

**Three distinct patterns of mental health response following
accidents in mountain sports – a follow-up study of individuals
treated at a tertiary trauma center**

Supplementary Material

Psychiatry Study Team

2023-08-31

Supplementary Methods

Software

Data analysis and transformation was accomplished with R version 4.2.3. The study dataset was imported from an SPSS file with the *foreign* package (1). Tabular data were handled with the *tidyverse* package bundle (2), and the packages *rlang* (3) and *trafo*. Text variables were handled with *stringi* (4).

For distribution testing (normality, variance, Gini index), the packages *ExDA*, *rstatix* (5) and *DescTools* (6) were utilized. Consistency of psychometric tools was assessed with the packages *psych* (7), *ExDA* and *clustTools*. Clustering tendency was investigated with *factoextra* (8) and *clustTools*. Statistical hypothesis testing and correlation analyses were accomplished with *ExDA* and *rstatix* (5).

For semi-supervised clustering, diagnostic and performance of the clustering analysis, the package *clustTools* with implementation of algorithms and distance measures from *factoextra*, *cluster* and *philentropy* (8–10) was utilized. UMAP (uniform manifold approximation and projection) was done with the *umap* and *clustTools* packages (11,12). Cross-distances between the training and test subsets of the study cohort and cross-distances between clusters in the training and test subsets of the study cohort were computed with *proxy* (13).

Multi-variable classifiers of cluster assignment were developed with the following algorithms: random forests (14,15), neural network (16), support vector machines with radial kernel (17,18), recursive partitioning (19,20), conditional random forests (21–23), shrinkage discriminant analysis (24,25), and elastic net multinomial regression (26,27). For tuning, training, prediction and assessment of performance of the classifiers, the packages *caret* (28) and *caretExtra* were used.

Results were visualized with *ggplot* (bar plots, box plots, heat maps of cross-distances, scatter plots) (29), *plotroc* (30), *ExDA* (violin, stack and ribbon plots) and *clustTools* (cluster quality control plots, heat maps of clustering features, distance heat maps, scatter plots of UMAP layouts) and *ComplexUpset* (visualization of overlap with upset plots) (31). Figures were created with the packages *cowplot* (32) and *patchwork* (33). Tables were generated with *flextable* (34). The manuscript and Supplementary Material were written in the *rmarkdown* environment (35) with the package *bookdown* (36). Figures, tables and R expressions in the markdown documents were managed with the development package *figur*. The markdown documents were rendered with the *knitr* (37) and *bookdown* (36) packages and with *author-info-blocks.lua* and *scholarly-metadata.lua* scripts by Albert Krewinkel and Jörn Krenzer.

Data import and transformation

The study data set was imported from an SPSS file with raw study data (function `read.spss()`, package *foreign*). The list of extracted variables with their description is available in **Supplementary Table S1**. A total of 307 participants with the complete set of psychometric battery variables (**Supplementary Table S2**) were included in the analysis (**Figure 1**).

The psychometric battery (**Supplementary Table S2**) consisted of German versions of assessment tools for anxiety (GAD-7: 7-item general anxiety disorder scale) (38), depression (PHQ: patient health questionnaire, PHQ-9) (39,40), panic (PHQ-panic module) (39,40), persistent somatic symptoms (PHQ-15) (41), resilience (RS13: 13-item resilience scale) (42), loss of sense of coherence (SOC-9L: Leipzig 9-item sense of coherence questionnaire) (43), quality of life (EUROHIS-QOL 8: 8-item EUROHIS project quality of life scale) (44), post-traumatic growth (PTGI: post-traumatic growth inventory) (45) and post-traumatic stress disorder (PCL-5 DSM-5: PTSD checklist for DSM-5) (46).

Clinically relevant symptoms of anxiety were defined as GAD-7 ≥ 10 , clinically relevant symptoms of depression were defined as PHQ-9 ≥ 11 (47), significant persistent somatic symptoms were defined as PHQ-15 ≥ 11 (41). Resilience classes were defined as follows: low: RS-13 0 - 65, moderate: 66 - 72, high: ≥ 73 (42).

Items of the EUROHIS QOL 8 tool were scored as with 1 - 5 likert scales (1: extreme concerns, no satisfaction at all, 5: no concerns/full satisfaction) with each item representing a single domain of quality of life (quality of life, health, energy, finances, activity, self-esteem, relationship and housing). The total EUROHIS QOL score was defined as the arithmetic mean of all items (44).

Separate scores were computed for each domain of the PTGI tool (I: relations, II: possibilities, III: personal strength, IV: spiritual strength, V: life appreciation) with each item scores as 0: none, 1: very little, 2: little, 3: moderate, 4: great, 5: extremely great. In addition, the total PTGI score was calculated as the sum of all items (45).

Separate scores were calculated for domains B, C, D and E of the PCL-5 DSM-5 tool along with the total score being the sum of all items. Each PCL-5 DSM-5 item was scored as 0: not at all, 1: a little bit, 2: moderate, 3: quite a bit and 4: extremely. Participants positive for the B domain or C domain PTSD symptoms were identified by at least one item per domain scored with 'moderate' or higher. Participants positive for the D or E domain PTSD symptoms were identified by at least two items per domain scored with 'moderate' or higher. Significant PTSD symptoms were assumed in participants screened positive for at least one of the B, C, D or E PCL-5 DSM-5 domains (46).

Traumatic events prior to the sport accident were assessed with the DIA-X tool (Diagnostic Expert System) (48). Direct personal experience or being a witness of a traumatic event

specified by the DIA-X questionnaire or by an additional yes/no item ('other traumatic events') was scored as 1. Prior traumatic event was assumed with at least one DIA-X item scored with 1. Prior sport accidents, flashbacks during mountain sport activity, confusion during mountain sport activity, self-reported need for psychological support, psychological/psychiatric support/therapy after the accident and presence of persistent physical health consequences after the accident were surveyed as single yes/no items. Flashbacks frequency during mountain sport activity was assessed in the following categories: none, more than one per year and more than one per month. Smoking was surveyed as a single yes/no question. Alcohol use was investigated with the CAGE tool with ≥ 2 points indicative of problematic alcohol consumption (49). Data on the type of the accident date and daytime, accident mountain sport, injury diagnosis, injured body regions, injury severity, hospital treatment, surgery and number of ICD-10 surgical diagnoses were extracted from electronic patient's records. Injury severity was assessed with the abbreviated injury scale (AIS) (50).

Additional information on study variables and their stratification schemes are presented in **Supplementary Table S1**.

Consistency of psychometric tools and power analysis

Consistency of psychometric tools was assessed by factor analysis and McDonald's ω (function `omega()`, package *psych*) (7,51). The number of latent factors for calculation of ω was identified by inspection of loadings determined by factor analysis (function `reduce_data()`, *clustTools*) (52). All psychometric tools used in the study except for the stress PSS-4 scale (53) and the BRCS resilience tool (54) exhibited good-to-excellent consistency with $\omega > 0.8$ (**Supplementary Table S3**). The poorly performing PSS-4 and BRCS scales were excluded from the further analysis.

To find the optimal size of a training subset of the study cohort, random subsamples of the study dataset of varying sizes were investigated (50, 100, 150, 200, 250, 300 observations, 50 random draws per subsample size). For each random subsample, clustering tendency was assessed by Hopkins statistic (function `get_clust_tendency()`, package *clustTools*). For $n = 250$ observations, the median Hopkins statistic was 0.73 (interquartile range: 0.73 - 0.74) indicative of good clustering tendency. The Hopkins statistic value for this subset was comparable with the value for the entire subset (0.74). For these reasons, 3/4 of the entire dataset or $n = 230$ was considered as the adequate size of the training subset of the study cohort for a reproducible clustering analysis.

Training/test subset definition

The study participants were assigned to the training and test subsets with the 3:1 size ratio (**Figure 1**). To this end, 100 random splits of the datasets were generated. The subset assignment scheme was chosen with the possibly smallest differences in

sociodemographic, medical history, clinical and accident- and injury-related variables assessed by Gower distance between the training and test subsets (function `dist()`, package *proxy*) (13).

Statistical hypothesis testing, effect size and multiple testing

correction

Differences in numeric variables were assessed with Mann-Whitney test with r effect size statistic or Kruskal-Wallis test with η^2 effect size statistic for two and more than two analysis groups, respectively. Differences in frequency of categories of qualitative variables between analysis groups were investigated with χ^2 test with Cramer's V effect size statistic. P values were corrected for multiple testing with the false discovery rate method separately for each analysis task (e.g. comparison of clusters) (55). Effects with $p < 0.05$ following the false discovery rate adjustment were considered significant. Effect size of accuracy of predicted cluster assignment by machine learning classifiers was assessed by Cohen's κ inter-rater reliability statistic (56,57). Intervals of effect sizes were defined as follows (57–59):

- r statistic: weak: < 0.3 , moderate: $0.3 - 0.5$, large: ≥ 0.5
- Cramer's V statistic: weak: < 0.3 , moderate: $0.3 - 0.5$, large: ≥ 0.5
- η^2 statistic: weak: < 0.06 , moderate: $0.06 - 0.14$, large: ≥ 0.14
- Cohen's κ reliability: none: < 0.2 , minimal: $0.2 - 0.4$, weak: $0.4 - 0.6$, moderate: $0.6 - 0.8$, strong: ≥ 0.8

Assessment of selection bias

To assess the possible selection bias, demographic, socioeconomic, clinical, accident- and recovery-related parameters were compared between individuals excluded due to missingness of psychometric data or denying survey participation and participants included in the analysis (**Figure 1**). Categorical variables were compared by χ^2 test with Cramer V effect size statistic and numeric variables were compared by Mann-Whitney test with r effect size statistic (function `compare_variables()`, package *ExDA*). Potential differences between the training and test subset of the study cohort (**Figure 1**) were analyzed in an analogous way. Significant ($p < 0.05$) differences between the included/excluded participants are presented in **Supplementary Table S4** and **S5**.

Semi-supervised clustering

Observations of the training subset of the study cohort were subjected to clustering in respect to the numeric psychometric scores (**Supplementary Table S2**). The score values were normalized and median-centered (function `center_data()`, package *clustTools*). For clustering, the PAM (partition around medoids) algorithm (9) with the cosine distance measure between observations (10) was employed. The clustering object was constructed with the function `kcluster()` from the *clustTools* package. The choice of the clustering algorithm was motivated by its good explanatory performance measured by the fraction of explained clustering variance (ratio of between-cluster sum of squares to total sum of squares, method `var()`, *clustTools*) and superior reproducibility in 10-fold cross-validation (60) (cluster assignment in the folds by an inverse distance weighted 27-nearest neighbors classifier, method `cv()`, *clustTools*) in a comparison with several other clustering algorithms presented in **Supplementary Figure S3A**. The number of cluster was chosen based on the bend of the curve of within-cluster sum of squares and the peak of mean silhouette statistic (method `plot()`, package *ExDA*) (8,61). By this means, three mental clusters were defined: the neutral, PTG (post-traumatic growth) and PTS (post-traumatic stress) cluster (**Supplementary Figure S3B**).

Assignment of the training subset observations to the mental clusters was accomplished with an inverse distance weighted 27-nearest neighbor classifier. This semi-supervised procedure yielded clustering structures with similar fractions of explained variance in the training ($V = 0.55$) and test subset ($V = 0.52$). Comparably good separation of the clusters could be discerned in the training and test subsets by a visual analysis of UMAP layouts and pairwise distance heat maps (**Supplementary Figure S4**). Distribution of the cluster frequency was similar in the training and test subset (**Supplementary Figure S5**). An analysis of cosine cross-distances between the clusters in the training and test subsets revealed far higher similarity of the corresponding mental clusters (i.e. neutral vs neutral, PTG vs PTG, PTS vs PTS) as compared with similarity of non-analogous clusters (e.g. neutral vs PTG) (distances computed with the function `dist()`, package *proxy*; **Supplementary Figure S5B**). Finally, quality of semi-supervised clustering was investigated by comparison of normalized levels of the clustering variables in the training and test subset by Kruskal-Wallis test with η^2 effect size statistic (function `compare_variables()`, package *ExDA*; **Figure 2, Supplementary Figure S6, Supplementary Table S7**).

