JMIR Publications
The Leading eHealth Publisher

Predicting Therapy Outcomes for Stress-Related Disorders using Machine Learning: A Study Protocol

Ludwig Franke Föyen, Victoria Sennerstam, Evelina Kontio, Oskar Flygare, Magnus Boman, Elin Lindsäter

Submitted to: JMIR Research Protocols
on: August 26, 2024

Table of Contents

Predicting Therapy Outcomes for Stress-Related Disorders using Machine Learning: A Study Protocol

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

[1]Division of Psychology Department of Clinical Neuroscience Karolinska Institutet Stockholm SE
[2]Stress Research Institute Department of Psychology Stockholm University Stockholm SE
[3]Gustavsberg University Primary Care Center Academic Primary Care Center Region Stockholm Stockholm SE
[4]Department of Clinical Neuroscience Osher Center for Integrative Health Karolinska Institutet Stockholm SE
[5]Centre for Psychiatry Research Department of Clinical Neuroscience Karolinska Institutet and Stockholm Health Care Services Stockholm SE
[6]BioClinicum MedTechLabs Karolinska University Hospital Stockholm SE
[7]Division of Psychiatry University College London London GB
[8]Division of Clinical Epidemiology Department of Medicine Solna Karolinska Institutet Stockholm SE

Corresponding Author:
Ludwig Franke Föyen MSc
Division of Psychology
Department of Clinical Neuroscience
Karolinska Institutet
Nobels väg 9
Stockholm
SE

Abstract

Background: Stress-related disorders, such as adjustment disorder and exhaustion disorder, are prevalent and associated with significant personal suffering and societal costs. While cognitive behavioral therapy has shown efficacy in treating these conditions, knowledge about factors contributing to treatment response is limited. Improved identification of such factors could enhance assessment procedures and treatment strategies.

Objective: This study aims to (1) evaluate putative predictors of treatment outcome in patients with stress-related disorders using traditional prediction methods and (2) model treatment outcomes utilizing a machine learning approach. The primary outcome of interest is responder status on the Perceived Stress Scale-10, evaluated based on the reliable change index post-treatment.

Methods: Data from a randomized controlled trial comparing two internet-delivered treatments for patients diagnosed with adjustment disorder or exhaustion disorder (N = 300) will be analyzed. Putative predictors include sociodemographic and clinical information, clinician-assessed data, self-rated symptoms, and cognitive test scores. 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, four 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 to be used in the present study was collected between April 2021 and July 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 67% or higher, thus being considered clinically useful.

Conclusions: This study will contribute to the limited research on predictors of treatment outcome in stress-related disorders. By comparing traditional and machine learning approaches, it aims to enhance our understanding of factors influencing treatment response. 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 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. Clinical Trial: ClinicalTrials ID: NCT04797273. Trial registration

date 15 March 2021.

(JMIR Preprints 26/08/2024:65790)
DOI: https://doi.org/10.2196/preprints.65790

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Predicting Therapy Outcomes for Stress-Related Disorders using Machine Learning: A Study Protocol

Ludwig Franke Föyen*[1,2,3]; Victoria Sennerstam [1,3,4]; Evelina Kontio[1,3]; Oskar Flygare, PhD [5]; Magnus Boman, Prof [6,7,8]; Elin Lindsäter, PhD [1,3,5]

[1] Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet, Stockholm, Sweden.
[2] Department of Psychology, Stress Research Institute, Stockholm University, Stockholm, Sweden
[3] Academic Primary Care Center, Gustavsberg University Primary Care Center, Region Stockholm, Sweden
[4] Osher Center for Integrative Health, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.
[5] Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet and Stockholm Health Care Services, Stockholm, Sweden
[6] Department of Medicine Solna, Division of Clinical Epidemiology, Karolinska Institutet, Stockholm, Sweden.
[7] MedTechLabs, BioClinicum, Karolinska University Hospital, Stockholm, Sweden
[8] University College London, Division of Psychiatry, London, UK

*Corresponding author: Ludwig.franke.foyen@ki.se, +46 70 497 73 91, Nobels väg 9, 171 65, Stockholm, Sweden

Abstract

Background

Stress-related disorders, such as adjustment disorder and exhaustion disorder, are prevalent and associated with significant personal suffering and societal costs. While cognitive behavioral therapy has shown efficacy in treating these conditions, knowledge about factors contributing to treatment response is limited. Improved identification of such factors could enhance assessment procedures and treatment strategies.

