

Protocol

# Predicting Therapy Outcomes in Patients With Stress-Related Disorders: Protocol for a Predictive Modeling Study

Ludwig Franke Föyen<sup>1,2,3,4</sup>, MSc; Victoria Sennerstam<sup>1,3,4</sup>, MSc; Evelina Kontio<sup>1,3</sup>, MSc; Oskar Flygare<sup>5</sup>, PhD; Magnus Boman<sup>6,7,8</sup>, PhD; Elin Lindsäter<sup>1,3,5</sup>, PhD

<sup>1</sup>Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>Stress Research Institute, Department of Psychology, Stockholm University, Stockholm, Sweden

<sup>3</sup>Gustavsberg University Primary Care Center, Academic Primary Care Center, Region Stockholm, Stockholm, Sweden

<sup>4</sup>Department of Clinical Neuroscience, Osher Center for Integrative Health, Karolinska Institutet, Stockholm, Sweden

<sup>5</sup>Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet and Stockholm Health Care Services, Stockholm, Sweden

<sup>6</sup>Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

<sup>7</sup>BioClinicum, MedTechLabs, Karolinska University Hospital, Stockholm, Sweden

<sup>8</sup>Division of Psychiatry, University College London, London, United Kingdom

**Corresponding Author:**

Ludwig Franke Föyen, MSc

Division of Psychology

Department of Clinical Neuroscience

Karolinska Institutet

Nobels väg 9

Stockholm, 171 65

Sweden

Phone: 46 8 123 395 50

Email: [ludwig.franke.foyen@ki.se](mailto:ludwig.franke.foyen@ki.se)

## Abstract

**Background:** While cognitive behavioral therapy has shown efficacy in treating stress-related disorders, such as adjustment disorder and exhaustion disorder, knowledge about factors contributing to treatment response is limited. Improved identification of such factors could enhance assessment procedures and treatment strategies. In addition, evaluating how traditional prediction methods and machine learning can complement each other may help bridge gaps in understanding and predicting treatment response.

**Objective:** This study aims to (1) evaluate putative predictors of treatment response in patients with stress-related disorders using traditional prediction methods and (2) model treatment outcomes using a machine learning approach. This design combines the interpretability of traditional methods with the ability of machine learning to identify complex patterns.

**Methods:** We will analyze data from a randomized controlled trial comparing 2 internet-delivered treatments, cognitive behavioral therapy versus an active control treatment, for patients diagnosed with adjustment disorder or exhaustion disorder (N=300). Prediction models will be based on pooled data from both treatment arms due to the limited sample size and lack of knowledge on predictors of treatment effects. Putative predictors include sociodemographic and clinical information, clinician-assessed data, self-rated symptoms, and cognitive test scores. The primary outcome of interest is responder status on the Perceived Stress Scale-10, evaluated based on the reliable change index posttreatment. For the traditional approach, univariate logistic regressions will be conducted for each predictor, followed by an ablation study for significant predictors. For the machine learning approach, 4 classifiers (logistic regression with elastic net, random forest, support vector machine, and AdaBoost) will be trained and evaluated. The dataset will be split into training (70%) and testing (30%) sets. Hyperparameter tuning will be conducted using 5-fold cross-validation with randomized search. Model performance will be assessed using balanced accuracy, precision, recall, and area under the curve.

**Results:** All data were collected between April 2021 and September 2022. We hypothesize that key predictors will include younger age, education level, baseline symptom severity, treatment credibility, and history of sickness absence. We anticipate that the machine learning models will outperform a dummy model predicting the majority class and achieve a balanced accuracy of  $\geq 67\%$ , thus indicating clinical usefulness.

**Conclusions:** This study will contribute to the limited research on predictors of treatment outcome in stress-related disorders. The findings could support the development of more personalized and effective treatments for individuals diagnosed with adjustment disorder or exhaustion disorder, potentially improving clinical practice and patient outcomes. If successful, this dual approach may encourage future studies with larger datasets and the implementation of machine learning models in clinical settings, ultimately enhancing precision in mental health care.

**International Registered Report Identifier (IRRID):** DERR1-10.2196/65790

(*JMIR Res Protoc* 2025;14:e65790) doi: [10.2196/65790](https://doi.org/10.2196/65790)

## KEYWORDS

adjustment disorder; cognitive behavioral therapy; exhaustion disorder; machine learning; predictive modeling; psychological stress; therapy outcome

## Introduction

### Background

Mental disorders have a negative effect on quality of life, often precipitating personal suffering and work disability [1]. Around 23% of all who receive a psychiatric diagnosis in Swedish primary care receive a stress-related diagnosis [2], and these account for most psychiatric long-term sickness absences [3]. In Sweden, disorders believed to stem from persistent or overwhelming subtraumatic life events are often categorized using the diagnostic labels adjustment disorder (AD) or exhaustion disorder (ED). Even though ED is only recognized as a medical diagnosis in the Swedish version of the *International Classification of Diseases, Tenth Revision*, the clinical picture of ED is similar to the internationally acknowledged burnout construct [4], a condition that is often associated with significant suffering and work disability [5].

According to diagnostic definitions of AD and ED, these conditions develop in the context of one or several subtraumatic life events (stressors), resulting in mixed symptoms of anxiety, depressed mood, disturbed sleep, fatigue, and impaired memory and concentration. They share symptomatology with other mental disorders, and their diagnostic validity is debated [6,7]. Despite evidence indicating the efficacy of cognitive behavioral therapy (CBT) [8-10] and problem-solving interventions [11] on symptoms of stress, many studies have suffered from significant attrition, and knowledge regarding the factors that contribute to treatment response is still limited [8,12]. Improved identification of such factors could facilitate development of improved assessment procedures and adaptive treatment strategies that might improve outcomes [13].

Research on predictors of psychiatric treatment outcomes is limited [14,15] but demographic factors (eg, age and education level) [16,17], clinical characteristics (eg, use of medication and symptom severity) [17-19], treatment-related factors (eg, treatment credibility and adherence) [16,18] and cognitive functioning [20] have been associated with treatment outcomes.

When it comes to studies investigating predictors of treatment for stress-related disorders, Kocalevent et al [15] found that symptoms of anxiety but not perceived stress, depressive symptoms, or demographic variables predicted self-rated mental health following treatment for patients diagnosed with AD. In a study investigating burnout, Pallich et al [21] identified emotional competence, but not demographic characteristics, as

a predictor of treatment response. However, both of these studies suffer from limited generalizability due to their inadequate description of the treatment offered, the fact that the interventions were conducted in an inpatient setting, and the lack of control groups. In ED patients, one study identified several predictors of treatment outcome following multimodal rehabilitation, including younger age, baseline symptom severity (insomnia, anxiety, and depression), perfectionism, physical activity level, treatment credibility, and a history of sickness absence due to ED [22]. However, the effects of demographics and pretreatment symptoms were so small that they offered limited clinical utility. In sum, at the current stage of research, it is a challenge for clinicians to determine who will benefit from treatment, underscoring the imperative for more sophisticated predictive studies.

Traditionally, prediction in psychiatry has relied on interpretable linear or logistic regression models. The aim has been to identify variables explaining a statistically significant portion of the variance in outcome, under the premise that such variables should inform researchers and clinicians. For example, the presence of previous sickness absence and earlier unsuccessful treatment attempts might lead a psychologist to conclude that a patient requires additional support, possibly extending the treatment duration. Although this approach of identifying predictors has offered some clinical utility, it often falls short in practice; the predictive power of specific variables in isolation is typically inadequate to inform assessment, treatment selection, and adaptations of interventions. Given the inherent complexity of mental disorders, the likelihood of pinpointing strong predictors with clinical utility is small, thus limiting the practical value of this approach [23,24].

Machine learning (ML) represents a promising methodological shift in psychiatric prediction modeling, transitioning from the identification of statistically significant predictors to an emphasis on quantifiable model performance, characterized by ensemble methods and adaptability to new datasets. This approach often sacrifices explainability in favor of enhanced predictive performance but offers unique value in handling the complex, nonlinear, high-dimensional data characteristic of mental disorders [25]. With this approach, a model generates a prediction (eg, remission, yes or no) intended to be actionable for a clinician. For example, patients predicted to have low probability of treatment success could be offered additional psychological support or an alternative intervention, thus increasing the likelihood of remission [26,27].

Forsell et al [28] have proposed a balanced accuracy (BACC) threshold of 67% as a benchmark for clinical utility in psychiatric applications, offering a tangible goal for ML implementation. However, the efficacy of ML in this domain remains an ongoing area of inquiry, and its capacity to surpass conventional methods in clinical usefulness is yet to be established [29].

Given the high prevalence and substantial societal costs associated with stress-related disorders, it is imperative to critically evaluate both the applicability and the limitations of ML within this specific context. Such an assessment will not only contribute to the broader understanding of the role of ML in precision psychiatry but also inform the development of more effective diagnostic and treatment strategies for stress-related disorders.

### Objective of the Study

The overall objective of this study is to predict treatment outcomes in patients with stress-related disorders. Due to limitations in existing methods for prediction analyses, this study aims to first evaluate putative predictors using a traditional prediction paradigm, and second to model treatment outcomes using an ML approach. Our primary outcome of interest is responder status after treatment on the Perceived Stress Scale-10 (PSS-10), evaluated using the reliable change index (RCI; further described in the Planned Statistical Analysis and Data Cleaning and Preparation sections). On the basis of earlier research on predictors of treatment outcome, we hypothesize that key predictors will include younger age, education level, baseline symptom severity, treatment credibility, and history of sickness absence. Furthermore, we anticipate that the ML models will outperform a dummy model predicting the majority class and achieve a BACC of  $\geq 67\%$ , thus being indicated clinically useful [28].

