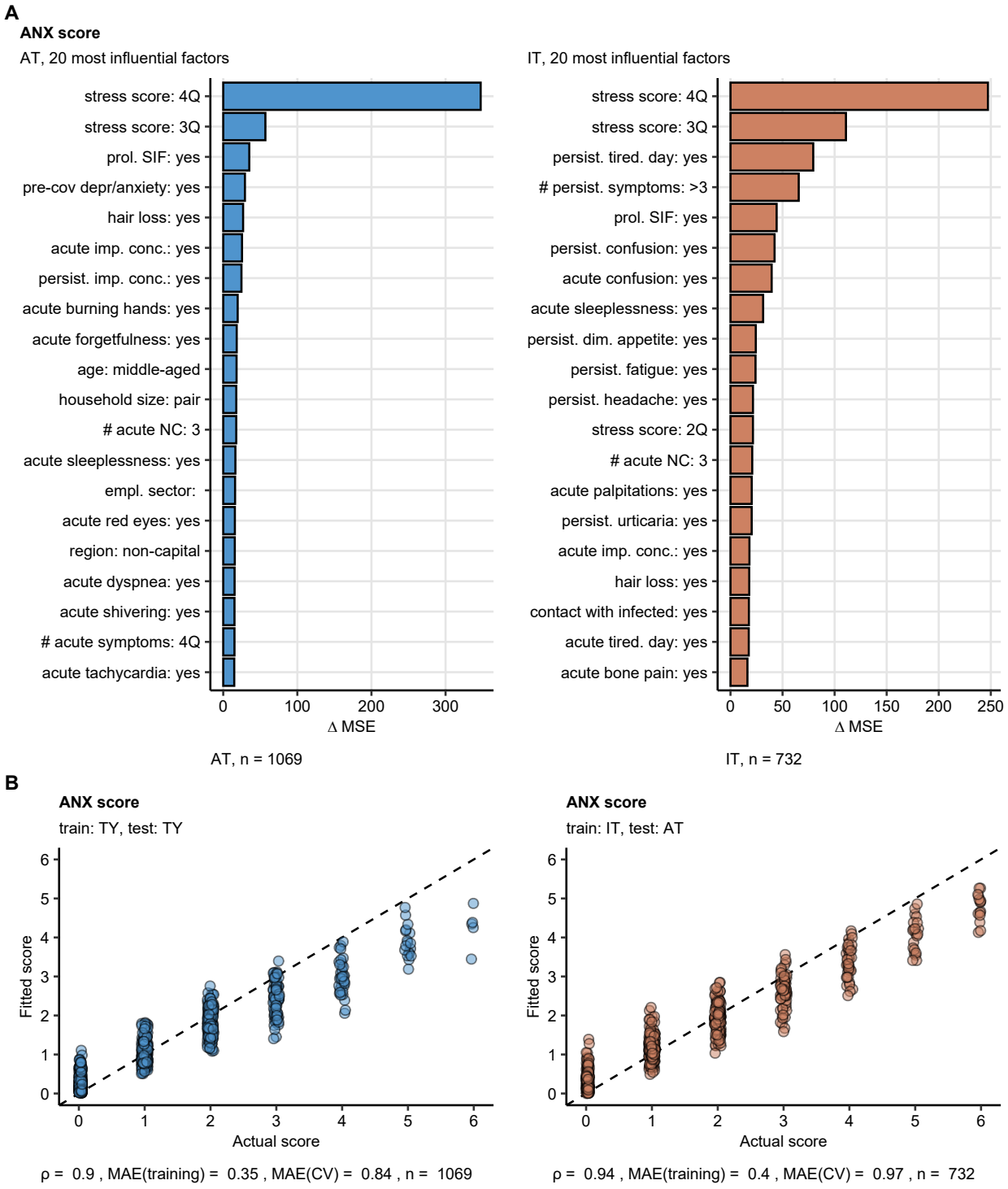

Figure S1: Construction and performance of the anxiety scoring random forest models.

Supplementary Figure S1. Construction and performance of the anxiety scoring random forest models.

Random forest models fitting 144 survey variables (**Supplementary Table S1**) to the anxiety scoring in the Austria/Tyrol (AT) and Italy/South Tyrol (IT) cohorts were constructed and validated by the 10-fold cross-validation (CV) technique.

(**A**) Top 20 most influential factors contributing to the improvement of to model fit measured as difference in mean squared error (ΔMSE). N numbers of observations are indicated below the plot.

(**B**) Fitted versus true scoring values in the Austria/Tyrol and Italy/South Tyrol cohorts. Spearman's ρ correlation coefficients, means absolute errors (MAE) for the whole-cohort and cross-validation data sets and n numbers of observations are indicated below the plot.

imp.: impaired, pre-cov depr/anxiety: depression or anxiety before COVID-19, tired. day: tiredness at day, prol.: prolonged, SIF: severe illness feeling, #: number, NC: neurocognitive symptoms, GP: general practitioner, persist.: persistent, dim.: diminished, 2Q, 3Q, 4Q: 2^{nd}, 3^{rd}, 4^{th} quartile, empl.: employment, conc.: concentration, daily medic.: daily medication, # cov in household: number of COVID-19 cases in the household, subj. cov percept.: subjective perception of acute COVID-19, pre-cov sleep disord.: sleep disorder before COVID-19.