Objectives

This study aims to (1) evaluate putative predictors of treatment outcome in patients with stress-related disorders using traditional prediction methods and (2) model treatment outcomes utilizing a machine learning approach. The primary outcome of interest is responder status on the Perceived Stress Scale-10, evaluated based on the reliable change index post-treatment.

Methods

Data from a randomized controlled trial comparing two internet-delivered treatments for patients diagnosed with adjustment disorder or exhaustion disorder (N = 300) will be analyzed. Putative predictors include sociodemographic and clinical information, clinician-assessed data, self-rated symptoms, and cognitive test scores. 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, four 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 to be used in the present study was collected between April 2021 and July 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 67% or higher, thus being considered clinically useful.

Conclusions

This study will contribute to the limited research on predictors of treatment outcome in stress-related disorders. By comparing traditional and machine learning approaches, it aims to enhance our understanding of factors influencing treatment response. 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 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.

Trial Registration

ClinicalTrials ID: NCT04797273. Trial registration date 15 March 2021.

Keywords

Adjustment Disorder, Cognitive Behavioral Therapy, Exhaustion Disorder, Machine Learning, Predictive Modeling, Stress-Related Disorders, Therapy Outcomes

Introduction

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 the majority of psychiatric long-term sickness absences.[3] In Sweden, disorders believed to stem from persistent or overwhelming sub-traumatic 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-10 (ICD-10), 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 sub-traumatic life events (stressors), with 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 what factors contribute to treatment response is still limited.[8,12] Improved identification of factors that contribute to symptom development could potentiate 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 (e.g. age and education level),[16,17] clinical characteristics (e.g. use of medication and symptom severity),[17,18] treatment-related factors (e.g. treatment credibility and adherence)[16,18] and cognitive functioning[19] 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.[20] 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.[21] However, the effects of demographics and pre-treatment 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.[22,23]

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, high-dimensional data characteristic of mental disorders.[24] With this approach, a model generates a prediction (e.g. remission, yes/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.[25,26]

Forsell et al.[27] have proposed a balanced accuracy 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 evolving area of inquiry, and its capacity to surpass conventional methods in clinical usefulness is yet to be established.[28]

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

Objective of the study

The objective of this study is to predict treatment outcomes in patients with stress-related disorders. To this end, the current study aims to firstly evaluate putative predictors using a traditional prediction paradigm, and secondly to model treatment outcomes using a ML approach. Our primary outcome of interest is responder status post treatment on the Perceived Stress Scale (PSS-10), evaluated using the reliable change index (RCI; further described in a later section). Based on 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 balanced accuracy (BACC) of 67% or higher, thus being indicated clinically useful.[27]

Methods

Study Design

We will use collected data from a randomized controlled trial (RCT) 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 (GHP). 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 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) + AI statement.[29] The study was pre-registered on ClinicalTrials ID: NCT04797273 and was approved by the Swedish Ethical Review Authority (DNR 2020–03198; 2023-06857-02).