## Methods

### Study Design

We will use collected data from a randomized controlled trial (RCT; N=300) of internet-delivered CBT for patients diagnosed with AD or ED compared to an active, internet-delivered control condition consisting of general health-promoting advice. A priori power analysis conducted for the main outcome in the RCT indicated that 300 study participants would be needed for a 90% power to detect a between-group effect size of Cohen  $d=0.4$  with a significance level of .05 and an expected attrition rate of 10%. Due to the limited sample size and general lack of knowledge on predictors of treatment effect, prediction models in this study will be based on pooled data from both treatment arms. The study design is prospective, and predictors will include sociodemographic and clinical information, clinician-assessed data, self-rated symptoms, and results from cognitive test scores. The results will be reported in line with the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence) statement [30].

### Ethical Considerations

The study was approved by the Swedish Ethical Review Authority (registration 2020-03198; 2023-06857-02) and was preregistered on ClinicalTrials.gov (NCT04797273). All participants provided written informed consent before inclusion, and their data are pseudonymized and securely stored on an encrypted server. Participants received no monetary compensation but accessed study interventions free of charge. No identifying information of participants will be included in the manuscript or supplementary materials.

### Procedure

#### Participants

In total, 300 nationally recruited individuals were diagnosed with a primary diagnosis of AD (n=142, 47.3%) or ED (n=158, 52.7%) and were included in the RCT. Participant recruitment was carried out through social media, newspaper advertisements, and information provided to health care clinics. Participants self-referred to the study web page, where they signed digital informed consent and completed a screening battery consisting of sociodemographic and clinical background questions as well as self-report symptom questionnaires. Participants were subsequently clinically assessed by a psychologist using a structured diagnostic interview, including Mini International Neuropsychiatric Interview [31], self-rated ED [32], and the Adjustment Disorder New Module-8 (ADNM-8) [33]. For inclusion, participants needed to (1) fulfill the criteria for a primary diagnosis of AD or ED, (2) be aged between 18 and 65 years, (3) have regular access to a computer with internet access, and (4) be able to read and write in the Swedish language. Exclusion criteria included (1) drug use or addiction during the past 6 months, (2) current or past psychosis or bipolar disorder, (3) current risk of suicide, (4) changed psychopharmacological treatment in the past month, (5) other ongoing psychological treatment, and (6) previous experience of CBT for AD or ED in the past year.

#### Treatment

Patients were randomized to one out of two 12-week internet-delivered treatments (CBT and general health-promoting advice). They both consisted of web-based text-based modules with related exercises and assignments. Patients were guided sequentially through the modules by a therapist via a secure web-based platform. The therapists' primary role was to provide feedback on exercises, support in problem-solving, and to give emotional and technical support via weekly asynchronous text messages. Therapists were licensed clinical psychologists or clinical psychology students in their final year of training. Because this study will not evaluate the individual treatments, they will not be further described here. A full description of the treatments is described in the study by Sennerstam et al [34].

### Outcomes

The primary outcome in this study and the original RCT is PSS-10 [35]. The PSS-10 is a self-report questionnaire developed to evaluate an individual's perception of life as unpredictable, uncontrollable, and overwhelming. Responses are recorded on an ordinal scale ranging from 0 *never* to 4 *very often*, reflecting the individual's feelings and thoughts over the

past month. It contains statements, such as *‘In the last month, how often have you been upset because of something that happened unexpectedly?’* and sum scores range from 0 to 40. The PSS is the most commonly used outcome measure of stress-management interventions globally [8,36,37]. For this study, a Swedish version of the PSS-10 was used. The PSS-10 has been found to exhibit high internal consistency (Cronbach  $\alpha=0.84$ ) and adequate construct validity [38]. The PSS-10 was administered digitally through the web-based study platform before randomization to treatment, every 3 weeks during the treatment phase, and at treatment completion (12 weeks). During treatment, the instructions for the PSS-10 were modified to have patients consider the last week instead of the last month. For this study, the sum score of the PSS-10 will be dichotomized

into responder or nonresponder after treatment based on the RCI criteria [39] to differentiate between statistically significant change and those attributable to measurement error or natural variability. The PSS-10 baseline and 3-week measurement will also be used as predictors.

## Putative Predictors

### Overview

Predictors were gathered through self-report measures that were administered in the web-based study platform, clinical assessment conducted before inclusion to the study, and remote cognitive testing. [Table 1](#) presents all predictors included in the study.

**Table 1.** Putative predictors of treatment outcome in stress-related disorders

| Predictor                             | Construct measured    | Type        | Clinician-rated | Scoring range   |
|---------------------------------------|-----------------------|-------------|-----------------|-----------------|
| <b>Sociodemographics</b>              |                       |             |                 |                 |
| Age (y)                               | — <sup>a</sup>        | Interval    |                 | 18-65           |
| Sex                                   | —                     | Categorical |                 | Male or female  |
| Relationship status                   | —                     | Categorical |                 | 3 categories    |
| Number of children                    | —                     | Interval    |                 | 0-∞             |
| Educational attainment                | —                     | Ordinal     |                 | 9 categories    |
| Employment status                     | —                     | Categorical |                 | 8 categories    |
| Employment type                       | —                     | Categorical |                 | 11 categories   |
| Self-rated computer skills            | —                     | Ordinal     |                 | 5 categories    |
| Self-rated reading skills             | —                     | Ordinal     |                 | 5 categories    |
| Swedish native speaker                | —                     | Categorical |                 | Yes or no       |
| <b>Clinical characteristics</b>       |                       |             |                 |                 |
| Number of medications                 | Medication            | Interval    | ✓               | 0-4             |
| Antidepressants                       | Medication            | Categorical | ✓               | Yes or no       |
| Sleep medication                      | Medication            | Categorical | ✓               | Yes or no       |
| Pain medication                       | Medication            | Categorical | ✓               | Yes or no       |
| Anxiolytics                           | Medication            | Categorical | ✓               | Yes or no       |
| Diagnosis                             | Primary diagnosis     | Categorical | ✓               | 2 categories    |
| Secondary diagnosis                   | Secondary diagnosis   | Interval    | ✓               | 0-4             |
| Depression                            | Secondary diagnosis   | Categorical | ✓               | Yes or no       |
| Anxiety disorder                      | Secondary diagnosis   | Categorical | ✓               | Yes or no       |
| Insomnia                              | Secondary diagnosis   | Categorical | ✓               | Yes or no       |
| Other disorders                       | Secondary diagnosis   | Categorical | ✓               | Yes or no       |
| S-ED <sup>b</sup>                     | Exhaustion disorder   | Ordinal     | ✓               | 3 categories    |
| ADNM-8 <sup>c</sup> criteria          | Adjustment disorder   | Categorical | ✓               | Yes or no       |
| ADNM-8 number of stressors            | Adjustment disorder   | Interval    | ✓               | 0-11            |
| ADNM-8 stressors                      | Adjustment disorder   | Categorical | ✓               | 16 categories   |
| Duration of current episode           | —                     | Interval    | ✓               | 0-∞             |
| Age of first episode (y)              | —                     | Interval    | ✓               | 0-65            |
| Sick-leave status                     | Sickness absence      | Interval    |                 | 0%-100% 5 steps |
| Sick-leave duration                   | Sickness absence      | Ordinal     |                 | 5 categories    |
| <b>Self-rated symptoms</b>            |                       |             |                 |                 |
| AUDIT <sup>d</sup>                    | Alcohol consumption   | Interval    |                 | 0-40            |
| GAD-7 <sup>e</sup>                    | Anxiety symptoms      | Interval    |                 | 0-21            |
| SMBQ <sup>f</sup> cognitive weariness | Burnout               | Continuous  |                 | 0-7             |
| SMBQ exhaustion                       | Burnout               | Continuous  |                 | 0-7             |
| SMBQ listlessness                     | Burnout               | Continuous  |                 | 0-7             |
| MADRS-S <sup>g</sup>                  | Depression            | Interval    |                 | 0-54            |
| KEDS <sup>h</sup>                     | Exhaustion disorder   | Interval    |                 | 0-54            |
| WHODAS <sup>i</sup> 2.0               | Functional disability | Continuous  |                 | 0%-100%         |

| Predictor                           | Construct measured             | Type       | Clinician-rated | Scoring range |
|-------------------------------------|--------------------------------|------------|-----------------|---------------|
| EQ-5D-5L                            | Quality of Life                | Interval   |                 | 5-25          |
| BBQ <sup>j</sup>                    | Quality of life                | Interval   |                 | 0-96          |
| ISI <sup>k</sup>                    | Insomnia severity              | Interval   |                 | 0-28          |
| SRH-5 <sup>l</sup>                  | Self-rated health              | Interval   |                 | 0-5           |
| PSS-10 <sup>m</sup>                 | Perceived stress               | Interval   |                 | 0-40          |
| PHQ-15 <sup>n</sup>                 | Somatoform symptoms            | Interval   |                 | 0-30          |
| 6-QEMP <sup>o</sup>                 | Subjective memory impairment   | Interval   |                 | 0-30          |
| <b>3-week measurement</b>           |                                |            |                 |               |
| SMBQ cognitive weariness            | Burnout                        | Continuous |                 | 0-7           |
| SMBQ exhaustion                     | Burnout                        | Continuous |                 | 0-7           |
| SMBQ listlessness                   | Burnout                        | Continuous |                 | 0-7           |
| ISI                                 | Insomnia severity              | Interval   |                 | 0-28          |
| PSS-10                              | Perceived stress               | Interval   |                 | 0-40          |
| <b>Treatment-related predictors</b> |                                |            |                 |               |
| Clinician treatment expectancy      | —                              | Interval   | ✓               | 0-10          |
| Treatment credibility scale         | —                              | Interval   |                 | 0-10          |
| <b>Cognitive functioning</b>        |                                |            |                 |               |
| SDMT <sup>p</sup>                   | Attention and processing speed | Interval   |                 | 0-∞           |
| FAS <sup>q</sup>                    | Executive functions            | Interval   |                 | 0-∞           |
| Stroop index                        | Executive functions            | Continuous |                 | 0-∞           |
| Stroop inhibition                   | Executive functions            | Continuous |                 | 0-∞           |
| CERAD <sup>r</sup> learning         | Memory                         | Interval   |                 | 0-30          |
| CERAD recognition                   | Memory                         | Interval   |                 | 0-10          |
| Corsi forward                       | Memory                         | Interval   |                 | 0-9           |

<sup>a</sup>Not applicable.