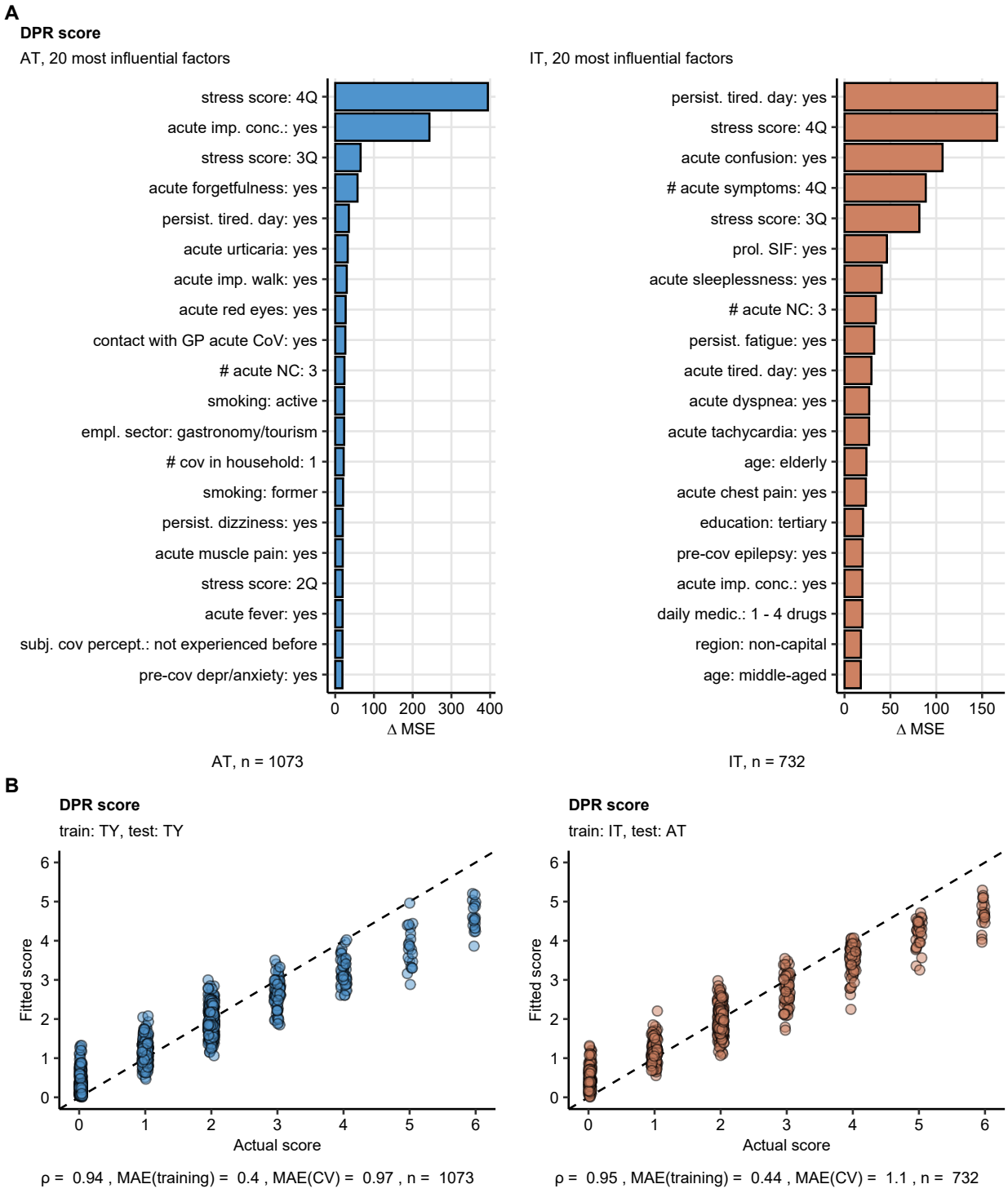

Figure S2: Construction and performance of the depression scoring random forest models.

Supplementary Figure S2. Construction and performance of the depression scoring random forest models.

Random forest models fitting 144 survey variables (**Supplementary Table S1**) to the depression scoring in the Austria/Tyrol (AT) and Italy/South Tyrol (IT) cohorts were constructed and validated by the 10-fold cross-validation (CV) technique.

(**A**) Top 20 most influential factors contributing to the improvement of to model fit measured as difference in mean squared error (ΔMSE). N numbers of observations are indicated below the plot.

(**B**) Fitted versus true scoring values in the Austria/Tyrol and Italy/South Tyrol cohorts. Spearman's ρ correlation coefficients, means absolute errors (MAE) for the whole-cohort and cross-validation data sets and n numbers of observations are indicated below the plot.

imp.: impaired, pre-cov depr/anxiety: depression or anxiety before COVID-19, tired. day: tiredness at day, prol.: prolonged, SIF: severe illness feeling, #: number, NC: neurocognitive symptoms, GP: general practitioner, persist.: persistent, dim.: diminished, 2Q, 3Q, 4Q: 2^{nd}, 3^{rd}, 4^{th} quartile, empl.: employment, conc.: concentration, daily medic.: daily medication, # cov in household: number of COVID-19 cases in the household, subj. cov percept.: subjective perception of acute COVID-19, pre-cov sleep disord.: sleep disorder before COVID-19.