Differences in frequencies of mental disorder symptoms as well as demographic, socioeconomic, clinical, accident- and recovery-related factors between the mental clusters were assessed in the entire cohort by χ^2 test with Cramer V effect size statistic and by Kruskal-Wallis test with η^2 effect size statistic for categorical and numeric variables, respectively (function `compare_variables()`, package *ExDA*; **Supplementary Table S8**).

Cluster assignment classifiers

Two types of multi-parameter machine learning classifiers of the mental cluster assignment were developed in the training subset:

1. models employing candidate early predictors of cluster assignment, i.e. demographic, socioeconomic, medical history and accident-related explanatory factors available during acute medical management of the accident
2. models including additionally recovery-related predictors such as persistent physical health consequences, flashbacks or cautious behavior during sport

Of note, psychometric variables used for definition of the mental clusters as well as mental disorder symptoms, presence and frequency of flashbacks were excluded from the explanatory variable sets. The explanatory variables are listed in **Supplementary Table S9**.

The models employed the following algorithms: canonical random forests (14,15), regularized neural networks with a single hidden layer (16), support vector machines with radial kernel (17,18), recursive partitioning (19,20), shrinkage discriminant analysis (24,25), conditional random forest (21–23), and elastic net multinomial regression (26,27). The optimal values of the algorithms' parameters were found by 10-fold cross-validation-based tuning with the maximal value of Cohen's κ (56) as the tuning criterion. The tuning and fitting ('training') in the training subset of the study cohort was done with the wrapper function `train()` provided by the package *caret* (28). For the random forest and conditional random forest algorithms, 1000 random trees each were constructed. Test statistic, p values, number of splits and other parameters of the conditional forest models were controlled with the `ForestControl` object returned by the convenience wrapper `cforest_unbiased()` provided by the *party* package (23). The optimal algorithm parameter sets are listed in **Supplementary Table S10**. Predictions of the cluster assignment in the test subset of the study cohort were obtained with the `predict()` method from the *caretExtra* package.

Overall accuracy and Cohen's κ statistics (56) were computed with the `summary()` method from the *caretExtra* package. Brier scores (62) were computed with the following formula:

$$BS = \frac{1}{N} \times \sum_{i=1}^N \sum_{c=1}^R (f_{ic} - o_{ic})^2$$

where f_{ic} is the output probability of assignment of the i-th observation to the c-th cluster, o_{ic} is the numeric-coded actual assignment of the i-th observation to the c-th cluster, R is the total cluster number and N is the total observation number. Brier skill scores (62) were computed with the following formula:

$$BSS = 1 - \frac{BS_{model}}{BS_{random}}$$

where BS_{model} is the Brier score of the model and BS_{random} is the Brier score obtained for a purely random cluster assignment. Brier scores, Brier skill scores, sensitivity and specificity of the PTS cluster assignment were computed with in-house developed R functions. Performance statistics for the training subset, 10-fold cross-validation and test subset of the study cohort are listed in **Supplementary Table S11** and **Supplementary Table S12** for the classifiers employing the early and full predictor set, respectively.

Variable importance statistic specific for the machine learning algorithm were extracted from the *caret* models with the `varImp()` function from package *caret* (28). They were: permutation importance for the random forest and conditional forest algorithm (14,23), connection weight importance statistic for the neural network (63), sum reduction in classification error attributed to each variable at each split for recursive partitioning (20), and linear model coefficient β for elastic net regression (26). For support vector machines and discriminant analysis, the variable importance was computed based on area under the ROC curve for single predictors for discrimination between the cluster pairs (28).

Data and code availability

An RData 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/PiotrTymoszek/mental_accident.

Supplementary Tables

Supplementary Table S1: Variables used in the analysis pipeline.

Supplementary Table S2: Mental health assessment battery.

Section	Variable ^a	Description ^a
PTSD assessment	PCL-5 score	PTSD total score, PCL-5, sum of all items
	PTSD+ (at least one PCL-5 domain)	At least one PCL-5 domain positive
	PCL-5 domain B score	PTSD rating, PCL-5 domain B
	PTSD domain B symptoms	PTSD symptoms, PCL-5 domain B positive
	PCL-5 domain C score	PTSD rating, PCL-5 domain C
	PTSD domain C symptoms	PTSD symptoms, PCL-5 domain C positive
	PCL-5 domain D score	PTSD rating, PCL-5 domain D
	PTSD domain D symptoms	PTSD symptoms, PCL-5 domain D positive
	PCL-5 domain E score	PTSD rating, PCL-5 domain E
	PTSD domain E symptoms	PTSD symptoms, PCL-5 domain E positive
PTG assessment	PTGI score	Post-traumatic growth, PTGI total score, sum of all items
	PTGI I relations score	Post-traumatic growth, PTGI scoring, domain I, relations
	PTGI II possibilities score	Post-traumatic growth, PTGI scoring, domain II, new possibilities
	PTGI III personal strength score	Post-traumatic growth, PTGI scoring, domain III, personal strength
	PTGI IV spiritual score	Post-traumatic growth, PTGI scoring, domain IV, spiritual
	PTGI V life appreciation score	Post-traumatic growth, PTGI scoring, domain V, appreciation of life
Mental health, resilience, coherence	RS13 score	Resilience, RS13 score
	RS13 resilience class	Resilience, RS13 class
	SOC-9L score	Lack of sense of coherence, SOC-9L score
	PHQ-9 score	PHQ-9 score, depression
	Depression symptoms (PHQ-9 ≥ 11)	PHQ-9 score, depression symptoms
	GAD-7 score	GAD-7 score, anxiety
	Anxiety symptoms (GAD-7 ≥ 10)	GAD-7 score, anxiety symptoms
	PHQ-panic score	PHQ panic 4 item score

Section	Variable ^a	Description ^a
	Panic symptoms (PHQ-panic)	PHQ panic positivity
	PHQ-15 score	PHQ-15 health problems, somatic symptoms
	Somatic symptoms (PHQ-15 ≥11)	PHQ-15 health problems, clinically relevant somatic symptoms
Quality of life	EUROHIS-QOL 8 mean score	Quality of life, EUROHIS-QOL 8 score, mean of all items
	EUROHIS-QOL 8 QoL score	Quality of life, EUROHIS-QOL 8 score QoL
	EUROHIS-QOL 8 health score	Quality of life, EUROHIS-QOL 8 score health
	EUROHIS-QOL 8 energy score	Quality of life, EUROHIS-QOL 8 score energy
	EUROHIS-QOL 8 finances score	Quality of life, EUROHIS-QOL 8 score financial aspects
	EUROHIS-QOL 8 activity score	Quality of life, EUROHIS-QOL 8 score activity
	EUROHIS-QOL 8 self-esteem score	Quality of life, EUROHIS-QOL 8 score self-esteem
	EUROHIS-QOL 8 relationship score	Quality of life, EUROHIS-QOL 8 score relationship
	EUROHIS-QOL 8 housing score	Quality of life, EUROHIS-QOL 8 score housing
^a GAD-7: 7-item general anxiety disorder scale; PHQ: patient health questionnaire; EUROHIS-QOL 8: 8-item EUROHIS project quality of life scale; SOC-9L: Leipzig 9-item sense of coherence questionnaire; RS13: 13-item resilience scale; PCL-5 DSM-5: PTSD checklist for DSM-5; PTGI: post-traumatic growth inventory		

Supplementary Table S3: Consistency of the psychometric tools used in the study measured by McDonald's omega.

Scale ^a	Number of latent factors	Total omega
PTGI	4	0.97
RS13	3	0.94
PCL-5	4	0.92
SOC-9L	3	0.89
GAD-7	3	0.89
PHQ-panic	1	0.88
EUROHIS-QOL 8 mean	4	0.88
PHQ-9	4	0.87
PHQ-15	4	0.84
^a GAD-7: 7-item general anxiety disorder scale; PHQ: patient health questionnaire; EUROHIS-QOL 8: 8-item EUROHIS project quality of life scale; SOC-9L: Leipzig 9-item sense of coherence questionnaire; RS13: 13-item resilience scale; PCL-5 DSM-5: PTSD checklist for DSM-5; PTGI: post-traumatic growth inventory		

Supplementary Table S4: Significant differences between patients who did not respond to the study invitation and the analyzed study participants. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are presented as percentages and counts within the complete observation set.

Variable ^a	Included	No response	Significance ^b	Effect size ^b
sport type	ski/snowboard/cross-country: 64% (n = 197) sledding: 3.9% (n = 12) climbing/hiking/mountaineering/ski touring: 14% (n = 42) biking: 16% (n = 48) other: 2.6% (n = 8) n = 307	ski/snowboard/cross-country: 52% (n = 2179) sledding: 5% (n = 207) climbing/hiking/mountaineering/ski touring: 11% (n = 457) biking: 27% (n = 1127) other: 4.8% (n = 202) n = 4172	p < 0.001	V = 0.078
injury severity class, AIS	1: 37% (n = 108) 2: 35% (n = 103) 3+: 28% (n = 83) 0: 0% (n = 0) n = 294	1: 46% (n = 124) 2: 35% (n = 93) 3+: 17% (n = 46) 0: 1.5% (n = 4) n = 267	p = 0.0019	V = 0.16
injury severity, AIS	2 [IQR: 1 - 3] range: 1 - 5 n = 294	2 [IQR: 1 - 2] range: 0 - 4 n = 267	p < 0.001	r = 0.15
hospitalized	26% (n = 80) n = 307	9.4% (n = 393) n = 4172	p < 0.001	V = 0.14
surgery	14% (n = 43) n = 307	4.2% (n = 175) n = 4172	p < 0.001	V = 0.12
number of surgical ICD-10 diagnoses	none: 86% (n = 264) 1: 8.5% (n = 26) 2+: 5.5% (n = 17) n = 307	none: 96% (n = 3997) 1: 2.8% (n = 116) 2+: 1.4% (n = 59) n = 4172	p < 0.001	V = 0.12

^aAIS: abbreviated injury scale

^bNumeric variables: Mann-Whitney test with r effect size statistic; categorical variables: χ^2 test with Cramer V effect size statistic. P values corrected for multiple testing with the false discovery rate method.

Supplementary Table S5: Significant differences between the study survey responders excluded from analysis due to missingness of psychometric data and the analyzed study participants. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are presented as percentages and counts within the complete observation set.

Variable ^a	Included	Incomplete variables	Significance ^b	Effect size ^b
income/year	none: 21% (n = 63) < 30K EUR: 18% (n = 56) 30K - 45K EUR: 19% (n = 59) ≥ 45K EUR: 42% (n = 129) n = 307	none: 44% (n = 35) < 30K EUR: 14% (n = 11) 30K - 45K EUR: 16% (n = 13) ≥ 45K EUR: 26% (n = 21) n = 80	p < 0.001	V = 0.22
injury severity, AIS	2 [IQR: 1 - 3] range: 1 - 5 n = 294	2 [IQR: 1 - 2] range: 1 - 4 n = 70	p = 0.031	r = 0.11
upper limb injury	41% (n = 120) n = 294	57% (n = 40) n = 70	p = 0.019	V = 0.13
hospitalized	26% (n = 80) n = 307	14% (n = 11) n = 80	p = 0.03	V = 0.12
surgery	14% (n = 43) n = 307	5% (n = 4) n = 80	p = 0.045	V = 0.11
physical health consequences	37% (n = 115) n = 307	22% (n = 12) n = 55	p = 0.037	V = 0.12

^aAIS: abbreviated injury scale; K: 1000 Euro; EUR: Euro.