<sup>b</sup>S-ED: self-rated exhaustion disorder.

<sup>c</sup>ADNM-8: The Adjustment Disorder New Module-8.

<sup>d</sup>AUDIT: Alcohol Use Disorder Identification Test.

<sup>e</sup>GAD-7: General Anxiety Disorder-7.

<sup>f</sup>SMBQ: Shirom-Melamed Burnout Questionnaire.

<sup>g</sup>MADRS-S: Montgomery-Åsberg Depression Rating Scale.

<sup>h</sup>KEDS: Karolinska Exhaustion Disorder Scale.

<sup>i</sup>WHODAS: World Health Organization Disability Assessment Schedule.

<sup>j</sup>BBQ: Brunnsvikens Brief Quality of Life Scale.

<sup>k</sup>ISI: Insomnia Severity Index.

<sup>l</sup>SRH-5: Self-Rated Health-5.

<sup>m</sup>PSS-10: Perceived Stress Scale-10.

<sup>n</sup>PHQ-15: Patient Health Questionnaire-15.

<sup>o</sup>6-QEMP: 6-item Questionnaire of Everyday Memory Problems.

<sup>p</sup>SDMT: Symbol Digit Modality Test.

<sup>q</sup>FAS: Verbal Fluency Test.

<sup>r</sup>CERAD: Consortium to Establish a Registry for Alzheimer's Disease.

### **Sociodemographic Variables**

Information on age (interval), sex (male, female, other, or prefer not to disclose), relationship status (in relationship, single, or widowed), number of children, educational attainment (in 9 categories between <9 years of school to PhD), employment status (eg, student, unemployed, or full-time work), and employment type (in 11 categories, eg, employed in the private sector, by the municipality, or other) was gathered before the start of treatment using the web-based study platform. Self-rated reading and computer skills were rated separately on a 5-step ordinal scale from *poor* to *very good*. Patients also reported if they were Swedish native speakers.

### **Clinical Characteristics**

During the clinical interview, patients reported their medication regimen, specifying both the number (0-4) and type of medication (antidepressants, anxiolytics, sleep medication, and pain medication and yes or no). Primary diagnosis (AD or ED), and possible secondary psychiatric diagnosis (eg, anxiety or depressive disorder) was assessed by the clinician using Mini International Neuropsychiatric Interview, self-rated ED (ordinal categories ranging from *no* to *yes—to a high degree*) [32], and the ADN-8 [33]. Using ADN-8, the patient was asked about which specific stressors had been present in the past 2 years (in 16 options, eg, *too much or too little work* or *financial difficulties*). The clinician assessed the length of the current episode (in months), and the age of the patients first episode (in years). Sick-leave status upon inclusion in the study (0%-100% in 5 steps), length of current sick-leave episode (*0-1 months* to *>12 months* in 5 categories) was self-reported.

### **Self-Rated Symptoms**

*Alcohol use* was assessed using the Alcohol Use Disorder Identification Test [40,41]. This 10-item screening instrument evaluates alcohol consumption, drinking behavior, and alcohol-related problems over the past year. It contains items, such as *How often do you have six or more drinks on one occasion?* rated on various ordinal scales, typically ranging from 0 to 4.

*Symptoms of anxiety* were measured using the Generalized Anxiety Disorder-7 scale [42]. This screening tool assesses generalized anxiety symptoms over the past 2 weeks. It comprises 7 items, such as *not being able to stop or control worrying* rated on a 4-point ordinal scale ranging from 0 *not at all* to 3 *nearly every day*.

*Symptoms of burnout* were measured using the Shirom-Melamed Burnout Questionnaire [43,44]. It aims to measure 3 components of burnout; emotional and physical fatigue, cognitive weariness, and listlessness and contains statements such as *I have difficulty concentrating* rated on a 7-point scale ranging from 1 *never or almost never* to 7 *always or almost always* with some items using reversed scoring.

*Symptoms of depression* were measured using Montgomery-Åsberg Depression Rating Scale [45]. It is a 9-item questionnaire used to measure different aspects of depression such as concentration difficulties, suicidal thoughts, sadness,

and affected appetite with answers rated on a 7-point ordinal scale from 0 to 6.

*Symptoms of exhaustion disorder* were measured using the 9-item Karolinska Exhaustion Disorder Scale [46]. Measuring different aspects of exhaustion such as fatigue, endurance, and sleep impairment, answers are rated on an ordinal scale from 0 to 6 (eg, ability to concentrate; ranging from 0 “I do not have any difficulty concentrating, and can read, watch TV and converse normally” to 6 “I cannot concentrate on anything at all.”)

*Functional disability* was measured using The World Health Organization Disability Assessment Schedule (2.0) [47], developed to assess functioning in the last 30 days in 6 different life domains, including cognition, mobility, self-care, relationships, life activities, and societal participation. It contains statements, such as “I have difficulty standing for longer periods such as 30 minutes.” Answers are rated on a 5-point ordinal scale ranging from 0 *never* to 4 *extreme or unable*. A 12-item version was used.

*Quality of life* was assessed using the EQ-5D-5L [48,49] and the Brunnsvikens Brief Quality of Life Scale [50]. The EQ-5D-5L contains 5 dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression each rated on 5 levels of severity from *no problems* to *extreme problems*. The Brunnsvikens Brief Quality of Life Scale is a 12-item questionnaire that assesses 6 life areas (leisure time, view on life, learning, creativity, view of self, and friends and friendship). Ratings range from 0 *strongly disagree*, to 4 *strongly agree*, on statements of the importance and satisfaction of each area.

*Insomnia severity* was measured using the Insomnia Severity Index [51]. The Insomnia Severity Index is a 7-item questionnaire designed to assess aspects of insomnia, including difficulty falling asleep, difficulty staying asleep, and satisfaction with sleep. Ratings are given using an ordinal scale ranging from 0 to 4.

*Self-rated health* was assessed using Self-Rated Health 5 asking patients to rate their general health on a scale of 1, *very bad* to 5 *very good* [52].

*Somatoform symptoms* were assessed using the Patient Health Questionnaire [53]. It consists of 15 questions covering somatic symptoms commonly seen in primary care, such as back pain, headache, and nausea. Answers are rated on a 3-point ordinal scale ranging from *not at all bothered* to *bothered a lot*.

*Subjective memory impairment* was measured using the 6-item Questionnaire of Everyday Memory Problems (6-QEMP) [54]. A 5-item version has previously been used to assess subjective memory problems in this patient population [55,56]. The present version was adapted by Stigsdotter Neely for use in patients with stress-related disorders with statements, such as “How do you think your memory functions now compared to before your stress-related mental health problems?.” The answers are rated on a 5-point ordinal scale.

### **Treatment-Related Predictors**

*Clinician Treatment Expectancy* was judged after patient assessment, upon inclusion in the study, by clinicians rating the probability of the patient improving after treatment on a scale of 0 *no improvement* to 10 *full remission*.

The *Treatment Credibility Scale* was administered 3 weeks after the start of treatment [57]. Patients were asked questions about their impression of the treatment and if they thought they would improve. It included statements such as “How logical do you think this treatment is?” and “How confidently would you recommend this treatment to a friend with the same problems as you?” on a scale of 0 *not at all* to 10 *very logical, very confidently*.

### **Cognitive Functioning**

*Attention and processing speed* were measured using the Symbol Digit Modality Test. A test originally developed by Smith [58,59] that measures visual detection, attention, and processing speed. A key with 9 different symbols and matching numbers is shown on the upper part of the display. At the center one of these 9 symbols are shown and the task of the participant is to choose the corresponding number using the key as guidance. The test score is the number of correct entries in 90 seconds. Comparable substitution tasks are considered sensitive to treatment effects for patients with multiple sclerosis [60] and depression [61], and it has been used in patients with stress-related disorders [62].

*Executive functioning* was measured using the Verbal Fluency Test (FAS) Word Fluency Test and the Stroop test. FAS was first described by Spreen and Benton [63], and it measures spontaneous verbal fluency and selective attention and shifting. The participant is tasked with producing words beginning with a certain alphabet letter (F, A, and S). Names, numbers, or repeated words are not allowed. The test score is the number of correct words beginning with the letter. FAS and similar word fluency tasks have been shown to be impaired in patients with stress-related exhaustion [62].

The Stroop test, originally developed by Stroop [64] and described by Jensen and Rohwer [65], measures executive functioning, inhibition, as well as updating and processing speed [66]. The test has 2 parts, (1) 20 color words are presented (green, yellow, blue, or red) and they are colored congruent to their meaning (eg, the word red colored in red). In the bottom part of the display, the color words are displayed on 4 buttons. The task is to, as quickly and thoroughly as possible, click the correct button. (2) Twenty color words are presented but displayed in an incongruent color (eg, the word red colored in green). The task of the participant is to click the button containing the color of the word as quickly and thoroughly as possible. Test score is calculated as an index (number of correct answers in part 2 divided by average time in seconds from part 2) and for interference (average time in part one–average time in part one). Performance of Stroop in patients with stress-related disorders has been shown to be impaired in 2 studies [62,67], but not in others [68,69].