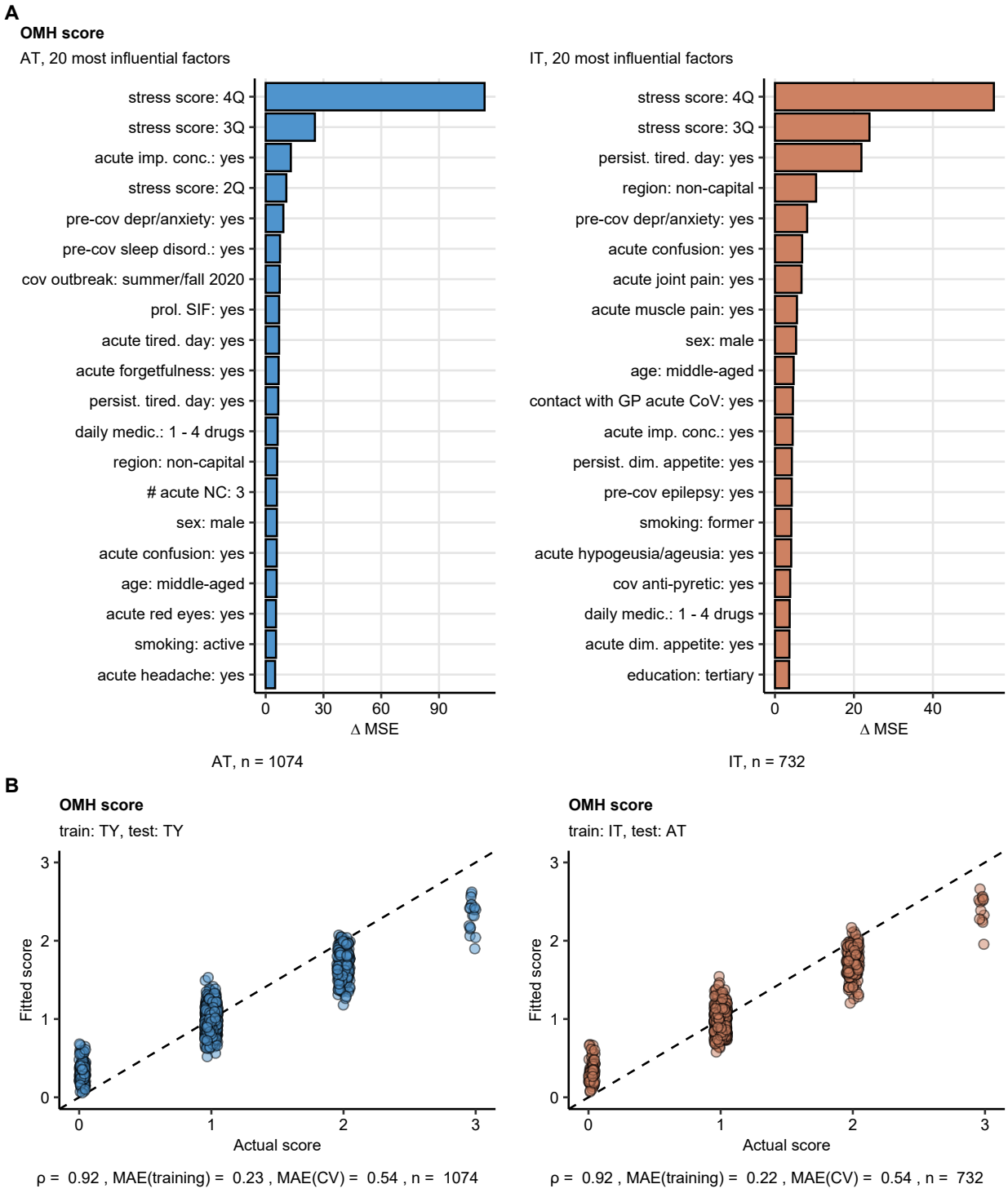

Figure S3: Construction and performance of the overall mental health scoring random forest models.

Supplementary Figure S3. Construction and performance of the overall mental health scoring random forest models.

Random forest models fitting 144 survey variables (**Supplementary Table S1**) to the overall mental health (OMH) scoring in the Austria/Tyrol (AT) and Italy/South Tyrol (IT) cohorts were constructed and validated by the 10-fold cross-validation (CV) technique.

(**A**) Top 20 most influential factors contributing to the improvement of to model fit measured as difference in mean squared error (ΔMSE). N numbers of observations are indicated below the plot.

(**B**) Fitted versus true scoring values in the Austria/Tyrol and Italy/South Tyrol cohorts. Spearman's ρ correlation coefficients, means absolute errors (MAE) for the whole-cohort and cross-validation data sets and n numbers of observations are indicated below the plot.

imp.: impaired, pre-cov depr/anxiety: depression or anxiety before COVID-19, tired. day: tiredness at day, prol.: prolonged, SIF: severe illness feeling, #: number, NC: neurocognitive symptoms, GP: general practitioner, persist.: persistent, dim.: diminished, 2Q, 3Q, 4Q: 2^{nd}, 3^{rd}, 4^{th} quartile, empl.: employment, conc.: concentration, daily medic.: daily medication, # cov in household: number of COVID-19 cases in the household, subj. cov percept.: subjective perception of acute COVID-19, pre-cov sleep disord.: sleep disorder before COVID-19.