^bNumeric variables: Mann-Whitney test with r effect size statistic; categorical variables: χ^2 test with Cramer V effect size statistic. P values corrected for multiple testing with the false discovery rate method.

Supplementary Table S6: Significant differences between the training and test subset of the study cohort. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are presented as percentages and counts within the complete observation set.

Variable ^a	Training	Test	Significance ^b	Effect size ^b
RS13 resilience class	low: 23% (n = 53) moderate: 13% (n = 29) high: 64% (n = 148) n = 230	low: 3.9% (n = 3) moderate: 17% (n = 13) high: 79% (n = 61) n = 77	p = 0.035	V = 0.22
EUROHIS-QOL 8 mean score	4.2 [IQR: 3.9 - 4.6] range: 2 - 5 n = 230	4.5 [IQR: 4.2 - 4.6] range: 2.8 - 5 n = 77	p = 0.049	r = 0.18
EUROHIS-QOL 8 QoL score	5 [IQR: 4 - 5] range: 3 - 5 n = 230	5 [IQR: 5 - 5] range: 3 - 5 n = 77	p = 0.0011	r = 0.25

^aRS13: 13-item resilience scale; EUROHIS-QOL 8: 8-item EUROHIS project quality of life scale.

^bNumeric variables: Mann-Whitney test with r effect size statistic; categorical variables: χ^2 test with Cramer V effect size statistic. P values corrected for multiple testing with the false discovery rate method.

Supplementary Table S7: Differences in psychometric clustering factors between the mental clusters. Numeric variables are presented as medians with interquartile ranges (IQR). Statistical significance was determined by false discovery rate-corrected Kruskal-Wallis test with eta-square effect size statistic. The table is available in a supplementary Excel file.

Supplementary Table S8: Significant differences in demographic, socioeconomic, clinical, accident- and recovery-related factors, and mental disorder symptoms between the mental clusters in the entire cohort. Numeric variables are presented as medians with interquartile ranges (IQR). Categorical variables are presented as percentages and counts within the complete observation set.

Variable ^a	Neutral cluster	PTG cluster	PTS cluster	Significance ^b	Effect size ^b
Participants, n	103	94	110		
age, years	53 [IQR: 34 - 61] range: 18 - 82	54 [IQR: 39 - 61] range: 18 - 81	45 [IQR: 29 - 56] range: 18 - 82	p = 0.023	$\eta^2 = 0.027$
pre-existing physical illness	5.8% (n = 6)	22% (n = 21)	18% (n = 20)	p = 0.013	V = 0.19
pre-existing mental disorder	0% (n = 0)	1.1% (n = 1)	14% (n = 15)	p < 0.001	V = 0.28
psychological support need	0% (n = 0)	8.5% (n = 8)	14% (n = 15)	p = 0.0039	V = 0.22
physical health consequences	27% (n = 28)	32% (n = 30)	52% (n = 57)	p = 0.0025	V = 0.23
caution post accident	no change: 50% (n = 52) more cautious: 49% (n = 50) less cautious: 0.97% (n = 1)	no change: 32% (n = 30) more cautious: 67% (n = 63) less cautious: 1.1% (n = 1)	no change: 22% (n = 24) more cautious: 78% (n = 86) less cautious: 0% (n = 0)	p = 0.0018	V = 0.19
flashbacks during sport	24% (n = 25)	35% (n = 33)	58% (n = 64)	p < 0.001	V = 0.3
flashback frequency	none: 76% (n = 78) > 1/year: 17% (n = 18) > 1/month: 6.8% (n = 7)	none: 65% (n = 61) > 1/year: 20% (n = 19) > 1/month: 15% (n = 14)	none: 42% (n = 46) > 1/year: 28% (n = 31) > 1/month: 30% (n = 33)	p < 0.001	V = 0.22
PTSD+ (at least one PCL-5 domain)	4.9% (n = 5)	15% (n = 14)	35% (n = 39)	p < 0.001	V = 0.33
PTSD domain B symptoms	1.9% (n = 2)	9.6% (n = 9)	20% (n = 22)	p < 0.001	V = 0.24
PTSD domain C symptoms	2.9% (n = 3)	5.3% (n = 5)	15% (n = 17)	p = 0.0083	V = 0.2
PTSD domain D symptoms	0.97% (n = 1)	3.2% (n = 3)	11% (n = 12)	p = 0.012	V = 0.2
PTSD domain E symptoms	0% (n = 0)	5.3% (n = 5)	17% (n = 19)	p < 0.001	V = 0.27

Variable ^a	Neutral cluster	PTG cluster	PTS cluster	Significance ^b	Effect size ^b
RS13 resilience class	low: 3.9% (n = 4) moderate: 5.8% (n = 6) high: 90% (n = 93)	low: 6.4% (n = 6) moderate: 5.3% (n = 5) high: 88% (n = 83)	low: 42% (n = 46) moderate: 28% (n = 31) high: 30% (n = 33)	p < 0.001	V = 0.43
Depression symptoms (PHQ-9 ≥11)	0% (n = 0)	2.1% (n = 2)	14% (n = 15)	p < 0.001	V = 0.27
Anxiety symptoms (GAD-7 ≥10)	0% (n = 0)	0% (n = 0)	6.4% (n = 7)	p = 0.0082	V = 0.2
Somatic symptoms (PHQ-15 ≥11)	0.97% (n = 1)	3.2% (n = 3)	10% (n = 11)	p = 0.023	V = 0.18

^aAIS: abbreviated injury score; PTSD: post-traumatic stress disorder; RS13: 13-item resilience scale; GAD-7: 7-item general anxiety disorder scale; PHQ: patient health questionnaire.

^bNumeric variables: Kruskal-Wallis test with η^2 effect size statistic. Categorical variables: χ^2 test with Cramer V effect size statistic. P values were corrected for multiple testing with the false discovery rate method.

Supplementary Table S9: Sets of explanatory factors used for modeling of the mental cluster assignment.

Classifier type ^a	Explanatory variables
early predictor model	age, age class, accident season, accident daytime, sex, education, employment, sport profession, trauma-risk profession, healthcare profession, income/year, residence in the Alps, smoking, pre-existing physical illness, pre-existing physical illness type, pre-existing mental disorder, number of prior traumatic events/DIA-X, problematic alcohol use (CAGE ≥ 2), prior sport accidents, sport type, alone during the accident, responsible for the accident, injured persons, rescue, rescue mode, injury severity class, injury severity, head injury, face injury, neck injury, chest injury, abdomen injury, spine region injury, upper limb injury, lower limb injury, other injury, number of injured body parts, hospitalized, surgery, number of surgical ICD-10 diagnoses, psychological support, psychological support need, physical health consequences, returned to same sport, caution post accident, confusion during sport
full set predictor model	age, age class, accident season, accident daytime, sex, education, employment, sport profession, trauma-risk profession, healthcare profession, income/year, residence in the Alps, smoking, pre-existing physical illness, pre-existing physical illness type, pre-existing mental disorder, number of prior traumatic events/DIA-X, problematic alcohol use (CAGE ≥ 2), prior sport accidents, sport type, alone during the accident, responsible for the accident, injured persons, rescue, rescue mode, injury severity class, injury severity, head injury, face injury, neck injury, chest injury, abdomen injury, spine region injury, upper limb injury, lower limb injury, other injury, number of injured body parts, hospitalized, surgery, number of surgical ICD-10 diagnoses
^a early predictors: variables available during acute medical management of the accident; full predictor set: variables available during acute medical management of the accident and during follow-up.	

Supplementary Table S10: The optimal combinations of machine learning algorithm parameters found in 10-fold cross-validation of the training subset of the study cohort.

Classifier type	Algorithm^a	Tuning parameters
full	RF	mtry = 2, splitrule = gini, min.node.size = 3
full	NNet	size = 2, decay = 0.01
full	SVM/radial	sigma = 0.02, C = 1.3
full	RPart	cp = 0.0325
full	SDA	diagonal = FALSE, lambda = 0.195
full	cForest	mtry = 4
full	ElasticNet	alpha = 0.5, lambda = 0.01
early	RF	mtry = 7, splitrule = gini, min.node.size = 1
early	NNet	size = 9, decay = 0.1
early	SVM/radial	sigma = 0.09, C = 3.1
early	RPart	cp = 0.0125
early	SDA	diagonal = FALSE, lambda = 0.25
early	cForest	mtry = 26
early	ElasticNet	alpha = 0.5, lambda = 0.0158
^a RF: random forest; NNet: neural network with a single hidden layer; SVM/radial: support vector machines with radial kernel; RPart: recursive partitioning; SDA: shrinkage discriminant analysis; cForest: conditional random forest; Elastic Net: elastic net multinomial regression.		

Supplementary Table S11: Performance statistics of machine learning classifiers at predicting the mental cluster assignment. Models employing early predictors available during acute medical management of the accident.

Algorithm ^a	Data subset ^b	Accuracy	Cohen's κ	Brier score	Brier skill score ^c	Sensitivity, PTS cluster	Specificity, PTS cluster
RF	training	1.00	0.990	0.150	0.88	1.00	1.00
	10-fold CV	0.41	0.110	0.670	0.48	0.42	0.68
	test	0.47	0.210	0.660	0.52	0.48	0.68
NNet	training	1.00	1.000	0.018	0.99	1.00	1.00
	10-fold CV	0.45	0.170	0.880	0.34	0.35	0.69
	test	0.38	0.078	0.960	0.32	0.33	0.72
SVM/radial	training	0.93	0.900	0.570	0.52	0.88	0.99
	10-fold CV	0.42	0.120	0.670	0.48	0.56	0.59
	test	0.47	0.210	0.660	0.55	0.56	0.64
RPart	training	0.64	0.450	0.500	0.61	0.62	0.84
	10-fold CV	0.38	0.070	0.770	0.44	0.38	0.76
	test	0.34	0.018	0.780	0.41	0.33	0.70
SDA	training	0.68	0.520	0.410	0.70	0.68	0.87
	10-fold CV	0.41	0.120	0.800	0.42	0.38	0.68
	test	0.39	0.094	0.740	0.41	0.33	0.68
cForest	training	0.70	0.550	0.550	0.57	0.75	0.84
	10-fold CV	0.38	0.058	0.660	0.52	0.40	0.67
	test	0.46	0.200	0.650	0.44	0.48	0.74
ElasticNet	training	0.72	0.580	0.410	0.70	0.69	0.86
	10-fold CV	0.42	0.130	0.760	0.40	0.38	0.66
	test	0.38	0.069	0.740	0.52	0.41	0.62

^aRF: random forest; NNet: neural network with a single hidden layer; SVM/radial: support vector machines with radial kernel; RPart: recursive partitioning; SDA: shrinkage discriminant analysis; cForest: conditional random forest; Elastic Net: elastic net multinomial

Algorithm ^a	Data subset ^b	Accuracy	Cohen's κ	Brier score	Brier skill score ^c	Sensitivity, PTS cluster	Specificity, PTS cluster
regression.							
^b CV: cross-validation							
^c Brier Skill Score comparing the given classifier with the purely random cluster assignment.							