*Memory and learning* were assessed using the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) Word

List Learning Test and Corsi block-tapping test forward. CERAD was originally developed for use with Alzheimer disease [70] but is similar to other word-list tasks used in this patient population. It measures verbal learning and episodic memory. In the learning part of the test, a word list containing 10 words is presented over 3 trials and the task after every trial is to recall the words from the list. For every presentation the order is mixed. In the delayed recall part of the test (trial 4) that occurs after 5 to 10 minutes, the participant is asked to recall the words. Test score for the learning time is number of correct words in trial 1 to 3, and in the delayed recall part, number of correct words in trial 4. Similar word-list tasks have been used previously to assess memory functioning in patients with stress-related disorders [62,69].

Corsi block-tapping test forward gives information about visual ability of attention, short-term memory and working memory [71]. It contains 2 parts, but in this test battery, only the first part of the test is used. Nine blocks are displayed, and the testing platform starts by lighting up a sequence of blocks. The task is to repeat the sequence of blocks that the platform has displayed. The task starts out easy with only 2 blocks, but the difficulty increases by adding a longer sequence of blocks until the participant enters the incorrect sequence twice at the same number of blocks. The test score is the maximum number of correct repeated blocks. A cross-sectional study comparing patients with stress-related disorders to a healthy normative group found impaired performance on this test [62].

### **Planned Statistical Analysis**

All data will be prepared and analyzed using the latest version of Python [72] and the libraries NumPy [73], Pandas [74], and scikit-learn [75] or equivalent statistical packages. A notebook containing the analysis in documented code will be made available on Open Science Framework [76] for research transparency following the analysis.

### **Data Cleaning and Preparation**

We will transform categorical variables into a format suitable for numerical analysis. For binary categorical variables, we will use label encoding. For multinomial variables, we will apply one-hot encoding. In addition, for ordinal data, which have a natural order, we will transform the categories into integers.

Predictor variables with over 20% missing data will be excluded from the analysis. Categorical variables exhibiting low variance, as determined by predictors with <5% of a certain response will be removed. For instance, by removing the variable “Sleep medication” if it only occurs in 3 out of 300 patients. This approach aims to reduce unnecessary complexity in the predictions and to minimize the risk of overfitting. To control for multicollinearity, variables with a correlation coefficient  $\geq 0.8$  will be removed from the traditional prediction analysis but will be retained for the ML model development. Data that are highly skewed will be transformed if deemed appropriate.

Cognitive test results will be manually reviewed before model fitting to validate a proper result. Comments pertaining to technical difficulties and disturbances that may have affected the test result will be assessed by two of the authors and lead to exclusion if so judged. Participants who have noted during



screening that Swedish is not their native language will be excluded from the analysis for CERAD and FAS. We will standardize the raw scores from the cognitive tests using normative regression models with age, education and sex as covariates. This standardization process will convert raw scores into  $Z$  scores, as previously described by Franke Föyén et al [62] and for a full overview of the multiple linear regression models used and how they were calculated, see the studies by Mindmore [77] and van den Hurk et al [78].

Patients who have missing data for the posttreatment PSS-10, ie, the missing outcome variable for the primary aim, will be replaced by a PSS-10 process measurement at week 10 if available; If not, the patient will be excluded from the analysis. The number of participants excluded from the final models will be described.

To prepare our primary outcome, RCI for the PSS-10 before to after treatment will be computed using the following formula [39]:

$$RCI = \frac{(X_1 - X_2)}{S_{diff}}$$

$$S_{diff} = \sqrt{2 * (S_E)^2}$$

$$S_E = SD * \sqrt{1 - \alpha}$$

Cronbach  $\alpha=0.83$  from normative data will be used [38]. Patients exhibiting an RCI of  $-1.96$  will be classified as responders.

### **Descriptive Statistics**

Descriptive statistics will be used to summarize the sample characteristics and pretreatment variables, including mean or median, SDs and IQR for continuous variables, and proportions for categorical variables.

### **Predictor Analysis**

For the traditional regression analysis, data will be imputed using the KNN imputer. The imputer, a nonparametric imputation method, works by imputing missing values based on the  $k$ -nearest neighbors; in this study  $k$  will be determined by cross-validation. It uses the Euclidean distance metric to find the nearest neighbors and can be used for both numerical and categorical data. Each missing value is imputed using values from its  $k$ -nearest neighbors. After imputation, we will run univariate logistic regressions for each predictor listed in Table 1 using RCI as a target variable. Predictors that are statistically significant in the univariate analyses will then be included in an ablation study, a systematic approach to evaluate feature importance. This method involves iteratively removing each significant predictor from a full model, measuring the change in explained variance, and then reinserting it, thereby quantifying each predictor's unique contribution to the model's explanatory power in the context of all other features.

### **ML Model Development**

For an introduction on the technical terms introduced in this section, see the review article by Bzdok and Meyer-Lindenberg [13].

### **Train Test Split**

As the ultimate goal of any model is to predict an outcome in unseen data, the ML models will be developed using a training set, and then evaluated on a test set stratified on main diagnosis (AD or ED) and responder status. In total, 70% of the data will be used for selecting predictor variables and training the models, and 30% for testing the prediction accuracy of the models. The choice of 70 to 30 was due to the limited size of our sample, as fewer observations in the testing data makes it difficult to use uncommon predictors. No external validation set is currently available at the time of writing.

### **Standardization and Imputation**

Standardization and imputation will be applied on the training and test data separately to avoid data leakage. Numerical data will be standardized and all missing data will be imputed using the KNN imputer.

### **Model Descriptions**

We will train and evaluate 4 different ML classifiers, a multiple logistic regression (LogReg) classifier using elastic net, a random forest (RF) classifier, a support vector machine (SVM) classifier, and an AdaBoost classifier. For a review of the models used, see the textbook by Geron [79]. In short, the LogReg classifier works by modeling the probability of a binary outcome based on one or more predictor variables, using the logistic function to ensure the output is between 0 and 1. We will use elastic net regularization to facilitate feature selection and prevent overfitting. Elastic net combines L1 (lasso) and L2 (ridge) penalties, encouraging sparsity and maintaining stability in the model. The RF classifier works by building multiple decision trees on random subsets of data and predictors. Each tree's prediction is based on splits that minimize variance in the target variable, with the final model ensembling these predictions. The SVM classifier works by finding the hyperplane that maximizes the margin between different classes in the feature space. SVM is particularly effective in high-dimensional spaces and when the number of dimensions exceeds the number of samples. AdaBoost, the final classifier, works by combining multiple weak classifiers, typically decision trees, into a single strong classifier. It sequentially fits these weak learners on repeatedly modified versions of the data, focusing more on misclassified instances to improve overall accuracy.

### **Hyperparameter Tuning**

We will conduct 5-fold cross-validation using randomized search for hyperparameter tuning and training evaluation to enhance the external generalizability and robustness of the results. This process involves defining a hyperparameter space, then randomly selecting a predetermined number of samples—in this case, 10—from this space, and conducting 5-fold cross-validation for each selected set of hyperparameters. Fivefold cross-validation is done by partitioning the data into 5 subsets, training the model on 4 subsets, and validating it on

the remaining subset. This process is repeated 5 times, with each subset used exactly once as the validation data. The best performing hyperparameters will be chosen for the final models that are trained and then evaluated on the test set.

The hyperparameter ranges for the LogReg will include C values from 0.01 to 100 and l1\_ratio values from 0 to 1. For RF, the parameter ranges will include the number of estimators from 5 to 1200, minimum samples required to split a node from 10 to 200, maximum depths from 5 to 750, and a binary indicator for bootstrapping. For SVM, the parameter range for the randomized search will include regularization parameter C values from 0.01 to 1 and for AdaBoost, the parameter ranges for the randomized search will include the number of estimators, ranging from 1 to 1500, and learning rates from 0.001 to 2.5.