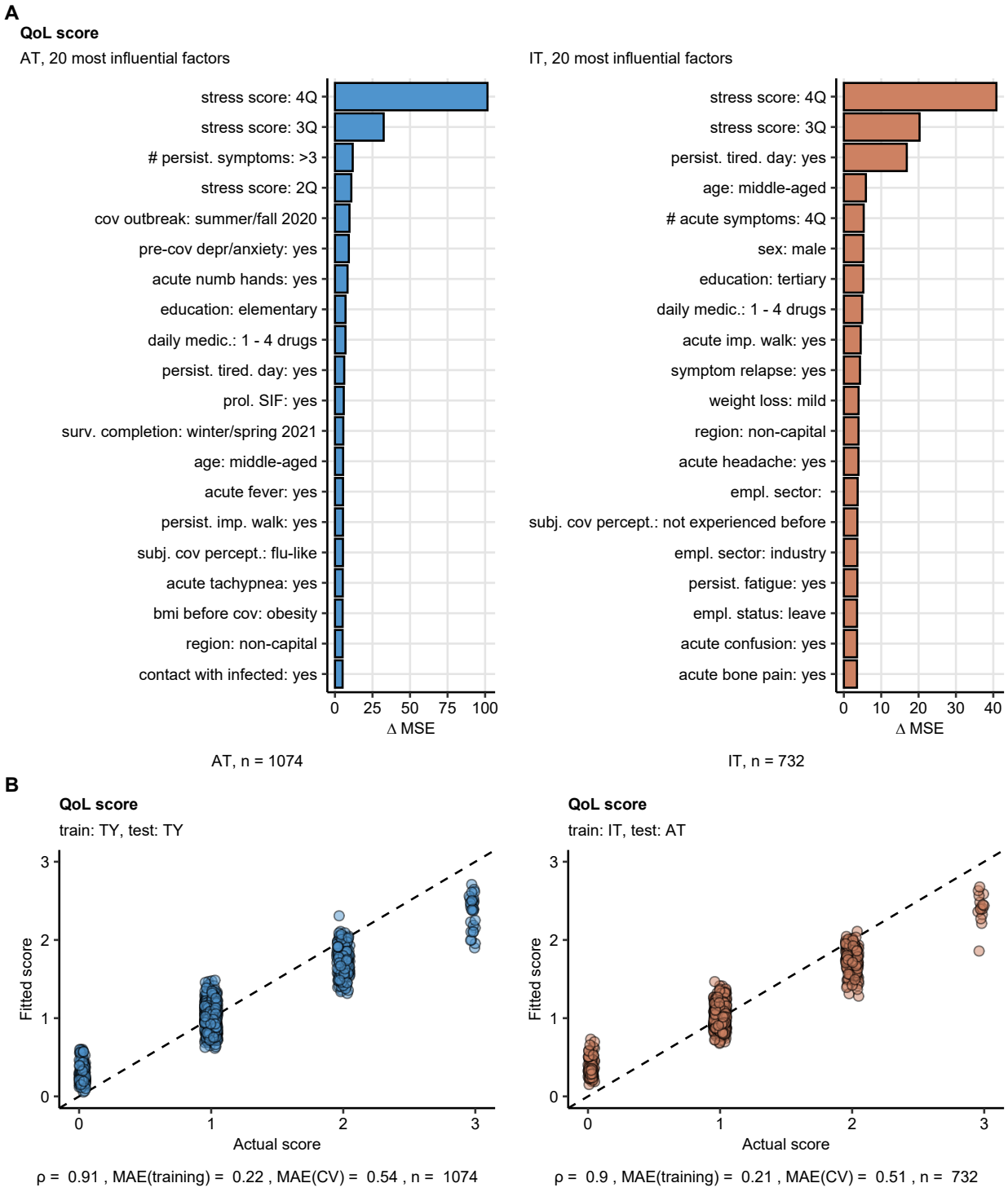

Figure S4: Construction and performance of the quality of life scoring random forest models.

Supplementary Figure S4. Construction and performance of the quality of life scoring random forest models.

Random forest models fitting 144 survey variables (**Supplementary Table S1**) to the quality of life (QoL) scoring in the austria/Tyrol (AT) and Italy/South Tyrol (IT) cohorts were constructed and validated by the 10-fold cross-validation (CV) technique.

(**A**) Top 20 most influential factors contributing to the improvement of to model fit measured as difference in mean squared error (ΔMSE). N numbers of observations are indicated below the plot.

(**B**) Fitted versus true scoring values in the Austria/Tyrol and Italy/South Tyrol cohorts. Spearman's ρ correlation coefficients, means absolute errors (MAE) for the whole-cohort and cross-validation data sets and n numbers of observations are indicated below the plot.

imp.: impaired, pre-cov depr/anxiety: depression or anxiety before COVID-19, tired. day: tiredness at day, prol.: prolonged, SIF: severe illness feeling, #: number, NC: neurocognitive symptoms, GP: general practitioner, persist.: persistent, dim.: diminished, 2Q, 3Q, 4Q: 2^{nd}, 3^{rd}, 4^{th} quartile, empl.: employment, conc.: concentration, daily medic.: daily medication, # cov in household: number of COVID-19 cases in the household, subj. cov percept.: subjective perception of acute COVID-19, pre-cov sleep disord.: sleep disorder before COVID-19.