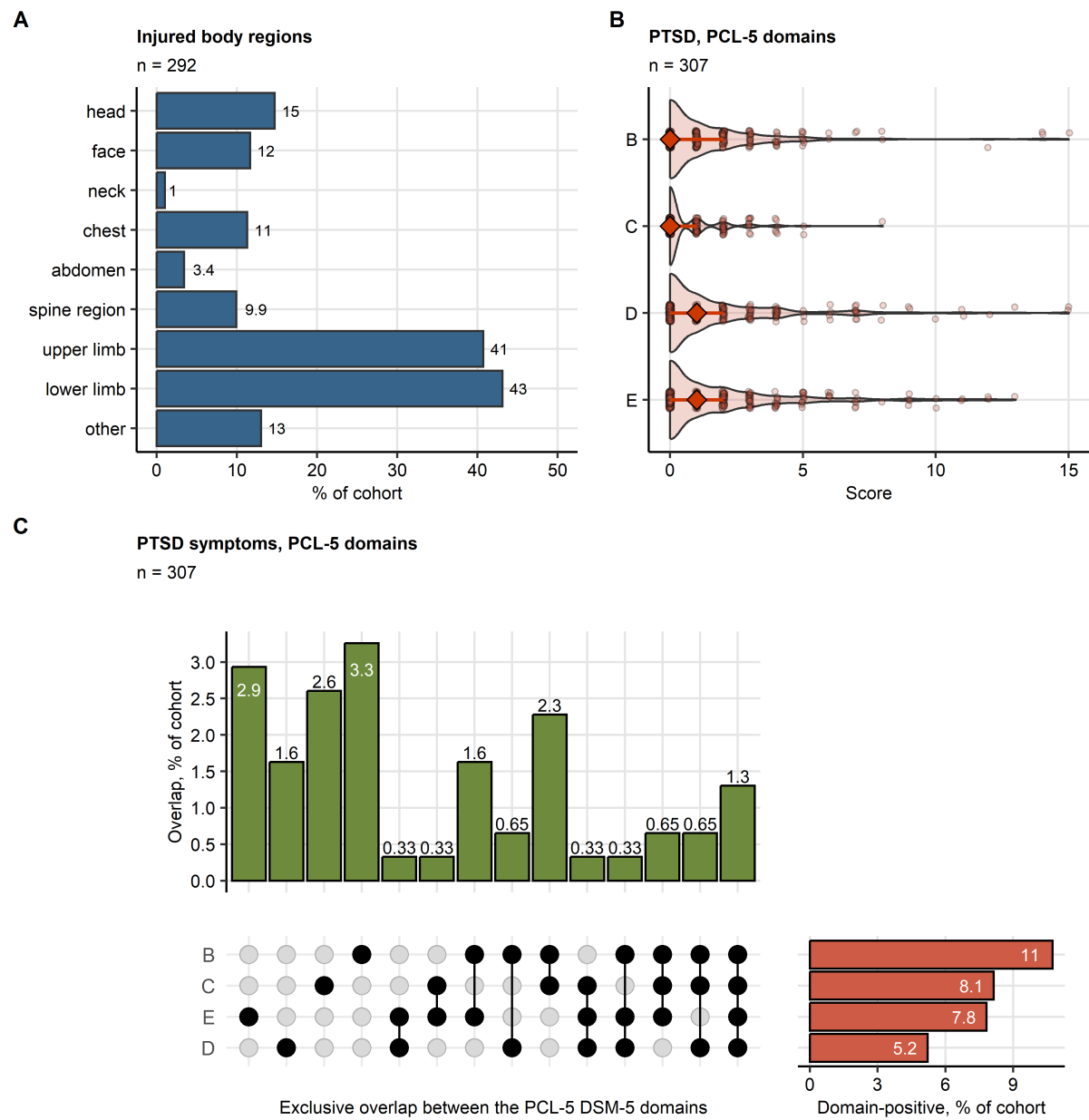
Supplementary Table S12: Performance statistics of machine learning classifiers at predicting the mental cluster assignment. Models employing the full predictor set available during acute medical management of the accident and follow-up.

Algorithm ^a	Data subset ^b	Accuracy	Cohen's κ	Brier score	Brier skill score ^c	Sensitivity, PTS cluster	Specificity, PTS cluster
RF	training	0.92	0.890	0.42	0.68	0.95	0.98
	10-fold CV	0.43	0.140	0.65	0.50	0.48	0.68
	test	0.46	0.200	0.64	0.50	0.52	0.70
NNet	training	0.90	0.850	0.18	0.87	0.91	1.00
	10-fold CV	0.48	0.230	0.90	0.28	0.48	0.75
	test	0.42	0.150	0.92	0.28	0.30	0.85
SVM/radial	training	0.66	0.490	0.52	0.59	0.75	0.80
	10-fold CV	0.47	0.200	0.63	0.50	0.59	0.68
	test	0.45	0.170	0.64	0.54	0.48	0.70
RPart	training	0.54	0.320	0.57	0.55	0.55	0.79
	10-fold CV	0.38	0.059	0.70	0.46	0.39	0.66
	test	0.36	0.049	0.70	0.36	0.41	0.72
SDA	training	0.76	0.640	0.34	0.74	0.71	0.88
	10-fold CV	0.48	0.220	0.79	0.42	0.42	0.77
	test	0.43	0.160	0.76	0.39	0.33	0.81
cForest	training	0.68	0.520	0.59	0.55	0.82	0.79
	10-fold CV	0.44	0.150	0.65	0.54	0.56	0.62
	test	0.39	0.098	0.65	0.49	0.48	0.68
ElasticNet	training	0.81	0.720	0.31	0.76	0.79	0.90
	10-fold CV	0.48	0.220	0.77	0.40	0.48	0.69
	test	0.39	0.100	0.81	0.42	0.30	0.74

^aRF: random forest; NNet: neural network with a single hidden layer; SVM/radial: support vector machines with radial kernel; RPart: recursive partitioning; SDA: shrinkage discriminant analysis; cForest: conditional random forest; Elastic Net: elastic net multinomial

Algorithm ^a	Data subset ^b	Accuracy	Cohen's κ	Brier score	Brier skill score ^c	Sensitivity, PTS cluster	Specificity, PTS cluster
regression.							
^b CV: cross-validation							
^c Brier Skill Score comparing the given classifier with the purely random cluster assignment.							

Supplementary Figures

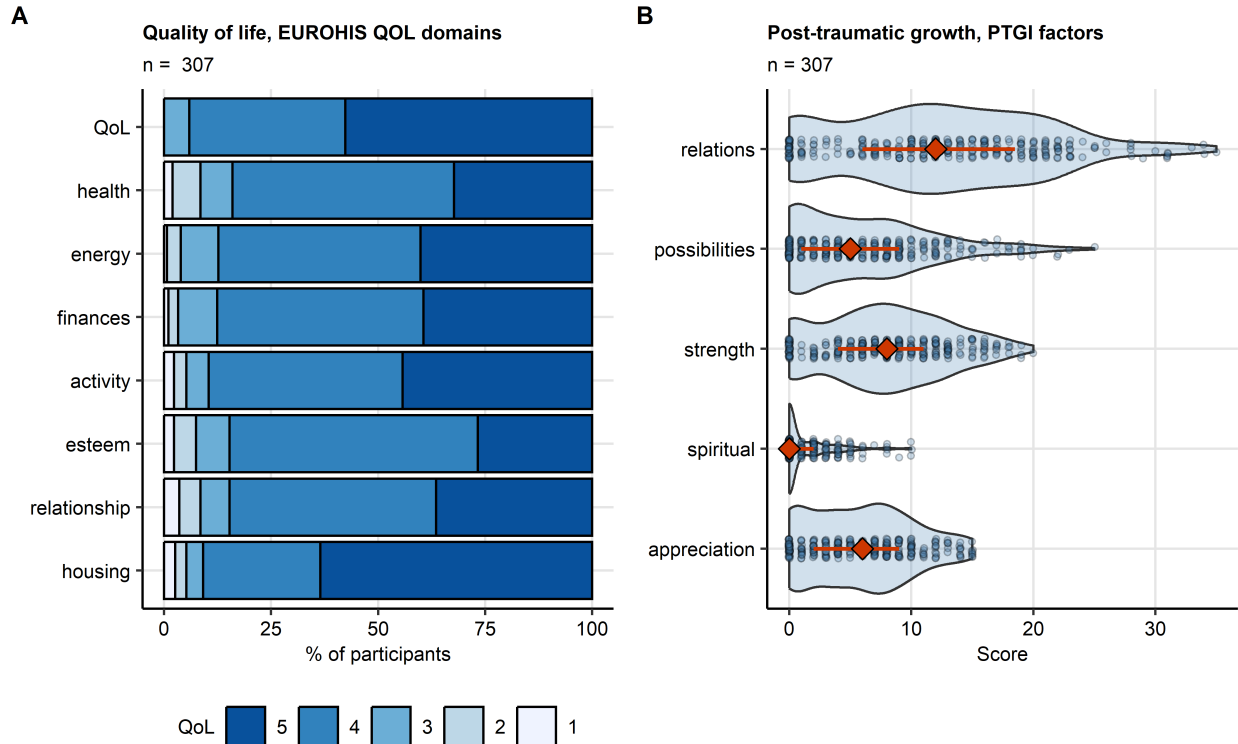


Supplementary Figure S1. Injured body regions and symptoms of PTSD in the study cohort.

(A) Distribution of injured body regions presented in a bar plot. The number of complete observations is indicated in the plot caption.

(B) Scores of particular domains of the PCL-5 tool evaluating symptoms of post-traumatic stress disorder (PTSD) presented in violin plots. Single observations are visualized as points. Medians and interquartile ranges are represented by red diamonds and whiskers. The number of complete observations is indicated in the plot caption.

(C) Frequency and overlap of clinically relevant PTSD symptoms captured by particular domains of the PCL-5 scale presented in an upset plot. Percentages of participants screened positive for the domain B, C, D or E PTSD symptoms are displayed in as red bars. Percentages of participants with single and overlapping domain B, C, D or E PTSD symptoms are presented as green bars. The number of complete observations is indicated in the plot caption.

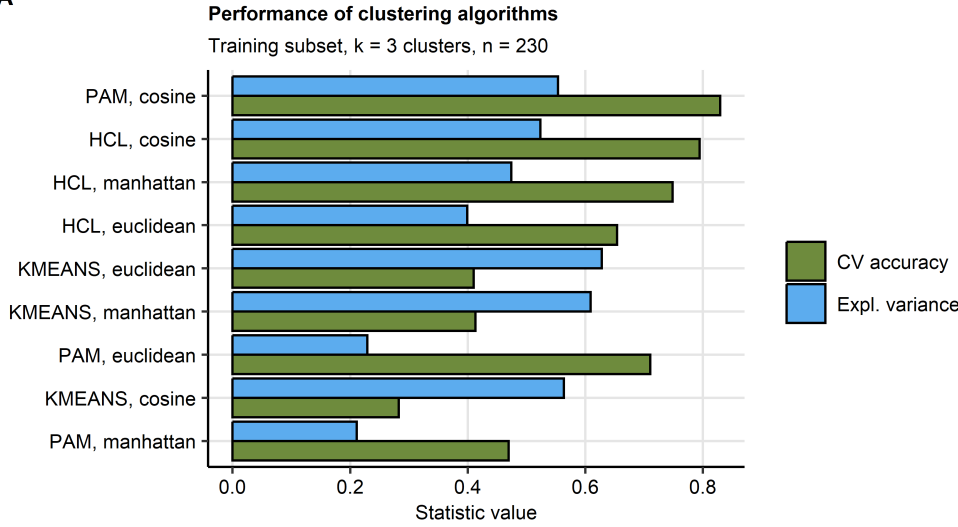


Supplementary Figure S2. Scores of quality of life and post-traumatic growth in the study cohort.