### ***Model Interpretation***

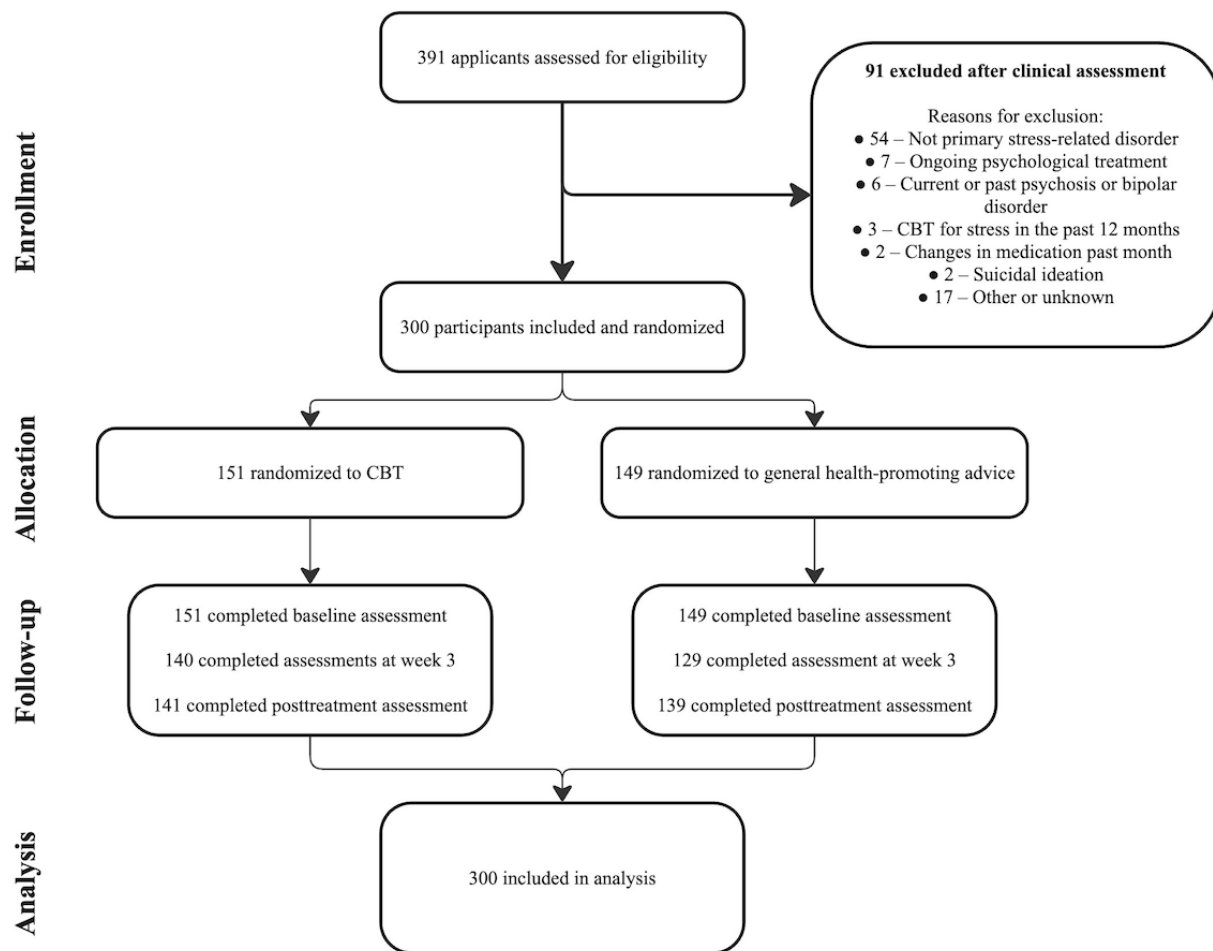
The models developed to identify the responder status will be evaluated using BACC, precision and recall, both in the training set obtained through k-fold cross-validation and in the test set. Predictor importance in the RF model will be determined using Scikit-learn's Feature importance function, which quantifies each predictor's contribution to the model's balanced classification accuracy. Area under the curve will be used to assess the models' capability to distinguish between classes

accurately. The approach will aim to provide a clear understanding of the models' effectiveness and the role of various predictors. Our primary outcome of interest for comparison will be BACC in each model in the test set with the aim that (1) the model should perform better than a dummy model that simply predicts the most common responder status, and (2) that the model should perform 67% BACC or above to be deemed clinically useful [28]. Furthermore, the models will be statistically compared using bootstrap sampling. Specifically, we will generate 5000 bootstrap samples from the test set, calculating the BACC for each model on each sample. The distributions of these bootstrap BACCs will be compared and we will conclude that there is a statistically significant difference between models if the CIs do not overlap.

### ***Results***

This study was funded by ALF medicin (20190148), Region Stockholm (SLSO 2022–1278; SLSO 2022–1276), and Region Stockholm in collaboration with Stockholm university (FoUI-939533). OF is supported by the Swedish innovation agency (No. 2022-00549). All data were collected (N=300) between April 2021 and September 2022. For a participant flow diagram throughout the study, see [Figure 1](#).

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) diagram showing participant flow through enrollment, allocation, follow-up, and analysis. CBT: cognitive behavioral therapy.



A cross-sectional study investigating baseline cognitive functioning as compared with a healthy reference group has been published indicating small-to-moderate objective cognitive impairments [62], raising the question of whether objective cognitive function serves as a predictor of treatment response. In addition, an interim analysis of pre- and postcomparisons was presented at a conference in September 2022. These earlier analyses addressed separate research questions and did not influence the design, methods, or objectives of the current protocol. As of March 2025, data have not been analyzed for this study.

## Discussion

### Overview

This study will use a high-quality dataset from an RCT to investigate potential treatment predictors using both traditional prediction methods and an ML paradigm. This dual approach will enable the identification of predictors of treatment response in a patient population where prior research is limited. In addition, it will facilitate comparisons between different methodological approaches to prediction research.

### Comparison to Prior Work

To the best of our knowledge, this is the first study to apply an ML approach to study predictors of treatment outcome in patients diagnosed with AD or ED. In line with previous traditional prediction research of treatment outcomes in stress-related disorders, we hypothesize that younger age, education level, symptom severity, treatment credibility, and history of sickness absence will predict treatment response [15,22]. Furthermore, we anticipate that the ML models will outperform a dummy model and achieve a BACC of 67% or higher, surpassing the benchmark indicated by Forsell and others [28]. If confirmed, our findings would support the notion that predictive models using sociodemographic, clinical, self-rated, treatment-related, and potentially cognitive variables are valuable when predicting therapy outcomes, as have been suggested in other patient populations [17,18,80]. In subsequent research, these models should be externally validated and tested in implementation trials to assess their utility as decision support tools. Such trials could evaluate whether integrating predictions into treatment planning improves outcomes and supports personalized care.

## Strengths and Limitations

The study's strengths include the use of a multimodal dataset from an RCT, including objective cognitive functioning. By comparing traditional prediction methods with advanced ML models and employing techniques, such as cross-validation and hyperparameter tuning, the study has the potential to generate robust and generalizable insights into treatment outcomes, contributing to methodological advancements in prediction research.

A limitation is the modest sample size (N=300) which may increase the risk of overfitting. While smaller sample sizes have been used in previous ML studies [80-82], it is well established that limited sample sizes can hinder generalization [23]. The minimal sample size required for ML prediction in mental health research depends on the explanatory power of the predictors, with some researchers advocating for at least 300 observations [83], while others recommend a larger sample of 500 to 1500 for studies involving predictors with low explanatory power [84]. External validation is widely regarded as the gold standard to ensure model generalizability [85], but such data are not currently available for this study. However, ongoing data collection by the research group may enable external validation in the near future. In the meantime, k-fold cross-validation on the training set and validation on a separate test set will be used to estimate and mitigate overfitting, providing a basis for model evaluation within the study's constraints.

In addition, the recruitment strategy, which relied on social media, newspaper advertisements, and health care clinic referrals, may introduce selection bias and limit the generalizability of the findings. Participants recruited through these channels may not fully represent the broader population of individuals with stress-related disorders, potentially

overrepresenting individuals with higher internet access, health literacy, or willingness to participate in internet-delivered interventions. These factors should be considered when interpreting the applicability of the study's results to other settings or populations.

Finally, ML models, such as RF, while effective at handling complex datasets, often prioritize predictive performance at the expense of interpretability. Unlike traditional statistical methods, their inclusion of numerous variables can make it challenging to understand the relationships between predictors and outcomes, limiting their integration into clinical practice where transparency is essential. Efforts to address this, such as using feature importance metrics, will be necessary to bridge this gap moving forward.

## Implications for Clinical Practice

The study's findings could significantly impact clinical practice by contributing to the limited research on predictors of treatment outcome for stress-related disorders. Given the current lack of a gold standard treatment for AD and ED, this research is particularly timely and relevant. The investigation into ML models for treatment outcome prediction may encourage future larger-scale studies and, potentially, the implementation of these models in clinical settings as decision support tools. These could help clinicians tailor treatments by integrating complex data, such as patient demographics, symptom severity, and treatment history, to recommend evidence-based options, guiding therapy selection, and monitoring progress in real time. By operationalizing predictive insights, decision support tools could enhance clinical precision, reduce trial-and-error in treatment, and improve patient outcomes for individuals with stress-related disorders.

## Acknowledgments

The research team gratefully acknowledges the contributions of all study patients, without whom this work would not have been possible.

## Data Availability

The data referenced in this study protocol are not yet available, as the study has not yet been conducted. Once the study is completed, data may be made available on reasonable request, subject to compliance with Swedish law (the Swedish Ethical Review Act: 2003:460). For such requests, please contact the corresponding author.

## Conflicts of Interest

LFF was previously employed by the cognitive testing company Mindmore until 2023 and runs a small-scale clinical psychology practice. All other authors declare no other conflicts of interest or financial involvement.