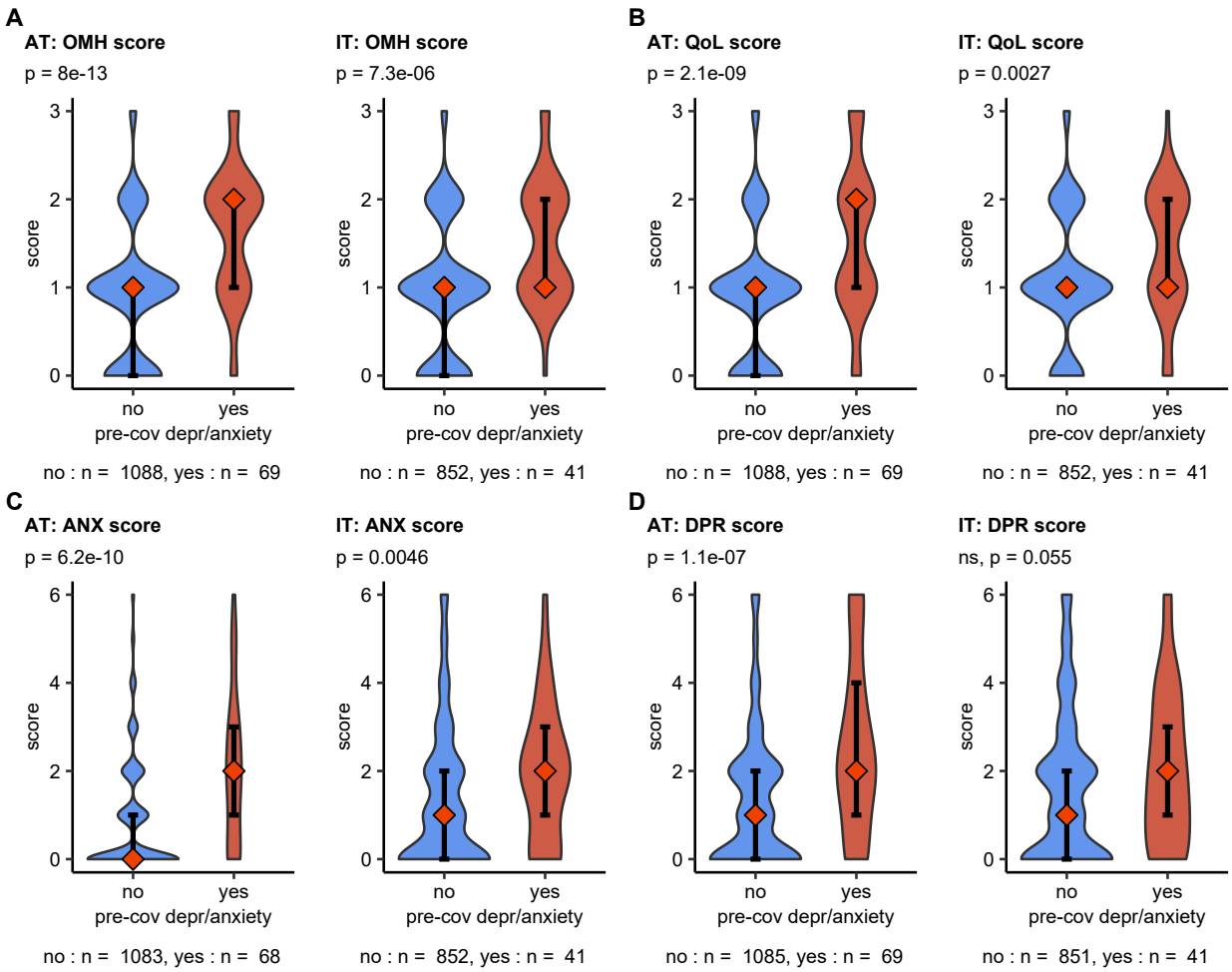

Figure S5: Depression or anxiety before COVID-19 and mental health scoring.

Supplementary Figure S5. Depression or anxiety before COVID-19 and mental health scoring.

Association of depression or anxiety before COVID-19 with overall mental health (OMH) (**A**), quality of life (QoL) (**B**), anxiety (ANX) (**C**) and depression (DPR) (**D**) scoring assessed by Mann-Whitney U test. The scoring is presented as violin plots, diamonds with whiskers represent medians with IQRs. P values corrected for multiple comparisons with Benjamini-Hochberg method are shown in plot sub-headings. N numbers of observations are indicated below the plot.