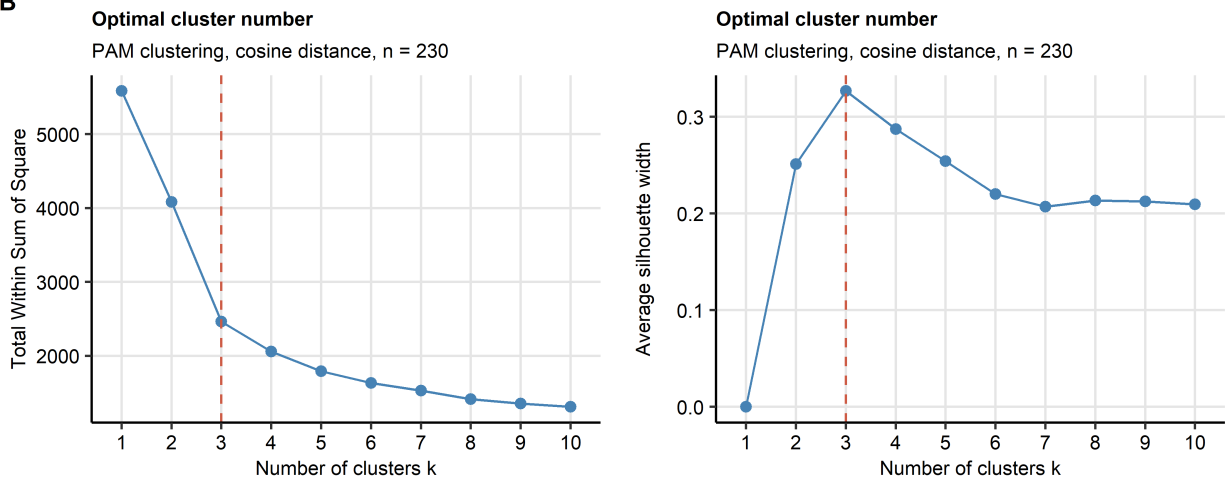
(A) Scores of the domains of the EUROHIS project 8-item quality of life scale (EUROHIS-QOL 8) presented in a bar plot. The number of complete observations is displayed in the plot caption.

(B) Scores of the domains of the post-traumatic growth inventory (PTGI) scale presented in a violin plots. Single observations are visualized as points. Red diamonds with whiskers represent medians with interquartile ranges. The number of complete observations is indicated in the plot caption.

A



B

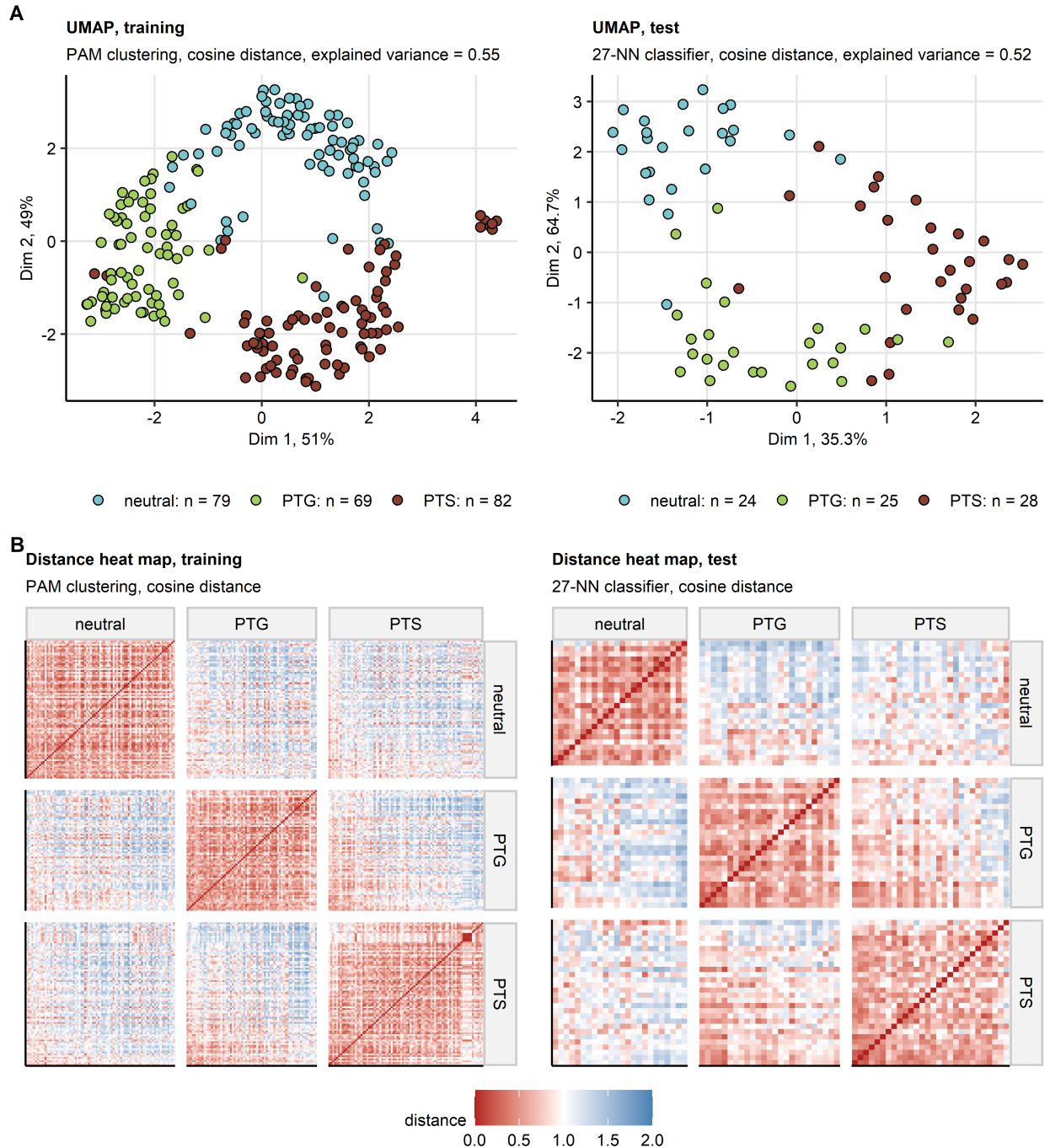


Supplementary Figure S3. Definition of the mental clusters in the training subset of the study cohort.

The mental clusters were defined in respect to psychometric scoring in the training subset of the study cohort by PAM (partition around medoids) with cosine distance between the observations.

(A) Comparison of explanatory performance and reproducibility of various clustering algorithms in the training subset. The explanatory performance was measured as a fraction of explained clustering variance (ratio of between-cluster sum of squares to total sum of squares). The reproducibility was assessed by the rate of correct cluster assignment in 10-fold cross-validation (CV) with cluster assignment in the folds by an inverse distance weighted 27-nearest neighbors classifier. Note the superior stability of the PAM/cosine distance algorithm.

(B) Determination of the cluster number by the bend of the within-cluster sum of squares curve and the peak mean silhouette statistic.



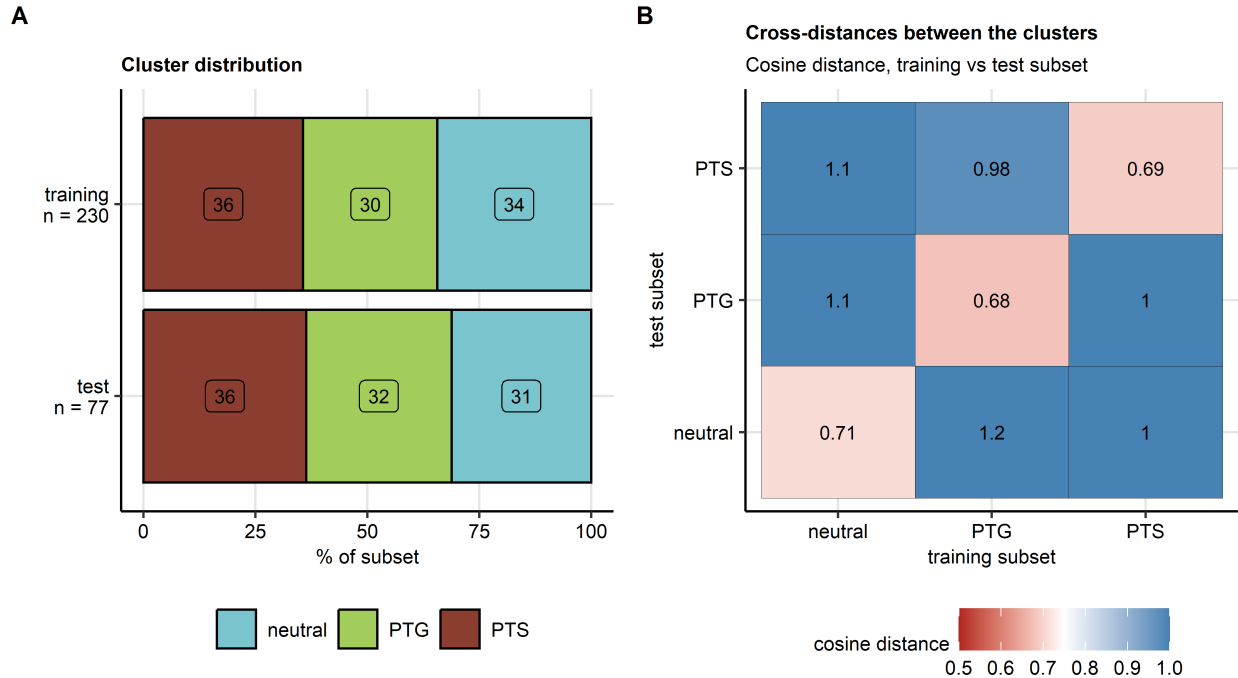
Supplementary Figure S4. Semi-supervised clustering.

The mental clusters were defined in respect to psychometric scoring in the training subset of the study cohort by PAM (partition around medoids) with cosine distance between the observations. Assignment of the test subset observations to the mental clusters was done with the inverse distance weighted 27-nearest neighbors classifier. By this procedure, three mental

clusters were identified: neutral, PTG (post-traumatic growth) and PTS (post-traumatic stress). Numbers of observations in the mental clusters are shown in the plot legend in (A).

(A) Observations in the training and test subsets of the study cohorts were subjected to two-dimensional UMAP (uniform manifold approximation and projection) in respect to the psychometric scores. UMAP layouts are shown in scatter plots. Points represent single samples. Point color codes for the cluster assignment.

(B) Pairwise cosine distances between observations in the mental clusters of the training and test subsets presented in heat maps.



Supplementary Figure S5. Distribution of the mental clusters and cosine distances between the mental clusters in the training and test subset of the study cohort.

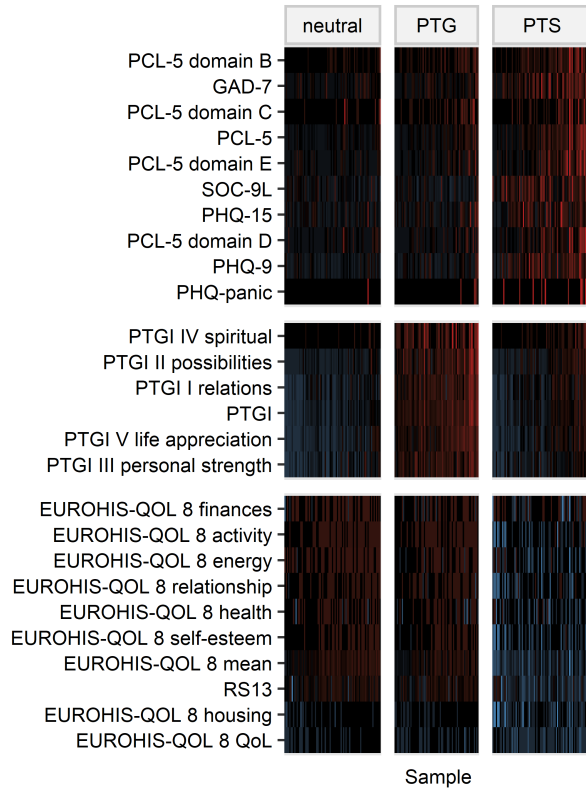
Study cohorts observations were assigned to the neutral, PTG (post-traumatic growth) and PTS (post-traumatic stress) clusters by semi-supervised clustering as presented in Supplementary Figure S4.