Procedure

Participants

Study inclusion and exclusion criteria

Three hundred nationally recruited individuals were diagnosed with a primary diagnosis of AD (*n* = 142) or ED (*n* = 158) and were included in the RCT. Participant recruitment was carried out through social media, by ads in the newspaper, and by information given to healthcare clinics. Participants self-referred to the study webpage, 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 (MINI),[30] self-rated exhaustion disorder (S-ED)[31] and the Adjustment Disorder New Module-8 (ADNM-8).[32] For inclusion, participants needed to (1) fulfill the criteria for a primary diagnosis of AD or ED, (2) be aged between 18-65 years old, (3) have regular access to a computer with internet access, and (4) be able to read and write in Swedish. 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 of two 12-week internet-delivered treatments (CBT and GHP). They both consisted of online text-based modules with related exercises and assignments. Patients were guided sequentially through the modules by a therapist via a secure online 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. A full description of the treatments is described by Sennerstam et al. (forthcoming) and an overview of each module in respective treatment will be presented in an online supplement. Due to the limited sample size and knowledge on predictors of treatment all patients regardless of treatment assignment will be included in this study.

Outcomes

The primary outcome in this study and the original RCT is the Perceived Stress Scale-10 (PSS-10).[33] 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 - 40. The PSS is the most commonly used outcome measure of stress-management interventions globally (see, e.g.[8,34]). For the present study, a Swedish version of the PSS-10 was used. The PSS-10 has been found to exhibit high internal consistency (Cronbach's α .84) and adequate construct validity.[35] The PSS-10 was administered digitally through the online study platform before randomization to treatment, every three 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 the purpose of this study, the sum score of the PSS-10 will be dichotomized into responder/non-responder post-treatment based on the Reliable Change Index (RCI) criteria[36] 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 a predictor.

Putative predictors

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

Sociodemographic variables

Information on age (interval), sex (male/female/other or prefer not to disclose), relationship status (in relationship, single, widowed), number of children, educational attainment (in nine categories between less than nine years of school to PhD), employment status (e.g., student, unemployed, full-time work), and employment type (in 11 categories, e.g. employed in the private sector, by the municipality) was gathered before the start of treatment using the online study platform. Self-rated reading and computer skills were rated separately on a five-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, yes/no). Primary diagnosis (AD/ED), possible secondary psychiatric diagnosis (e.g. anxiety or depression) was assessed by the clinician using MINI, S-ED (ordinal categories ranging from '*No*' to '*Yes - to a high degree*')[31] and the ADNM-8.[32] Using ADNM-8, the patient was asked about which specific stressors had been present in the past two years (in 16 options, e.g. '*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 five steps), length of current sick leave episode (0-1 month - >12 months in five categories) was self-reported.

Self-rated symptoms

Alcohol use was assessed using the Alcohol Use Disorder Identification Test (AUDIT).[37,38] 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 (GAD-7) scale.[39] This screening tool assesses generalized anxiety symptoms over the past two weeks. It is comprised of seven 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 (SMBQ-18).[40,41] It aims to measure three 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 (MADRS-S).[42] 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 nine item Karolinska Exhaustion Disorder Scale (KEDS).[43] Measuring different aspects of exhaustion such as fatigue, endurance, and sleep impairment, answers are rated on an ordinal scale from 0-6 (e.g. 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 (WHODAS 2.0).[44] WHODAS was developed to assess functioning in the last 30 days in six 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 "*Extreme/ Unable*". A 12-item version was used.

Quality of life was assessed using the EQ-5D-5L[45,46] and the Brunnsviken Brief Quality of Life Scale.[47] The EQ-5D-5L contains five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression each rated on five levels of severity from "*No problems*" to "*Extreme problems*". The Brunnsviken Brief Quality of Life scale (BBQ) is a 12-item questionnaire that assesses six 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 (ISI).[48] The ISI 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 0-4.

Self-rated health was assessed using SRH-5 asking patients to rate their general health on a scale of 1, "*Very bad*" to 5 "*Very good*".[49]

Somatoform symptoms were assessed using the Patient Health Questionnaire (PHQ-15).[50] 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).[51] A 5-item version has previously been used to assess subjective memory problems in this patient population.[52,53] 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.[54] 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*".

The *treatment assignment* in the RCT (ICBT & GHP) will also be used as a putative predictor.