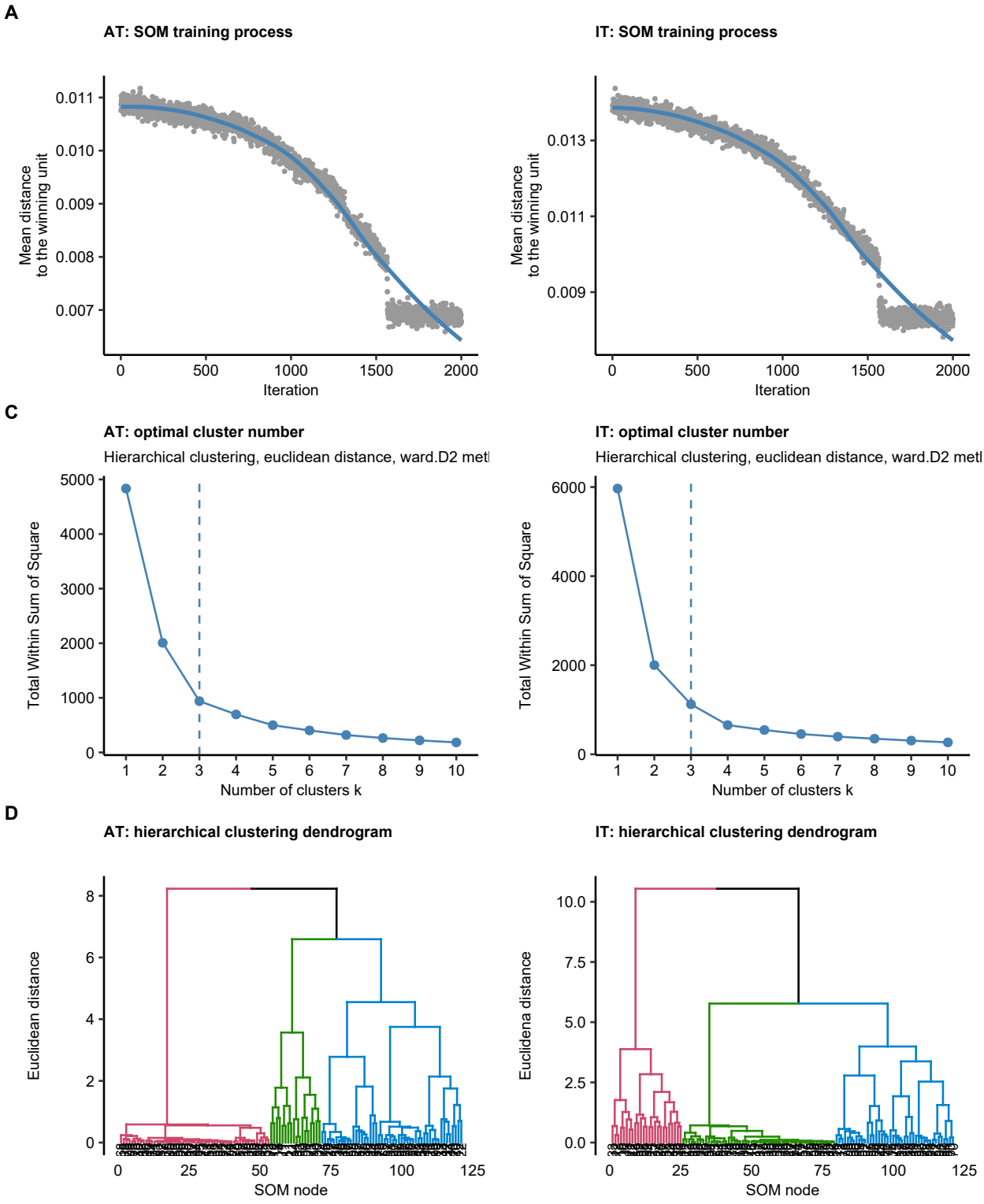

Figure S6: Development of the mental disorder risk clusters.

Supplementary Figure S6. Development of the mental disorder risk clusters.

Study participants were assigned to the Low Risk (LR), Intermediate Risk (IR) and High Risk (HR) subsets

by clustering analysis of the most influential factors impacting on the net mental health scoring (**Figure 2**) with the self-organizing map (SOM, 11 × 11 hexagonal grid, Jaccard distance between participants) and the hierarchical clustering (Ward D2 method, Euclidean distance between the SOM nodes) algorithms as presented in **Figure 4**.

(A) Progress of the SOM training procedure visualized as the drop of the mean distance to the winning unit with the algorithm iterations.

(B) Determination of the optimal cluster number in hierarchical clustering of the SOM nodes by finding the bend of the total within sum of square curve.

(B) Assignment of the SOM nodes to the clusters defined by hierarchical clustering presented in dendrograms.