(A) Percentages of observations in the training and test subsets of the study cohort in the mental clusters. Numbers of complete observations are indicated in the Y axis.

(B) Cosine cross-distances between the mental clusters in the training and test subset of the study dataset visualized as a heat map. Cross-distances are indicated in the tiles.

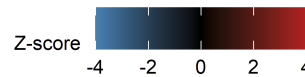
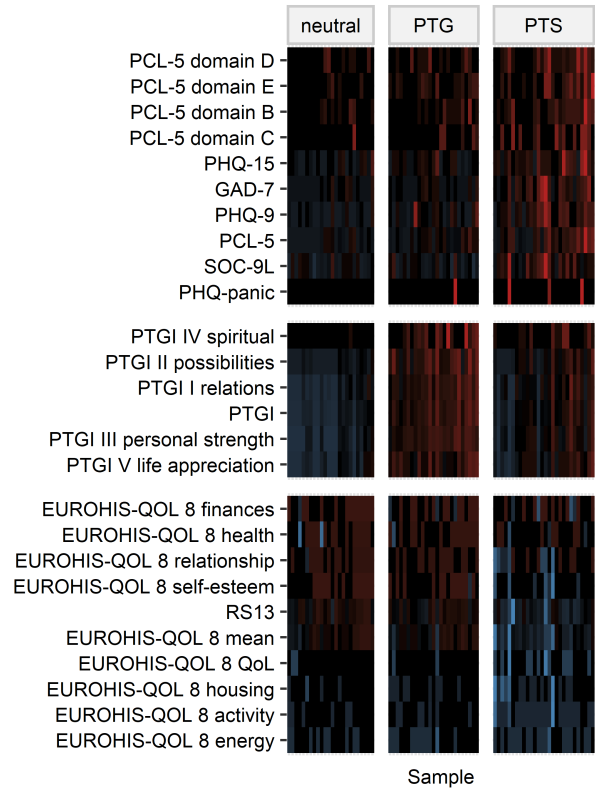
Training

neutral: n = 79, PTG: n = 69, PTS: n = 82



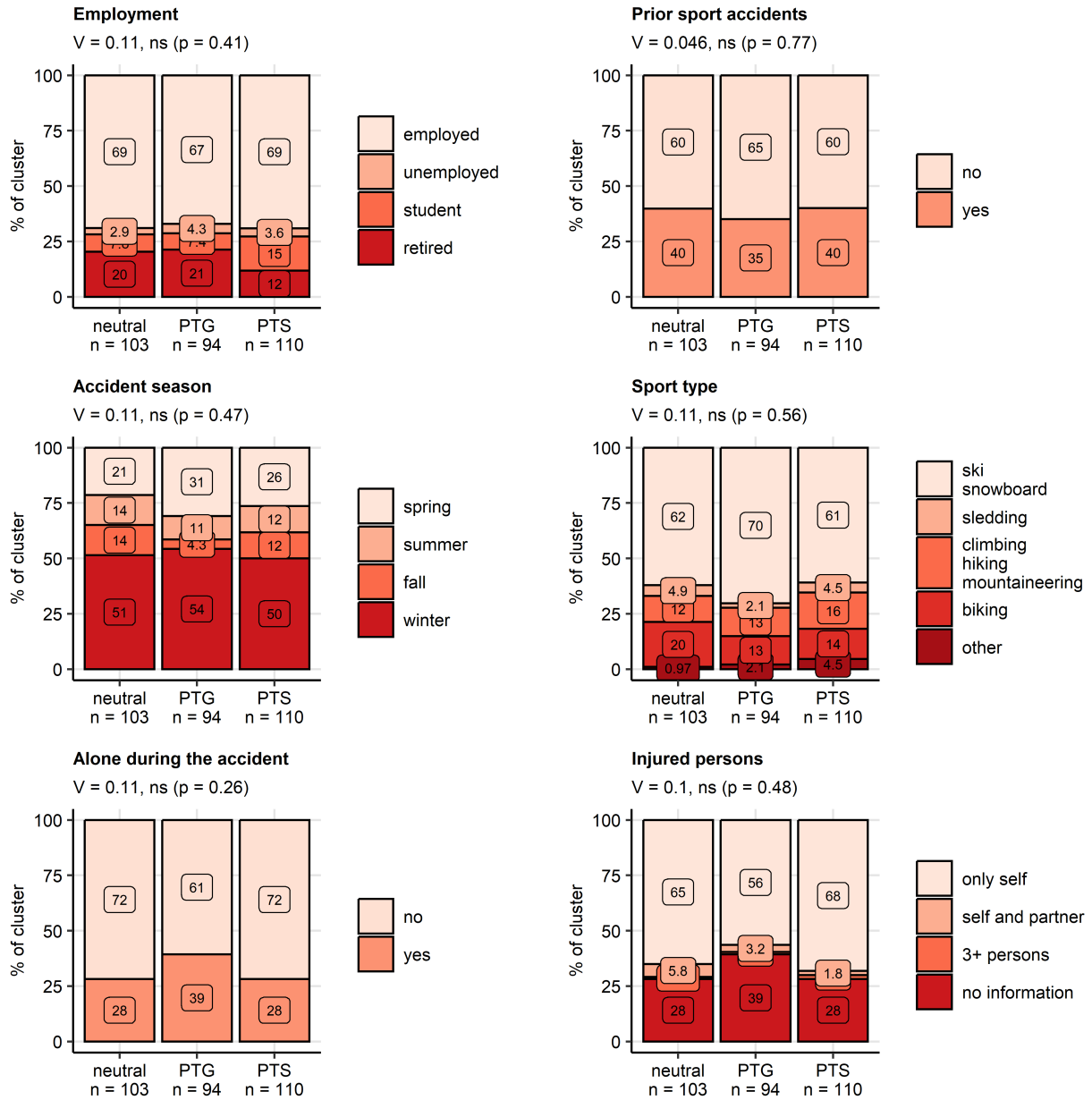
Test

neutral: n = 24, PTG: n = 25, PTS: n = 28



Supplementary Figure S6. Levels of psychometric scores used for the cluster definition in the mental clusters.

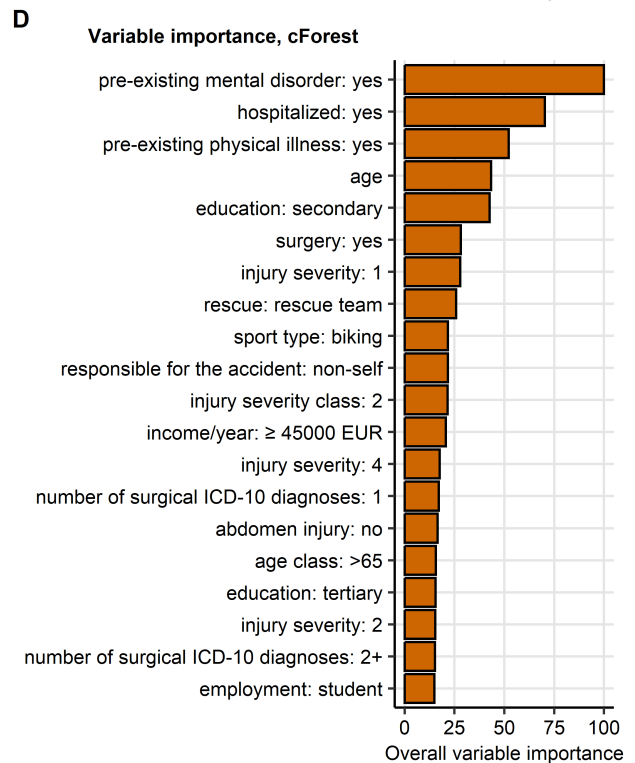
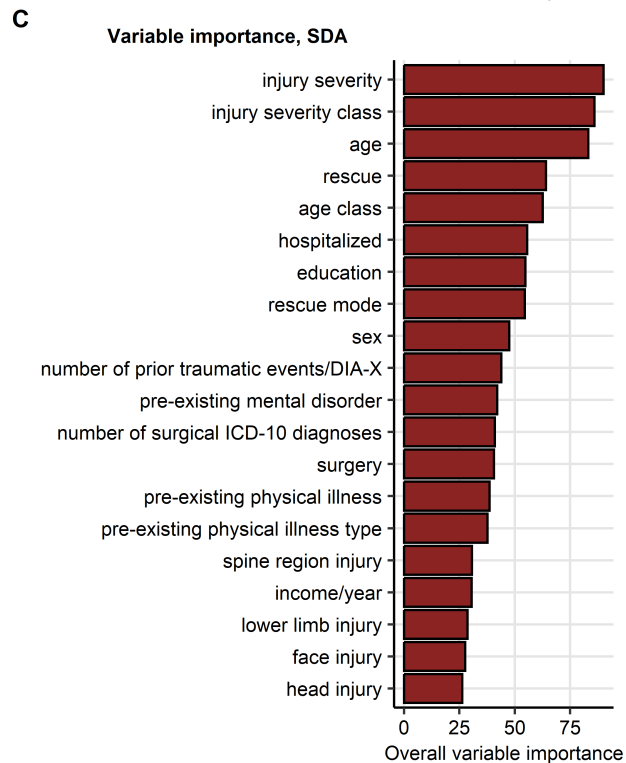
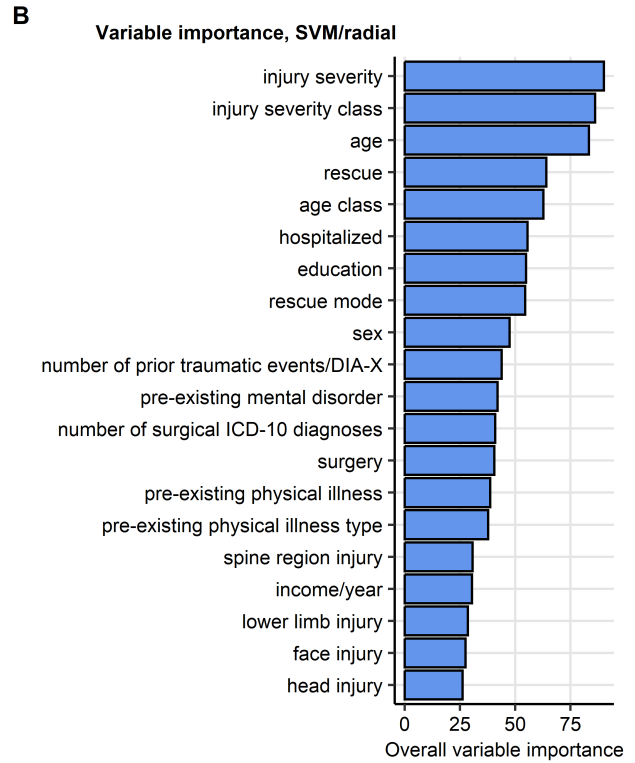
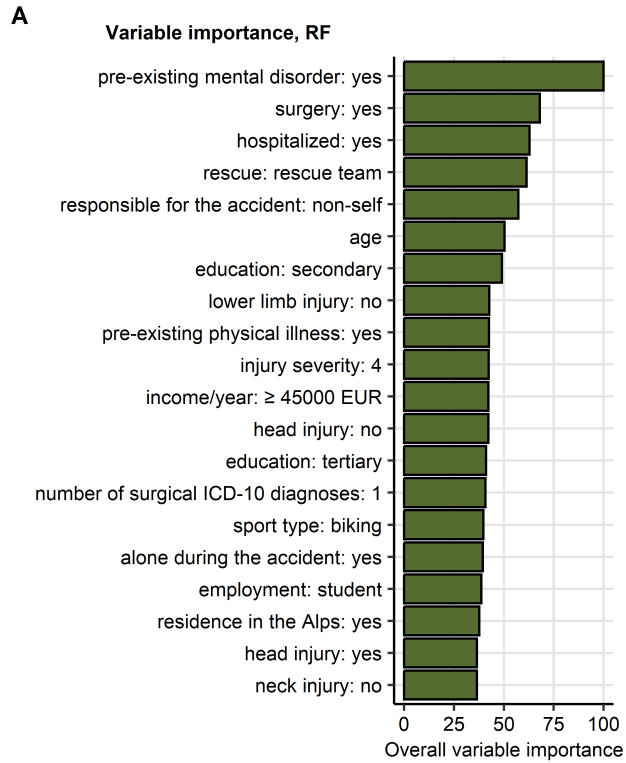
Levels of psychometric scores used in clustering in the mental clusters. Normalized, median-centered score values (Z scores) are presented as heat maps. Numbers of observations in the clusters are displayed in the plot captions. PSS4: 4-item perceived stress scale; GAD-7: 7-item general anxiety disorder scale; PHQ: patient health questionnaire; EUROHIS-QOL 8: 8-item EUROHIS project quality of life scale; SOC-9L: Leipzig 9-item sense of coherence questionnaire; RS13: 13-item resilience scale; PCL-5 DSM-5: PTSD checklist for DSM-5; PTGI: post-traumatic growth inventory; PTSD: post-traumatic stress disorder.