Cognitive functioning

Attention and processing speed was measured using the Symbol Digit Modality Test (SDMT). A test originally developed by Smith [55,56] that measures visual detection, attention and processing speed. A key with nine different symbols and matching numbers is shown on the upper part of the display. At the center one of these nine symbols is shown and the task of the participant is choosing 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 MS[57] and depression,[58] and it has been used in patients with stress-related disorders.[59]

Executive functioning was measured using the FAS Word Fluency Test and the Stroop test. FAS was first described by Spreen & Benton,[60] 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. 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.[59]

The Stroop test, originally developed by Stroop[61] and described by Jensen & Rohwer,[62] measures executive functioning, inhibition, as well as updating and processing speed.[63] The test has two parts, (1) twenty color words are presented (green, yellow, blue or red) and they are colored congruent to their meaning (e.g., the word red colored in red). In the bottom part of the display, the color words are displayed on four 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 (e.g., 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 two/average time in seconds from part two) 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 two studies,[59,64] but not in others.[65,66]

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's disease[67] 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 three 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 5-10 minutes, the participant is asked to recall the words. Test score for the learning time is number of correct words in trial 1-3, 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.[59,66]

Corsi block-tapping test forward gives information about visual ability of attention, short term memory and working memory. It contains two 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 two 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.[59]

Table 1.
Putative Predictors of Treatment Outcome in Stress-Related Disorders

Type of predictor	Predictor	Construct measured	Type	Clinician-rated	Scoring range
Sociodemographics	Age		Interval		18-65
	Sex		Categorical		Male/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/No
Clinical characteristics	Number of medications	Medication	Interval	x	0-4
	Antidepressants	Medication	Categorical	x	Yes/No
	Sleep medication	Medication	Categorical	x	Yes/No
	Pain medication	Medication	Categorical	x	Yes/No
	Anxiolytics	Medication	Categorical	x	Yes/No
	Diagnosis	Primary diagnosis	Categorical	x	2 Categories
	Secondary diagnosis	Secondary diagnosis	Interval	x	0-4
	Depression	Secondary diagnosis	Categorical	x	Yes/No
	Anxiety disorder	Secondary diagnosis	Categorical	x	Yes/No
	Insomnia	Secondary diagnosis	Categorical	x	Yes/No
	Other disorders	Secondary diagnosis	Categorical	x	Yes/No
	S-ED	Exhaustion disorder	Ordinal	x	3 Categories
	ADNM-8 criteria	Adjustment disorder	Categorical	x	Yes/No
	ADNM-8 no. stressors	Adjustment disorder	Interval	x	0-11
	ADNM- stressors	Adjustment disorder	Categorical	x	16 categories
	Duration of current episode		Interval	x	0-∞
	Age of first episode		Interval	x	0-65
	Sick-leave status	Sickness Absence	Interval		0-100% 5 steps
	Sick-leave duration	Sickness Absence	Ordinal		5 categories
Self-rated symptoms	AUDIT	Alcohol consumption	Interval		0-40
	GAD-7	Anxiety symptoms	Interval		0-21
	SMBQ cognitive weariness	Burnout	Continuous		0-7
	SMBQ exhaustion	Burnout	Continuous		0-7
	SMBQ listlessness	Burnout	Continuous		0-7
	MADRS-S	Depression	Interval		0-54
	KEDS	Exhaustion disorder	Interval		0-54
	WHODAS 2.0	Functional disability	Continuous		0-100%
	EQ-5D-5L	Quality of Life	Interval		5-25
	BBQ	Quality of life	Interval		0-96
	ISI	Insomnia severity	Interval		0-28
	SRH-5	Self-Rated Health	Interval		0-5
	PSS-10	Perceived stress	Interval		0-40
	PHQ-15	Somatoform symptoms	Interval		0-30
	6-QEMP	Subjective memory impairment	Interval		0-30
3-week measurement	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
Treatment related predictors	Clinician treatment expectancy		Interval	x	0-10
	Treatment credibility scale		Interval		0-10
	Treatment assignment		ICBT/GHP		2 Categories
Cognitive functioning	SDMT	Attention and processing speed	Interval		0-∞
	FAS	Executive functions	Interval		0-∞