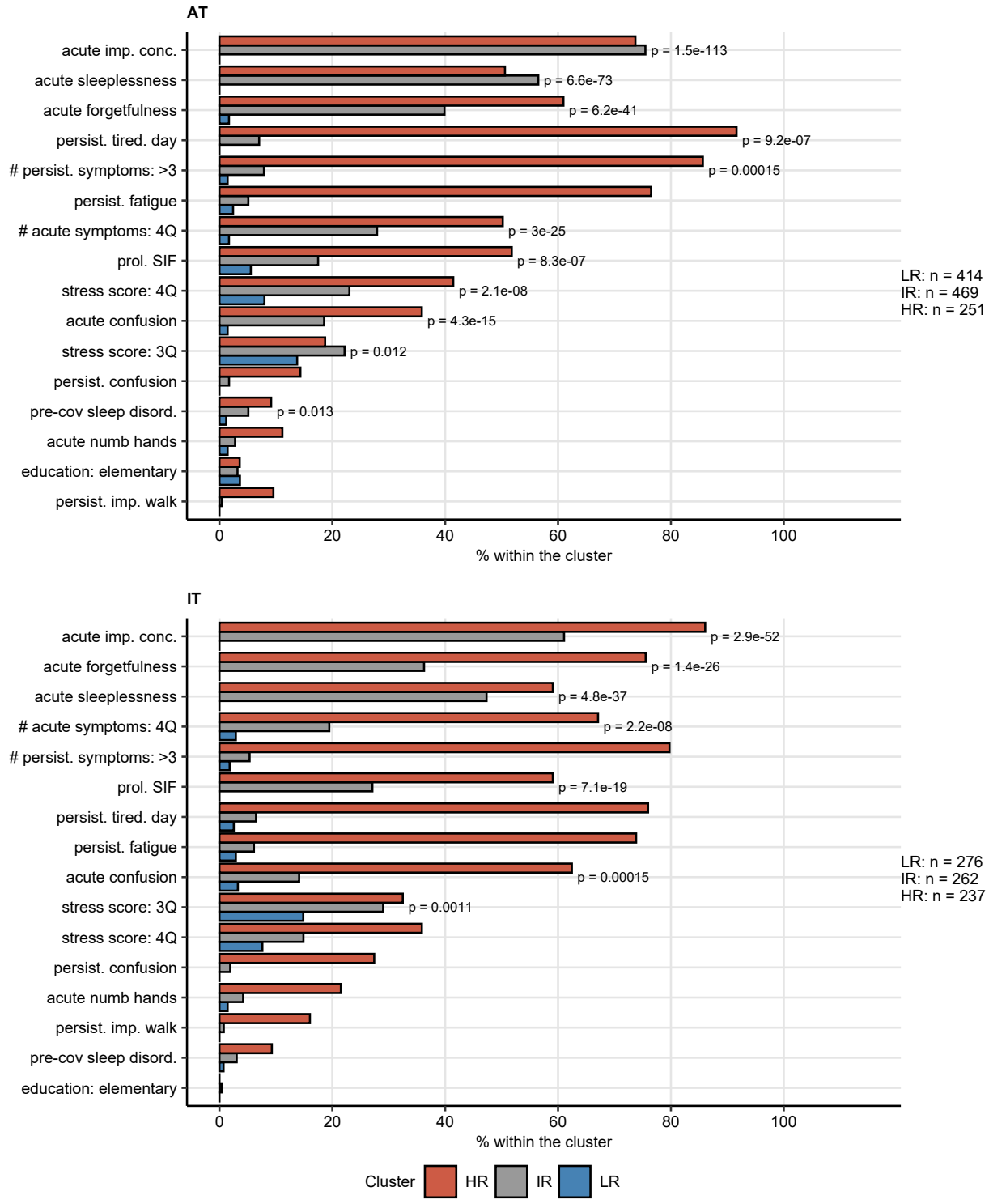

Figure S7: Frequency of the clustering features in the mental disorder risk clusters.

Supplementary Figure S7. Frequency of the clustering features in the mental disorder risk clusters.

Study participants were assigned to the Low Risk (LR), Intermediate Risk (IR) and High Risk (HR) subsets as presented in **Figure 4**. Differences in frequency of these features between the Low risk (LR), Intermediate Risk (IR) and High Risk (HR) clusters were assessed by χ^2 test. P values corrected for multiple comparisons with Benjamini-Hochberg method are presented for the significant comparisons. N numbers of individuals assigned to the clusters are presented next to the plots.

prol.: prolonged, SIF: severe illness feeling, imp.: impaired, conc.: concentration, #: number, tired.day.: tiredness at day, pre-cov sleep disord.: sleep disorder before COVID-19, 3Q, 4Q: 3^{rd} and 4^{th} quartile, persist.: persistent.