## References

1. Hassard J, Teoh KR, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. *J Occup Health Psychol.* Jan 2018;23(1):1-17. [doi: [10.1037/ocp0000069](https://doi.org/10.1037/ocp0000069)] [Medline: [28358567](https://pubmed.ncbi.nlm.nih.gov/28358567/)]
2. Cullen AE, Lindsäter E, Rahman S, Taipale H, Tanskanen A, Mittendorfer-Rutz E, et al. Patient factors associated with receipt of psychological and pharmacological treatments among individuals with common mental disorders in a Swedish primary care setting. *BJPsych Open.* Feb 28, 2023;9(2):e40. [FREE Full text] [doi: [10.1192/bjo.2023.8](https://doi.org/10.1192/bjo.2023.8)] [Medline: [36852532](https://pubmed.ncbi.nlm.nih.gov/36852532/)]
3. Sjukfrånvaro i psykiatriska diagnoser - En registerstudie av Sveriges arbetande befolkning i åldern 20-69 år. Försäkringskassan, Analys och rapport. Swedish Social Insurance Agency. URL: <https://tinyurl.com/4vhax7nj> [accessed 2024-04-29]

4. Guseva Canu I, Marca SC, Dell'Oro F, Balázs Á, Bergamaschi E, Besse C, et al. Harmonized definition of occupational burnout: a systematic review, semantic analysis, and Delphi consensus in 29 countries. *Scand J Work Environ Health*. Mar 01, 2021;47(2):95-107. [[FREE Full text](#)] [doi: [10.5271/sjweh.3935](https://doi.org/10.5271/sjweh.3935)] [Medline: [33258478](https://pubmed.ncbi.nlm.nih.gov/33258478/)]
5. Ahola K, Toppinen-Tanner S, Seppänen J. Interventions to alleviate burnout symptoms and to support return to work among employees with burnout: systematic review and meta-analysis. *Burn Res*. Mar 2017;4:1-11. [doi: [10.1016/j.burn.2017.02.001](https://doi.org/10.1016/j.burn.2017.02.001)]
6. Bachem R, Casey P. Adjustment disorder: a diagnosis whose time has come. *J Affect Disord*. Feb 2018;227:243-253. [doi: [10.1016/j.jad.2017.10.034](https://doi.org/10.1016/j.jad.2017.10.034)] [Medline: [29107817](https://pubmed.ncbi.nlm.nih.gov/29107817/)]
7. Lindsäter E, Svärdman F, Wallert J, Ivanova E, Söderholm A, Fondberg R, et al. Exhaustion disorder: scoping review of research on a recently introduced stress-related diagnosis. *BJPsych Open*. Aug 24, 2022;8(5):e159. [[FREE Full text](#)] [doi: [10.1192/bjo.2022.559](https://doi.org/10.1192/bjo.2022.559)] [Medline: [36458830](https://pubmed.ncbi.nlm.nih.gov/36458830/)]
8. Svärdman F, Sjöwall D, Lindsäter E. Internet-delivered cognitive behavioral interventions to reduce elevated stress: a systematic review and meta-analysis. *Internet Interv*. Sep 2022;29:100553. [[FREE Full text](#)] [doi: [10.1016/j.invent.2022.100553](https://doi.org/10.1016/j.invent.2022.100553)] [Medline: [35781929](https://pubmed.ncbi.nlm.nih.gov/35781929/)]
9. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. *J Med Internet Res*. Jan 27, 2016;18(1):e21. [[FREE Full text](#)] [doi: [10.2196/jmir.5112](https://doi.org/10.2196/jmir.5112)] [Medline: [26818683](https://pubmed.ncbi.nlm.nih.gov/26818683/)]
10. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. *J Occup Health Psychol*. Jan 2008;13(1):69-93. [doi: [10.1037/1076-8998.13.1.69](https://doi.org/10.1037/1076-8998.13.1.69)] [Medline: [18211170](https://pubmed.ncbi.nlm.nih.gov/18211170/)]
11. Arends I, Bruinvels DJ, Rebergen DS, Nieuwenhuijsen K, Madan I, Neumeyer-Gromen A, et al. Interventions to facilitate return to work in adults with adjustment disorders. *Cochrane Database Syst Rev*. Dec 12, 2012;12(12):CD006389. [doi: [10.1002/14651858.CD006389.pub2](https://doi.org/10.1002/14651858.CD006389.pub2)] [Medline: [23235630](https://pubmed.ncbi.nlm.nih.gov/23235630/)]
12. Rachyla I, Mor S, Cuijpers P, Botella C, Castilla D, Quero S. A guided Internet-delivered intervention for adjustment disorders: a randomized controlled trial. *Clin Psychol Psychother*. Mar 02, 2021;28(2):313-324. [doi: [10.1002/cpp.2518](https://doi.org/10.1002/cpp.2518)] [Medline: [32959481](https://pubmed.ncbi.nlm.nih.gov/32959481/)]
13. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. Mar 2018;3(3):223-230. [doi: [10.1016/j.bpsc.2017.11.007](https://doi.org/10.1016/j.bpsc.2017.11.007)] [Medline: [29486863](https://pubmed.ncbi.nlm.nih.gov/29486863/)]
14. Meehan AJ, Lewis SJ, Fazel S, Fusar-Poli P, Steyerberg EW, Stahl D, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Mol Psychiatry*. Jun 01, 2022;27(6):2700-2708. [[FREE Full text](#)] [doi: [10.1038/s41380-022-01528-4](https://doi.org/10.1038/s41380-022-01528-4)] [Medline: [35365801](https://pubmed.ncbi.nlm.nih.gov/35365801/)]
15. Kocalevent RD, Mierke A, Danzer G, Klapp BF. Adjustment disorders as a stress-related disorder: a longitudinal study of the associations among stress, resources, and mental health. *PLoS One*. May 13, 2014;9(5):e97303. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0097303](https://doi.org/10.1371/journal.pone.0097303)] [Medline: [24825165](https://pubmed.ncbi.nlm.nih.gov/24825165/)]
16. Salomonsson S, Santoft F, Lindsäter E, Ejeby K, Ingvar M, Öst LG, et al. Predictors of outcome in guided self-help cognitive behavioural therapy for common mental disorders in primary care. *Cogn Behav Ther*. Nov 2020;49(6):455-474. [[FREE Full text](#)] [doi: [10.1080/16506073.2019.1669701](https://doi.org/10.1080/16506073.2019.1669701)] [Medline: [31638472](https://pubmed.ncbi.nlm.nih.gov/31638472/)]
17. Wallert J, Boberg J, Kaldo V, Mataix-Cols D, Flygare O, Crowley JJ, et al. Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Transl Psychiatry*. Sep 01, 2022;12(1):357. [[FREE Full text](#)] [doi: [10.1038/s41398-022-02133-3](https://doi.org/10.1038/s41398-022-02133-3)] [Medline: [36050305](https://pubmed.ncbi.nlm.nih.gov/36050305/)]
18. Prasad N, Chien I, Regan T, Enrique A, Palacios J, Keegan D, et al. Deep learning for the prediction of clinical outcomes in internet-delivered CBT for depression and anxiety. *PLoS One*. Nov 27, 2023;18(11):e0272685. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0272685](https://doi.org/10.1371/journal.pone.0272685)] [Medline: [38011176](https://pubmed.ncbi.nlm.nih.gov/38011176/)]
19. Forsell E, Isacson N, Blom K, Jernelöv S, Ben Abdesslem F, Lindefors N, et al. Predicting treatment failure in regular care internet-delivered cognitive behavior therapy for depression and anxiety using only weekly symptom measures. *J Consult Clin Psychol*. Apr 2020;88(4):311-321. [doi: [10.1037/ccp0000462](https://doi.org/10.1037/ccp0000462)] [Medline: [31829635](https://pubmed.ncbi.nlm.nih.gov/31829635/)]
20. Groves SJ, Douglas KM, Porter RJ. A systematic review of cognitive predictors of treatment outcome in major depression. *Front Psychiatry*. 2018;9:382. [[FREE Full text](#)] [doi: [10.3389/fpsy.2018.00382](https://doi.org/10.3389/fpsy.2018.00382)] [Medline: [30210368](https://pubmed.ncbi.nlm.nih.gov/30210368/)]
21. Pallich G, Blättler L, Gomez Penedo JM, Grosse Holtforth M, Hochstrasser B. Emotional competence predicts outcome of an inpatient treatment program for burnout. *J Affect Disord*. Sep 01, 2020;274:949-954. [doi: [10.1016/j.jad.2020.05.139](https://doi.org/10.1016/j.jad.2020.05.139)] [Medline: [32664037](https://pubmed.ncbi.nlm.nih.gov/32664037/)]
22. Clason van de Leur J, Johansson F, McCracken LM, Åhs F, Brodda Jansen G, Buhman M. Predictors and sub-groups in the treatment of stress-induced exhaustion disorder. *Cogn Behav Ther*. Jul 2023;52(4):397-418. [[FREE Full text](#)] [doi: [10.1080/16506073.2023.2197148](https://doi.org/10.1080/16506073.2023.2197148)] [Medline: [37039046](https://pubmed.ncbi.nlm.nih.gov/37039046/)]
23. Chekroud AM, Bondar J, Delgado J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*. Jun 18, 2021;20(2):154-170. [[FREE Full text](#)] [doi: [10.1002/wps.20882](https://doi.org/10.1002/wps.20882)] [Medline: [34002503](https://pubmed.ncbi.nlm.nih.gov/34002503/)]
24. Vieira S, Liang X, Guiomar R, Mechelli A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. *Clin Psychol Rev*. Nov 2022;97:102193. [[FREE Full text](#)] [doi: [10.1016/j.cpr.2022.102193](https://doi.org/10.1016/j.cpr.2022.102193)] [Medline: [35995023](https://pubmed.ncbi.nlm.nih.gov/35995023/)]

25. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. May 07, 2018;14:91-118. [doi: [10.1146/annurev-clinpsy-032816-045037](https://doi.org/10.1146/annurev-clinpsy-032816-045037)] [Medline: [29401044](https://pubmed.ncbi.nlm.nih.gov/29401044/)]
26. Bjurner P, Isacson NH, Abdesslem FB, Boman M, Forsell E, Kaldov V. Study protocol for a triple-blind randomised controlled trial evaluating a machine learning-based predictive clinical decision support tool for internet-delivered cognitive behaviour therapy (ICBT) for depression and anxiety. *Open Science Framework*. URL: <https://osf.io/cs4bx/> [accessed 2024-07-08]
27. Forsell E, Jernelöv S, Blom K, Kraepelien M, Svanborg C, Andersson G, et al. Proof of concept for an adaptive treatment strategy to prevent failures in internet-delivered CBT: a single-blind randomized clinical trial with insomnia patients. *Am J Psychiatry*. Apr 01, 2019;176(4):315-323. [doi: [10.1176/appi.ajp.2018.18060699](https://doi.org/10.1176/appi.ajp.2018.18060699)] [Medline: [30696270](https://pubmed.ncbi.nlm.nih.gov/30696270/)]
28. Forsell E, Jernelöv S, Blom K, Kaldov V. Clinically sufficient classification accuracy and key predictors of treatment failure in a randomized controlled trial of internet-delivered cognitive behavior therapy for insomnia. *Internet Interv*. Sep 2022;29:100554. [FREE Full text] [doi: [10.1016/j.invent.2022.100554](https://doi.org/10.1016/j.invent.2022.100554)] [Medline: [35799973](https://pubmed.ncbi.nlm.nih.gov/35799973/)]
29. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. Jun 2019;110:12-22. [doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004)] [Medline: [30763612](https://pubmed.ncbi.nlm.nih.gov/30763612/)]
30. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Apr 18, 2024;385:q902. [FREE Full text] [doi: [10.1136/bmj.q902](https://doi.org/10.1136/bmj.q902)] [Medline: [38636956](https://pubmed.ncbi.nlm.nih.gov/38636956/)]
31. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Sheehan K, et al. Mini international neuropsychiatric interview. American Psychological Association. 2013. URL: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F18597-000> [accessed 2024-04-29]
32. Glise K, Hadzibajramovic E, Jonsdottir IH, Ahlberg G. Self-reported exhaustion: a possible indicator of reduced work ability and increased risk of sickness absence among human service workers. *Int Arch Occup Environ Health*. Jun 27, 2010;83(5):511-520. [doi: [10.1007/s00420-009-0490-x](https://doi.org/10.1007/s00420-009-0490-x)] [Medline: [19943058](https://pubmed.ncbi.nlm.nih.gov/19943058/)]
33. Kazlauskas E, Gegieckaite G, Eimontas J, Zelviene P, Maercker A. A brief measure of the international classification of diseases-11 adjustment disorder: investigation of psychometric properties in an adult help-seeking sample. *Psychopathology*. 2018;51(1):10-15. [FREE Full text] [doi: [10.1159/000484415](https://doi.org/10.1159/000484415)] [Medline: [29301130](https://pubmed.ncbi.nlm.nih.gov/29301130/)]
34. Franke Föyen L, Sennerstam V, Kontio E, Lekander M, Hedman-Lagerlöf E, Lindsäter E. Internet-delivered treatment for stress-related disorders: ICBT vs general health promotion, a randomized controlled trial. *Open Science Framework*. URL: <https://osf.io/w9apq/resources> [accessed 2024-04-29]
35. Cohen S. *Perceived Stress in a Probability Sample of the United States*. Thousand Oaks, CA: Sage Publications; 1988.
36. Lee EH. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res (Korean Soc Nurs Sci)*. Dec 2012;6(4):121-127. [FREE Full text] [doi: [10.1016/j.anr.2012.08.004](https://doi.org/10.1016/j.anr.2012.08.004)] [Medline: [25031113](https://pubmed.ncbi.nlm.nih.gov/25031113/)]
37. Franke Föyen L, Bianchi R, Lindsäter E. The Perceived Stress Scale revisited: longitudinal clinical implications of a two-factor structure. *Int J Stress Manag*. Nov 2024;31(4):367-374. [doi: [10.1037/str0000334](https://doi.org/10.1037/str0000334)]
38. Nordin M, Nordin S. Psychometric evaluation and normative data of the Swedish version of the 10-item perceived stress scale. *Scand J Psychol*. Dec 05, 2013;54(6):502-507. [doi: [10.1111/sjop.12071](https://doi.org/10.1111/sjop.12071)] [Medline: [24118069](https://pubmed.ncbi.nlm.nih.gov/24118069/)]
39. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. Feb 1991;59(1):12-19. [doi: [10.1037//0022-006x.59.1.12](https://doi.org/10.1037//0022-006x.59.1.12)] [Medline: [2002127](https://pubmed.ncbi.nlm.nih.gov/2002127/)]
40. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. *Addiction*. Jun 1993;88(6):791-804. [doi: [10.1111/j.1360-0443.1993.tb02093.x](https://doi.org/10.1111/j.1360-0443.1993.tb02093.x)] [Medline: [8329970](https://pubmed.ncbi.nlm.nih.gov/8329970/)]
41. Bergman H, Källmén H. Alcohol use among Swedes and a psychometric evaluation of the alcohol use disorders identification test. *Alcohol Alcohol*. 2002;37(3):245-251. [doi: [10.1093/alcalc/37.3.245](https://doi.org/10.1093/alcalc/37.3.245)] [Medline: [12003912](https://pubmed.ncbi.nlm.nih.gov/12003912/)]
42. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. May 22, 2006;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
43. Lundgren-Nilsson Å, Jonsdottir IH, Pallant J, Ahlberg G. Internal construct validity of the Shirom-Melamed Burnout Questionnaire (SMBQ). *BMC Public Health*. Jan 03, 2012;12(1):1. [FREE Full text] [doi: [10.1186/1471-2458-12-1](https://doi.org/10.1186/1471-2458-12-1)] [Medline: [22214479](https://pubmed.ncbi.nlm.nih.gov/22214479/)]
44. Melamed S, Kushnir T, Shirom A. Burnout and risk factors for cardiovascular diseases. *Behav Med*. Jun 1992;18(2):53-60. [doi: [10.1080/08964289.1992.9935172](https://doi.org/10.1080/08964289.1992.9935172)] [Medline: [1392214](https://pubmed.ncbi.nlm.nih.gov/1392214/)]
45. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. Apr 01, 1979;134(4):382-389. [doi: [10.1192/bjp.134.4.382](https://doi.org/10.1192/bjp.134.4.382)] [Medline: [444788](https://pubmed.ncbi.nlm.nih.gov/444788/)]
46. Besèr A, Sorjonen K, Wahlberg K, Peterson U, Nygren A, Asberg M. Construction and evaluation of a self rating scale for stress-induced exhaustion disorder, the Karolinska Exhaustion Disorder Scale. *Scand J Psychol*. Feb 2014;55(1):72-82. [FREE Full text] [doi: [10.1111/sjop.12088](https://doi.org/10.1111/sjop.12088)] [Medline: [24236500](https://pubmed.ncbi.nlm.nih.gov/24236500/)]
47. Üstün TB, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, et al. Developing the World Health Organization Disability assessment schedule 2.0. *Bull World Health Organ*. May 20, 2010;88(11):815-823. [doi: [10.2471/blt.09.067231](https://doi.org/10.2471/blt.09.067231)]

48. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. Jul 2001;33(5):337-343. [doi: [10.3109/07853890109002087](https://doi.org/10.3109/07853890109002087)] [Medline: [11491192](#)]
49. Sun S, Chuang LH, Sahlén KG, Lindholm L, Norström F. Estimating a social value set for EQ-5D-5L in Sweden. *Health Qual Life Outcomes*. Dec 23, 2022;20(1):167. [FREE Full text] [doi: [10.1186/s12955-022-02083-w](https://doi.org/10.1186/s12955-022-02083-w)] [Medline: [36564844](#)]
50. Lindner P, Frykheden O, Forsström D, Andersson E, Ljótsson B, Hedman E, et al. The Brunnsviken Brief Quality of Life Scale (BBQ): development and psychometric evaluation. *Cogn Behav Ther*. Apr 2016;45(3):182-195. [FREE Full text] [doi: [10.1080/16506073.2016.1143526](https://doi.org/10.1080/16506073.2016.1143526)] [Medline: [26886248](#)]
51. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med*. Jul 2001;2(4):297-307. [doi: [10.1016/s1389-9457\(00\)00065-4](https://doi.org/10.1016/s1389-9457(00)00065-4)] [Medline: [11438246](#)]
52. Eriksson I, Undén AL, Elofsson S. Self-rated health. Comparisons between three different measures. Results from a population study. *Int J Epidemiol*. Apr 2001;30(2):326-333. [doi: [10.1093/ije/30.2.326](https://doi.org/10.1093/ije/30.2.326)] [Medline: [11369738](#)]
53. Nordin S, Palmquist E, Nordin M. Psychometric evaluation and normative data for a Swedish version of the Patient Health Questionnaire 15-Item Somatic Symptom Severity Scale. *Scand J Psychol*. Apr 07, 2013;54(2):112-117. [doi: [10.1111/sjop.12029](https://doi.org/10.1111/sjop.12029)] [Medline: [23294182](#)]
54. Gavelin HM, Boraxbekk CJ, Stenlund T, Järholm LS, Neely AS. Effects of a process-based cognitive training intervention for patients with stress-related exhaustion. *Stress*. 2015;18(5):578-588. [FREE Full text] [doi: [10.3109/10253890.2015.1064892](https://doi.org/10.3109/10253890.2015.1064892)] [Medline: [26305186](#)]
55. Öhman L, Nordin S, Bergdahl J, Slunga Birgander L, Stigsdotter Neely A. Cognitive function in outpatients with perceived chronic stress. *Scand J Work Environ Health*. Jun 2007;33(3):223-232. [FREE Full text] [doi: [10.5271/sjweh.1131](https://doi.org/10.5271/sjweh.1131)] [Medline: [17572832](#)]
56. Österberg K, Skogsliden S, Karlson B. Neuropsychological sequelae of work-stress-related exhaustion. *Stress*. Jan 28, 2014;17(1):59-69. [FREE Full text] [doi: [10.3109/10253890.2013.862615](https://doi.org/10.3109/10253890.2013.862615)] [Medline: [24188506](#)]
57. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. Jun 2000;31(2):73-86. [doi: [10.1016/s0005-7916\(00\)00012-4](https://doi.org/10.1016/s0005-7916(00)00012-4)] [Medline: [11132119](#)]
58. Smith A. Symbol Digit Modalities Test (SDMT). Manual (Revised). Los Angeles. Western Psychological Services; 1982.
59. Smith A. Symbol digit modalities test. *Clin Neuropsychol* 1973. 2022. [doi: [10.1037/t27513-000](https://doi.org/10.1037/t27513-000)]
60. Strober L, DeLuca J, Benedict RH, Jacobs A, Cohen JA, Chiaravalloti N, et al. Multiple Sclerosis Outcome Assessments Consortium (MSOAC). Symbol digit modalities test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis. *Mult Scler*. Nov 18, 2019;25(13):1781-1790. [FREE Full text] [doi: [10.1177/1352458518808204](https://doi.org/10.1177/1352458518808204)] [Medline: [30334474](#)]
61. Baune BT, Brignone M, Larsen KG. A network meta-analysis comparing effects of various antidepressant classes on the digit symbol substitution test (DSST) as a measure of cognitive dysfunction in patients with major depressive disorder. *Int J Neuropsychopharmacol*. Feb 01, 2018;21(2):97-107. [FREE Full text] [doi: [10.1093/ijnp/pyx070](https://doi.org/10.1093/ijnp/pyx070)] [Medline: [29053849](#)]
62. Franke Föyen L, Sennerstam V, Kontio E, Lekander M, Hedman-Lagerlöf E, Lindsäter E. Objective cognitive functioning in patients with stress-related disorders: a cross-sectional study using remote digital cognitive testing. *BMC Psychiatry*. Aug 07, 2023;23(1):565. [FREE Full text] [doi: [10.1186/s12888-023-05048-5](https://doi.org/10.1186/s12888-023-05048-5)] [Medline: [37550693](#)]
63. Spreen O, Benton AL. Neurosensory center comprehensive examination for aphasia. In: Kreutzer JS, DeLuca J, Caplan B, editors. *Encyclopedia of Clinical Neuropsychology*. Victoria, BC. Springer; 1977:1772-1773.
64. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. Dec 1935;18(6):643-662. [doi: [10.1037/h0054651](https://doi.org/10.1037/h0054651)]
65. Jensen AR, Rohwer WD. The Stroop color-word test: a review. *Acta Psychol (Amst)*. Jan 1966;25(1):36-93. [doi: [10.1016/0001-6918\(66\)90004-7](https://doi.org/10.1016/0001-6918(66)90004-7)] [Medline: [5328883](#)]
66. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn Psychol*. Aug 2000;41(1):49-100. [doi: [10.1006/cogp.1999.0734](https://doi.org/10.1006/cogp.1999.0734)] [Medline: [10945922](#)]
67. Ellbin S, Engen N, Jonsdottir IH, Nordlund AI. Assessment of cognitive function in patients with stress-related exhaustion using the Cognitive Assessment Battery (CAB). *J Clin Exp Neuropsychol*. Aug 05, 2018;40(6):567-575. [doi: [10.1080/13803395.2017.1388359](https://doi.org/10.1080/13803395.2017.1388359)] [Medline: [29105564](#)]
68. Jonsdottir IH, Nordlund A, Ellbin S, Ljung T, Glise K, Währborg P, et al. Cognitive impairment in patients with stress-related exhaustion. *Stress*. Mar 2013;16(2):181-190. [FREE Full text] [doi: [10.3109/10253890.2012.708950](https://doi.org/10.3109/10253890.2012.708950)] [Medline: [22746338](#)]
69. Nelson A, Gavelin HM, Boraxbekk CJ, Eskilsson T, Josefsson M, Slunga Järholm L, et al. Subjective cognitive complaints in patients with stress-related exhaustion disorder: a cross sectional study. *BMC Psychol*. May 18, 2021;9(1):84. [FREE Full text] [doi: [10.1186/s40359-021-00576-9](https://doi.org/10.1186/s40359-021-00576-9)] [Medline: [34006315](#)]
70. Rossetti HC, Munro Cullum C, Hynan LS, Lacritz LH. The CERAD neuropsychologic battery total score and the progression of Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2010;24(2):138-142. [FREE Full text] [doi: [10.1097/WAD.0b013e3181b76415](https://doi.org/10.1097/WAD.0b013e3181b76415)] [Medline: [20505431](#)]
71. Corsi PM. Memory and the Medial Temporal Region of the Brain. Doctoral Thesis in Philosophy. McGill University. 1972.
72. van Rossum G, Drake F. The Python language reference. Release 3.0.1 [Repr.]. Hampton, NH. Python Software Foundation; 2010.

73. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. Sep 16, 2020;585(7825):357-362. [[FREE Full text](#)] [doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)] [Medline: [32939066](#)]
74. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010. Presented at: SciPy '10; June 28-July 3, 2010:56; Austin, TX. URL: <https://proceedings.scipy.org/articles/Majora-92bf1922-00a> [doi: [10.25080/majora-92bf1922-00a](https://doi.org/10.25080/majora-92bf1922-00a)]
75. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Machine learning for neuroimaging with scikit-learn. *J Mach Learn Res*. 2011;12:2825-2830. [[FREE Full text](#)] [doi: [10.3389/fminf.2014.00014](https://doi.org/10.3389/fminf.2014.00014)] [Medline: [24600388](#)]
76. Open Science Framework. URL: <https://osf.io/> [accessed 2024-04-29]
77. Mindmore användarmanual. Version 1.1.25. Mindmore. URL: <https://s3.sto2.safedc.net/mindmore-user-manual/Mindmore%20Anvandarmanual.pdf> [accessed 2021-12-11]
78. van den Hurk W, Bergman I, Machado A, Bjermo J, Gustavsson A. Swedish normative data for mindmore: a comprehensive cognitive screening battery, both digital and self-administrated. *J Int Neuropsychol Soc*. May 24, 2021;28(2):188-202. [doi: [10.1017/s135561772100045x](https://doi.org/10.1017/s135561772100045x)]
79. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 3rd edition. Cambridge, MA. Shroff/O'Reilly; 2023.
80. Flygare O, Enander J, Andersson E, Ljótsson B, Ivanov VZ, Mataix-Cols D, et al. Predictors of remission from body dysmorphic disorder after internet-delivered cognitive behavior therapy: a machine learning approach. *BMC Psychiatry*. May 19, 2020;20(1):247. [[FREE Full text](#)] [doi: [10.1186/s12888-020-02655-4](https://doi.org/10.1186/s12888-020-02655-4)] [Medline: [32429939](#)]
81. Lenhard F, Sauer S, Andersson E, Månsson KN, Mataix-Cols D, Rück C, et al. Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *Int J Methods Psychiatr Res*. Mar 2018;27(1):e320. [[FREE Full text](#)] [doi: [10.1002/mpr.1576](https://doi.org/10.1002/mpr.1576)] [Medline: [28752937](#)]
82. Månsson KN, Frick A, Boraxbekk CJ, Marquand AF, Williams SC, Carlbring P, et al. Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl Psychiatry*. Mar 17, 2015;5(3):e530. [[FREE Full text](#)] [doi: [10.1038/tp.2015.22](https://doi.org/10.1038/tp.2015.22)] [Medline: [25781229](#)]
83. Giesemann J, Delgado J, Schwartz B, Bennemann B, Lutz W. Predicting dropout from psychological treatment using different machine learning algorithms, resampling methods, and sample sizes. *Psychother Res*. Jul 20, 2023;33(6):683-695. [doi: [10.1080/10503307.2022.2161432](https://doi.org/10.1080/10503307.2022.2161432)] [Medline: [36669124](#)]
84. Zantvoort K, Nacke B, Görlich D, Hornstein S, Jacobi C, Funk B. Estimation of minimal data sets sizes for machine learning predictions in digital mental health interventions. *NPJ Digit Med*. Dec 18, 2024;7(1):361. [[FREE Full text](#)] [doi: [10.1038/s41746-024-01360-w](https://doi.org/10.1038/s41746-024-01360-w)] [Medline: [39695276](#)]
85. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. *Psychother Res*. Jan 29, 2021;31(1):92-116. [doi: [10.1080/10503307.2020.1808729](https://doi.org/10.1080/10503307.2020.1808729)] [Medline: [32862761](#)]

## Abbreviations

- AD:** adjustment disorder
- ADNM-8:** Adjustment Disorder New Module-8
- BACC:** balanced accuracy
- CBT:** cognitive behavioral therapy
- CERAD:** Consortium to Establish a Registry for Alzheimer's Disease
- ED:** exhaustion disorder
- FAS:** Verbal Fluency Test
- ML:** machine learning
- PSS-10:** Perceived Stress Scale-10
- RCI:** reliable change index
- RCT:** randomized controlled trial
- RF:** random forest
- SVM:** support vector machine
- TRIPOD+AI:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence



*Edited by A Schwartz; submitted 26.08.24; peer-reviewed by RS Goma Mahmoud, G Bucur, M Wagner; comments to author 20.12.24; revised version received 22.01.25; accepted 19.02.25; published 25.03.25*

*Please cite as:*

*Franke Föyen L, Sennerstam V, Kontio E, Flygare O, Boman M, Lindsäter E*

*Predicting Therapy Outcomes in Patients With Stress-Related Disorders: Protocol for a Predictive Modeling Study*

*JMIR Res Protoc 2025;14:e65790*

URL: <https://www.researchprotocols.org/2025/1/e65790>

doi: [10.2196/65790](https://doi.org/10.2196/65790)

PMID:

©Ludwig Franke Föyen, Victoria Sennerstam, Evelina Kontio, Oskar Flygare, Magnus Boman, Elin Lindsäter. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.