Stroop index	Executive functions	Continuous	0-∞
Stroop inhibition	Executive functions	Continuous	0-∞
CERAD learning	Memory	Interval	0-30
CERAD recognition	Memory	Interval	0-10
Corsi Forward	Memory	Interval	0-9

Note. Abbreviations. 6-QEMP, 6-item questionnaire of everyday memory problems; ADNM-8, The Adjustment Disorder New Module-8, BBQ, Brunnsviken Brief Quality of Life Scale; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; EQ-5D-5L, EuroQol 5-Dimension 5-Level; FAS, Verbal Fluency Test; GAD-7, General Anxiety Disorder-7; GHP, General Health Promoting advice; ICBT, Internet-Based Cognitive Behavioral Therapy; ISI, Insomnia Severity Index; KEDS, Karolinska Exhaustion Disorder Scale; MADRS-S, Montgomery-Åsberg Depression Rating Scale - Self report; PHQ-15, Patient Health Questionnaire-15; PSS-10, Perceived Stress Scale; S-ED, Self-Rated Exhaustion Disorder (clinician-assessed); SDMT, Symbol Digit Modality Test; SMBQ-18, Shirom-Melamed Burnout Questionnaire-18; SRH5, Self-Rated Health 5; WHODAS 2.0, The World Health Organization Disability Assessment Schedule 2.0.

Planned Statistical Analysis

All data will be prepared and analyzed using Python,[68] and the libraries NumPy,[69] Pandas[70] and scikit-learn[71] or equivalent statistical packages. A notebook containing the analysis in documented code will be made available on Open Science Framework (OSF.io) 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. Additionally, 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 three 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 ≧.8 will be removed from further analysis. Data that is highly skewed will be transformed if deemed appropriate.

Cognitive test results will be manually reviewed prior to model fitting to validate a proper result. Comments pertaining to technical difficulties and/or 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 Foyen et al.,[59] and for a full overview of the multiple linear regression models used and how they were calculated, see [72,73].

Patients who have missing data for the post-treatment PSS-10, i.e. 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, reliable change index (RCI) for the PSS-10 pre- to post- treatment will be computed using the following formula: $RCI = ¿¿$. $SEM = SD \star \sqrt{1-\alpha}$.[36] Cronbach's α = .83 from normative data will be used.[35] 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 pre-treatment variables, including mean/median, standard deviations and IQR range 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 non-parametric 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.

Machine learning model development

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/ED) and responder status. 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 (also called hold-out sample). The choice of 70-30 was due to the limited size of our sample, as fewer observations in the testing data makes it difficult to utilize uncommon predictors. No external validation set is currently available at the time of writing.

Preprocessing pipeline

The preprocessing pipeline 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 four 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. 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 ensambling these predictions. The support vector machine (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. 5-fold cross-validation is done by partitioning the data into five subsets, training the model on four subsets, and validating it on the remaining subset. This process is repeated five 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 0.0 to 1.0. 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 balanced accuracy (BACC), precision and recall, both in the train-set obtained through k-fold cross-validation and in the test set. Predictor importance in the Random Forest 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 (AUC) will be utilized 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.[27] Further, 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 confidence intervals 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). Dr Flygare is supported by the Swedish innovation agency (no. 2022-00549). All data was collected (*N* = 300) between April 2021 and July 2022. A cross-sectional study investigating baseline cognitive functioning as compared to normative material has been published,[59] and an interim analysis of pre- and post-comparisons was presented at a conference in September 2022. As of August 2024, data has not been analyzed for this study.