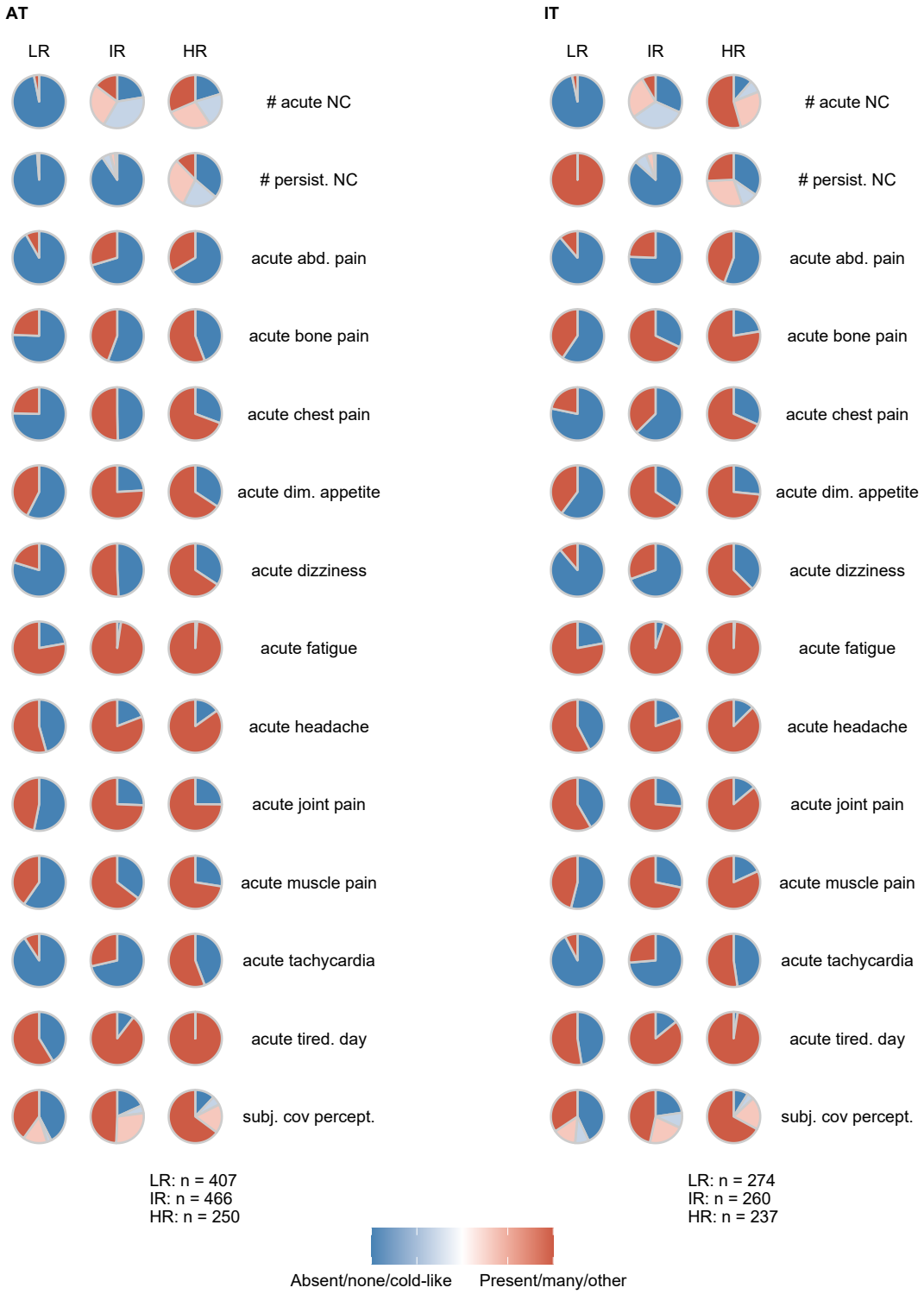

Figure S8: Frequency of the most significant differing features in the mental disorder risk clusters.

Supplementary Figure S8. Frequency of the most significant differing features in the mental disorder risk clusters.

Study participants were assigned to the Low Risk (LR), Intermediate Risk (IR) and High Risk (HR) subsets as presented in **Figure 4**. Differences in frequency of 129 survey variables not used for the cluster definition (**Figures 2** and **4**, **Supplementary Table S1**) between the risk clusters were compared by χ^2 test. P values were corrected for multiple comparisons with Benjamini-Hochberg method. Frequencies of the most significant features within the mental disorder risk clusters are shown. N numbers of individuals assigned to the clusters are presented next to the plots.

NC: neurocognitive symptoms, #: number, persist.: persistent, abd.: abdominal, dim.: diminished, subj. cov percept.: subjective perception of acute COVID-19.

References

1. Phenotyping of acute and persistent COVID-19 features in the outpatient setting: exploratory analysis of an international cross-sectional online survey. *medRxiv*. Published online August 2021:2021.08.05.21261677. doi:10.1101/2021.08.05.21261677

2. Löwe B, Wahl I, Rose M, et al. 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(1-2):86-95. doi:10.1016/j.jad.2009.06.019

3. Löwe B, Spitzer RL, Zipfel S, Herzog W. *Auflage Manual 17.07*. Vol 9.; 2002:44-45.

4. Beutel TF, Zwerenz R, Michal M. Psychosocial stress impairs health behavior in patients with mental disorders. *BMC Psychiatry*. 2018;18(1). doi:10.1186/s12888-018-1956-8

5. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. doi:10.21105/joss.01686

6. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. Springer-Verlag; 2016. https://ggplot2.tidyverse.org

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

8. 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(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

9. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

10. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software*. 2008;28(5):1-26. doi:10.18637/jss.v028.i05

11. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics*. 2008;9(1):1-11. doi:10.1186/1471-2105-9-307

12. Croux C, Filzmoser P, Oliveira MR. Algorithms for Projection-Pursuit robust principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 2007;87(2):218-225. doi:10.1016/j.chemolab.2007.01.004

13. Austrian Agency for Health and Food Safety (AGES). Epidemiologische Abklärung Covid 19. Accessed August 5, 2021. https://www.ages.at/themen/krankheitserreger/coronavirus/epidemiologische-abklaerung-covid-19/

14. Istituto Superiore di Sanitá (ISS). Integrated surveillance of COVID-19 in Italy. Accessed August 5, 2021. https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_7-luglio-2021.pdf

15. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*. 2010;1(2):97-111. doi:10.1002/jrsm.12

16. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*. 2019;22(4):153-160. doi:10.1136/ebmental-2019-300117

17. Vesanto J, Alhoniemi E. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*. 2000;11(3):586-600. doi:10.1109/72.846731

18. Kohonen T. *Self-Organizing Maps*. Vol 30. Springer Berlin Heidelberg; 1995. doi:10.1007/978-3-642-97610-0

19. Vesanto J, Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. Self-organizing map in Matlab: the SOM toolbox. *IN PROCEEDINGS OF THE MATLAB DSP CONFERENCE*. Published online 1999:35—-40. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.179

20. Health after COVID-19 in Tyrol. Mental Health after COVID-19 in Tyrol. Accessed September 9, 2021. https://im2-ibk.shinyapps.io/mental_health_dashboard/