Supplementary Figure S7. Employment status, prior sport accidents, and accident details in the mental clusters.

Distribution of employment status, frequency of sport accidents in the past, distribution of the accidents in the seasons, accident sport types, frequency of being alone during the accident and number of injured persons in the mental clusters. Statistical significance was determined by χ^2 test with Cramer V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Percentages of variable's categories in the entire study cohort

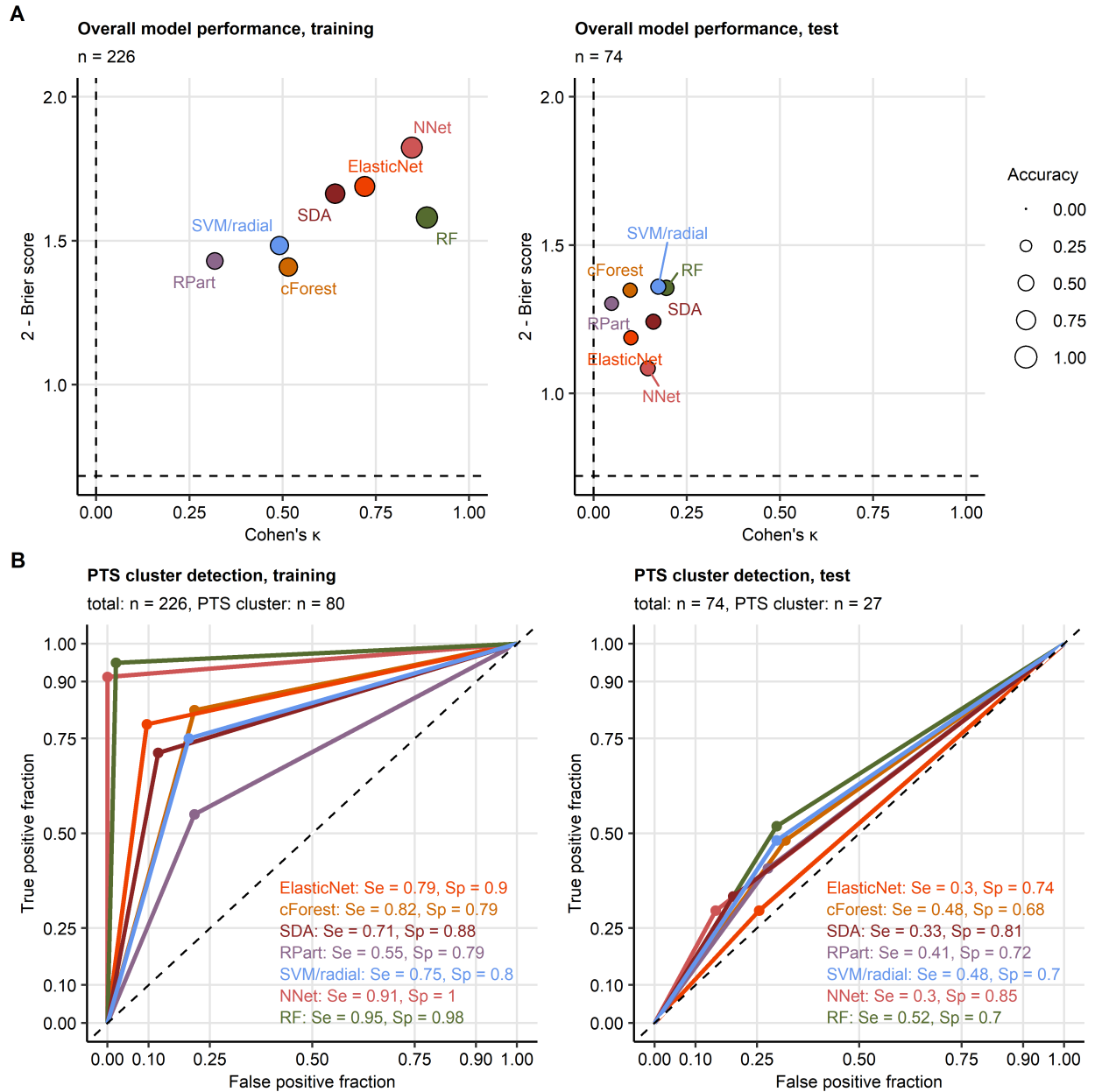
are presented in stack plots. Effect sizes and p-values are displayed in the plot captions. Numbers of observations in the clusters are presented in the X axes.



Supplementary Figure S8. Variable importance metrics for the random forest, support vector machine, discriminant analysis, and conditional random forest algorithms. Predictors available during acute medical management of the accident.

The cluster assignment was modeled with demographic, medical history and accident-related explanatory factors available during acute medical management of the accident.

Psychometric variables used for cluster definition, mental disorder symptoms, resilience classes as well as presence and frequency of flashbacks were excluded from the explanatory factor set. Variable importance metrics of machine learning algorithms with the best performance at predicting the mental cluster assignment in the test subset of the study cohorts were computed. Importance metrics for the top 20 most important variables are presented as bar plots. (A) Random forest, (B) support vector machines with radial kernel, (C) shrinkage discriminant analysis, (D) conditional random forest.



Supplementary Figure S9. Assignment of accident victims to the mental clusters based on explanatory factors available during acute medical management of the accident and long-term follow-up.

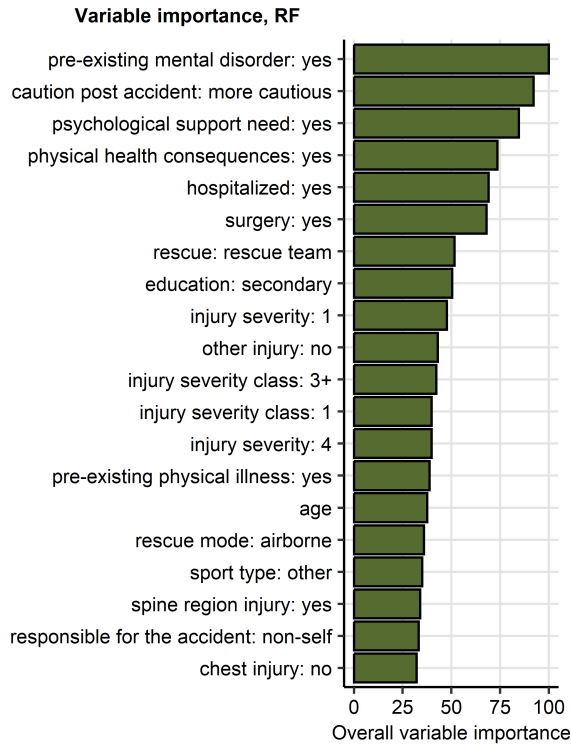
The cluster assignment was modeled with demographic, medical history, accident- and recovery-related explanatory factors available during acute medical management of the accident and follow-up. Psychometric variables used for cluster definition, mental disorder symptoms, resilience classes as well as presence and frequency of flashbacks were excluded from the explanatory factor set.

(A) Accuracy of the predicted Cluster assignment and predictive performance of the modeling algorithms was assessed by overall cluster assignment accuracy, Cohen's κ inter-rater accuracy metric (high values indicate good accuracy) and Brier score (low values indicate good performance) in the training and test subsets of the study cohort. Performance metrics are presented in scatter plots. Point size codes for the overall cluster assignment accuracy. Point color codes for the modeling algorithm. Numbers of complete observations are displayed in the plot captions.

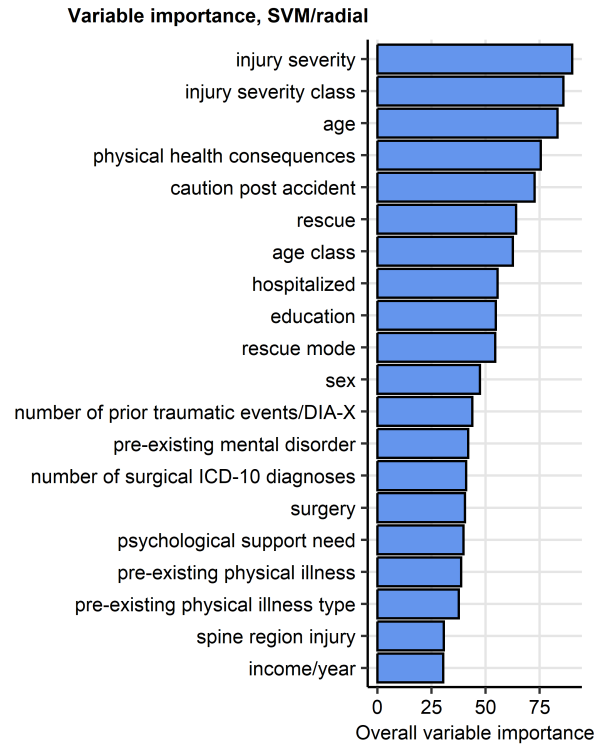
(B) Sensitivity (Se) and specificity (Sp) of detection of participants assigned to the PTS cluster (post-traumatic stress) investigated by receiver-operating characteristic in the training and test subset of the study cohort. Sensitivity and specificity values are indicated in the plots. Line color codes for the modeling algorithm. Numbers of complete observations and observations in the PTS cluster are indicated in the plot captions.

RF: random forest; NNet: neural network with a single hidden layer; SVM/radial: support vector machines with radial kernel; RPart: recursive partitioning; SDA: shrinkage discriminant analysis; cForest: conditional random forest; Elastic Net: elastic net multinomial regression.

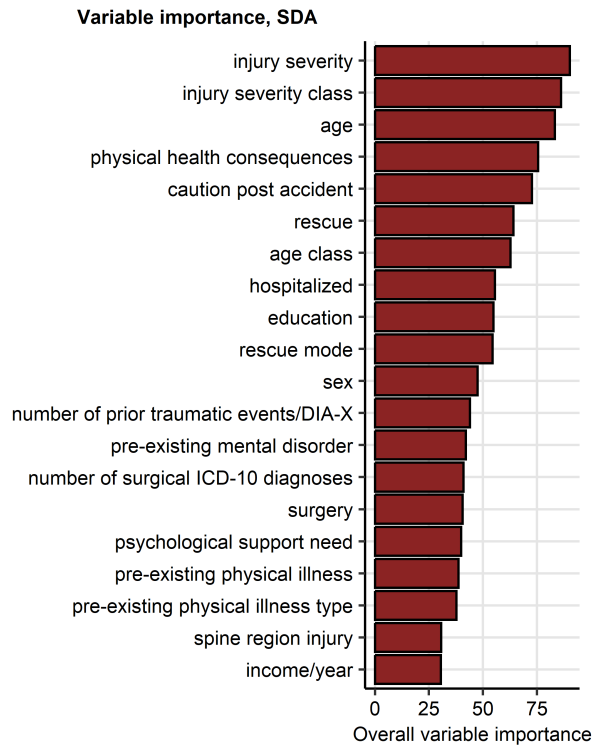
A



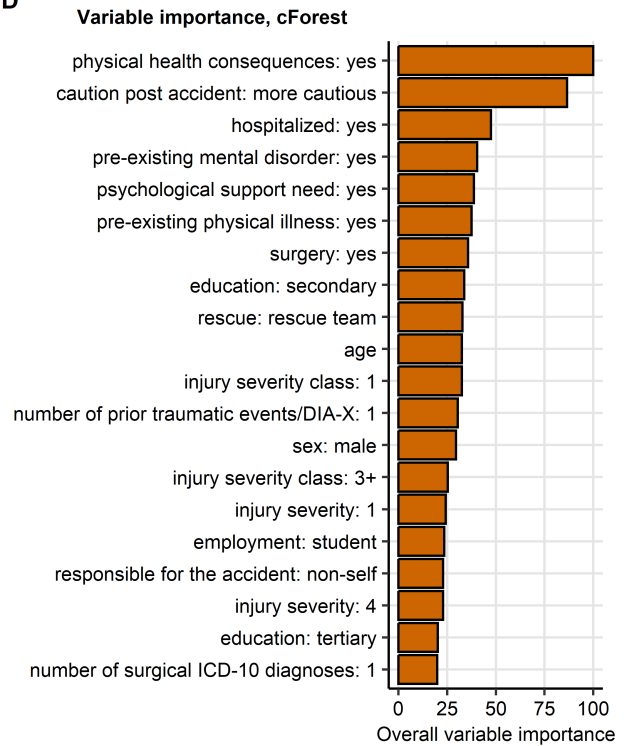
B



C



D



Supplementary Figure S10. Variable importance metrics for the random forest, support vector machine, discriminant analysis, and conditional random forest algorithms. Predictors available during acute medical management of the accident and during the follow-up.

The cluster assignment was modeled with demographic, medical history, accident- and recovery-related explanatory factors available during acute medical management of the accident and follow-up. Psychometric variables used for cluster definition, mental disorder symptoms, resilience classes as well as presence and frequency of flashbacks were excluded from the explanatory factor set. Variable importance metrics of machine learning algorithms with the best performance at predicting the mental cluster assignment in the test subset of the study cohorts were computed. Importance metrics for the top 20 most important variables are presented as bar plots. (A) Random forest, (B) support vector machines with radial kernel, (C) shrinkage discriminant analysis, (D) conditional random forest.

References

1. R Core Team, Bivand R, Carey VJ, DebRoy S, Eglen S, Guha R, Herbrandt S, Lewin-Koh N, Myatt M, Nelson M, et al. foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ... (2022) <https://cran.r-project.org/web/packages/foreign/index.html>
2. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund 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)
3. Henry L, Wickham Hadley. rlang: Functions for Base Types and Core R and 'Tidyverse' Features. (2022) <https://cran.r-project.org/web/packages/rlang/index.html>
4. Gagolewski M, Tartanus B. Package 'stringi'. (2021) <https://cran.r-project.org/web/packages/stringi/index.html> <http://cran.ism.ac.jp/web/packages/stringi/stringi.pdf>
5. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2021) <https://cran.r-project.org/package=rstatix>
6. Signorell A. DescTools: Tools for Descriptive Statistics. (2022) <https://cran.r-project.org/package=DescTools>
7. Revelle W. 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>
8. 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>
9. Schubert E, Rousseeuw PJ. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. Springer (2019). p. 171–187 doi: [10.1007/978-3-030-32047-8_16](https://doi.org/10.1007/978-3-030-32047-8_16)
10. 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)
11. Konopka T. umap: Uniform Manifold Approximation and Projection. (2022) <https://cran.r-project.org/web/packages/umap/index.html>

12. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018) <https://arxiv.org/abs/1802.03426v3>
<http://arxiv.org/abs/1802.03426>
13. Meyer D, Buchta C. proxy: Distance and Similarity Measures. (2022) <https://cran.r-project.org/web/packages/proxy/index.html>
14. Breiman L. Random forests. *Machine Learning* (2001) 45:5–32. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
15. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* (2017) 77:1–17. doi: [10.18637/JSS.V077.I01](https://doi.org/10.18637/JSS.V077.I01)
16. Ripley BD. *Pattern recognition and neural networks*. Cambridge University Press (2014). doi: [10.1017/CBO9780511812651](https://doi.org/10.1017/CBO9780511812651)
17. Weston J, Watkins C. Multi-Class Support Vector Machines. (1998)
18. Karatzoglou A, Hornik K, Smola A, Zeileis A. kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software* (2004) 11:1–20. doi: [10.18637/JSS.V011.I09](https://doi.org/10.18637/JSS.V011.I09)
19. Therneau TM, Atkinson B, Ripley BD. rpart: Recursive Partitioning and Regression Trees. (2022) <https://cran.r-project.org/web/packages/rpart/index.html>
20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. *Classification and Regression Trees* (2017) 1–358. doi: [10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN](https://doi.org/10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN)
21. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* (2006) 15:651–674. doi: [10.1198/106186006X133933](https://doi.org/10.1198/106186006X133933)
22. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* (2007) 8:1–21. doi: [10.1186/1471-2105-8-25/FIGURES/11](https://doi.org/10.1186/1471-2105-8-25/FIGURES/11)
23. Hothorn T, Hornik K, Strobl C, Zeileis A. party: A Laboratory for Recursive Partytioning. (2022) <https://cran.r-project.org/web/packages/party/index.html>
24. Ahdesmäki M, Strimmer K. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *https://doi.org/10.1214/09-AOAS277* (2010) 4:503–519. doi: [10.1214/09-AOAS277](https://doi.org/10.1214/09-AOAS277)
25. Ahdesmaki M, Zuber V, Gibb S, Strimmer K. sda: Shrinkage Discriminant Analysis and CAT Score Variable Selection. (2022) <https://cran.r-project.org/web/packages/sda/index.html>

26. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* (2005) 67:301–320. doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* (2010) 33:1–22. doi: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01)
28. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software* (2008) 28:1–26. doi: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)
29. Wickham Hadley. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. New York: Springer-Verlag (2016). <https://ggplot2.tidyverse.org>
30. Sachs MC. PlotROC: A tool for plotting ROC curves. *Journal of Statistical Software* (2017) 79:1–19. doi: [10.18637/jss.v079.c02](https://doi.org/10.18637/jss.v079.c02)
31. Krassowski M. ComplexUpset: Create Complex UpSet Plots Using 'ggplot2' Components. (2021) <https://cran.r-project.org/web/packages/ComplexUpset/index.html>
32. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st ed. Sebastopol: O'Reilly Media (2019).
33. Pedersen TL. patchwork: The Composer of Plots. (2023) <https://cran.r-project.org/web/packages/patchwork/index.html>
34. Gohel D. flextable: Functions for Tabular Reporting. (2022) <https://cran.r-project.org/web/packages/flextable/index.html>
35. Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J. rmarkdown: Dynamic Documents for R. (2022) <https://cran.r-project.org/web/packages/rmarkdown/index.html>
36. Xie Y. *Bookdown: Authoring books and technical documents with R Markdown*. (2016). doi: [10.1201/9781315204963](https://doi.org/10.1201/9781315204963)
37. Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. (2022) <https://cran.r-project.org/web/packages/knitr/index.html>
38. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* (2006) 166:1092–1097. doi: [10.1001/ARCHINTE.166.10.1092](https://doi.org/10.1001/ARCHINTE.166.10.1092)
39. Löwe B, Spitzer RL, Zipfel S, Herzog W. Auflage Manual 17.07. (2002).

40. 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)
41. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic medicine* (2002) 64:258–266. doi: [10.1097/00006842-200203000-00008](https://doi.org/10.1097/00006842-200203000-00008)
42. Leppert K, Koch B, Brähler E, Und BS-KD, 2008 U. Die Resilienzskala (RS)–Überprüfung der Langform RS-25 und einer Kurzform RS-13. *Klinische Diagnostik und Evaluation* (2008) 1:226–243.
https://www.academia.edu/download/44388154/A_406.pdf
43. Schumacher J, Wilz G, Gunzelmann T, Brähler E. Die sense of coherence scale von antonovsky: Teststatische überprüfung in einer repräsentativen bevölkerungsstichprobe und konstruktion einer kurzskala. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie* (2000) 50:472–482. doi: [10.1055/s-2000-9207](https://doi.org/10.1055/s-2000-9207)
44. Schmidt S, Mühlen H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. *European Journal of Public Health* (2006) 16:420–428. doi: [10.1093/EURPUB/CKI155](https://doi.org/10.1093/EURPUB/CKI155)
45. Tedeschi RG, Calhoun LG. The Posttraumatic Growth Inventory: measuring the positive legacy of trauma. *Journal of traumatic stress* (1996) 9:455–471. doi: [10.1007/BF02103658](https://doi.org/10.1007/BF02103658)
46. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, Keane TM. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychological assessment* (2016) 28:1379–1391. doi: [10.1037/PAS0000254](https://doi.org/10.1037/PAS0000254)
47. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *CMAJ* (2012) 184:E191. doi: [10.1503/CMAJ.110829/-/DC1](https://doi.org/10.1503/CMAJ.110829/-/DC1)
48. Maercker A, Bromberger F. Checklisten und Fragebogen zur Erfassung traumatischer Ereignisse in deutscher Sprache. *Trierer Psychologische Berichte* (2005) 32:
49. O'Brien CP. The CAGE Questionnaire for Detection of Alcoholism. *JAMA* (2008) 300:2054–2056. doi: [10.1001/JAMA.2008.570](https://doi.org/10.1001/JAMA.2008.570)
50. Gennarelli TA, Wodzin E. AIS 2005: A contemporary injury scale. *Injury* (2006) 37:1083–1091. doi: [10.1016/j.injury.2006.07.009](https://doi.org/10.1016/j.injury.2006.07.009)
51. McDonald RP. *Test theory: A unified treatment*. 1st Editio. New Yor: Psychology Press (1999). doi: [10.4324/9781410601087](https://doi.org/10.4324/9781410601087)

52. BARTLETT MS. THE STATISTICAL CONCEPTION OF MENTAL FACTORS. *British Journal of Psychology General Section* (1937) 28:97–104. doi: [10.1111/j.2044-8295.1937.tb00863.x](https://doi.org/10.1111/j.2044-8295.1937.tb00863.x)
53. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of health and social behavior* (1983) 24:385–396. doi: [10.2307/2136404](https://doi.org/10.2307/2136404)
54. Sinclair VG, Wallston KA. The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment* (2004) 11:94–101. doi: [10.1177/1073191103258144](https://doi.org/10.1177/1073191103258144)
55. 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)
56. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* (1960) 20:37–46. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)
57. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica* (2012) 22:276. doi: [10.11613/bm.2012.031](https://doi.org/10.11613/bm.2012.031)
58. Field AP. Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll, 4th edition. *Choice Reviews Online* (2013) 50:xviii, 908, xxxvi. <http://www.uk.sagepub.com/field4e/default.htm>
59. Cohen J. Statistical Power Analysis for the Behavioral Sciences. *Statistical Power Analysis for the Behavioral Sciences* (2013) doi: [10.4324/9780203771587](https://doi.org/10.4324/9780203771587)
60. 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)
61. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* (1987) 20:53–65. doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
62. Brier GW. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review* (1950) 78:1–3. doi: [10.1175/1520-0493\(1950\)078<0001:vofeit>2.0.co;2](https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2)
63. Garson GD. Interpreting neural-network connection weights. *AI Expert* (1991) 6:47–51. doi: [10.5555/129449.129452](https://doi.org/10.5555/129449.129452)