Discussion

Strengths and limitations

This study will use a high-quality dataset from an RCT to investigate potential treatment predictors using both traditional prediction methods and a machine learning paradigm. This dual approach will enable the identification of putative predictors of treatment response in patient populations where prior research is limited. Additionally, it will facilitate comparisons between different methodological approaches to prediction research. A limitation is that the available sample size is modest (*N* = 300) which may increase the risk of under/overfitting and thus limit conclusions that can be drawn. Further, machine learning models such as RF, in including such a multitude of variables, risk being difficult to interpret and integrate in clinical practice.

Implications for 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. It could support the development of more personalized and effective interventions, addressing a critical gap in mental health care. The investigation into machine learning models for identifying important treatment predictors may encourage future larger-scale studies and, potentially, the implementation of these models in clinical settings. This approach could enhance precision in mental health care by capturing the complexity of psychiatric disorders and identifying key factors for treatment response. Ultimately, this multifaceted analysis of predictors could lead to more targeted and efficacious interventions, improving outcomes for patients with stress-related disorders.

Acknowledgements

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

Conflicts of interest

The authors of this study protocol have no competing interests to declare.

References

1. Hassard J, Teoh K, Visockaite G, Dewe P, Cox T. The cost of work-related stress: a systematic review. *J Occup Health Psychol*. 2017;23. doi:10.1037/ocp0000069

2. Cullen AE, Lindsäter E, Rahman S, 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*. 2023;9(2):e40. doi:10.1192/bjo.2023.8

3. Swedish Social Insurance Agency. *Sjukfrånvaro i psykiatriska diagnoser - En registerstudie av Sveriges arbetande befolkning i åldern 20–69 år*. Försäkringskassan, Analys och rapport; 2020:108. https://www.forsakringskassan.se/!ut/p/z0/LcyxCsMwDATQb-ngUdihnbrlL0KWImI1NnYkI5mU_n0d6HS8Ozi_-sWvjGfesWdhrMPLZzueDXty0312YRPuxN3opWRN2PJJLrQB-8dBMVJFjsQuWL8qHe6kGKHZt2QrIAmrITwmaCrbOIQkuxLoNb5FoZyZWdS3Mt9-RkLcrQ!!/

4. Guseva Canu I, Marca SC, Dell'Oro F, et al. Harmonized definition of occupational burnout: A systematic review, semantic analysis, and Delphi consensus in 29 countries. *Scand J Work Environ Health*. 2021;47(2):95-107. doi:10.5271/sjweh.3935

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*. 2017;4:1-11. doi:10.1016/j.burn.2017.02.001

6. Bachem R, Casey P. Adjustment disorder: A diagnosis whose time has come. *J Affect Disord*. 2018;227:243-253. doi:10.1016/j.jad.2017.10.034

7. Lindsäter E, Svärdman F, Wallert J, et al. Exhaustion disorder: scoping review of research on a recently introduced stress-related diagnosis. *BJPsych Open*. 2022;8(5):e159. doi:10.1192/bjo.2022.559

8. Svärdman F, Sjöwall D, Lindsäter E. Internet-delivered cognitive behavioral interventions to reduce elevated stress: A systematic review and meta-analysis. *Internet Interv*. 2022;29:100553. doi:10.1016/j.invent.2022.100553

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*. 2016;18(1):e21-e21. doi:10.2196/JMIR.5112

10. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: A meta-analysis. *J Occup Health Psychol*. 2008;13(1):69-93. doi:10.1037/1076-8998.13.1.69

11. Arends I, Bruinvels DJ, Rebergen DS, et al. Interventions to facilitate return to work in adults with adjustment disorders. *Cochrane Database Syst Rev*. 2012;12:CD006389. doi:10.1002/14651858.CD006389.pub2

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*. 2021;28(2):313-324. doi:10.1002/cpp.2518

13. Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3(3):223-230. doi:10.1016/j.bpsc.2017.11.007

14. Meehan AJ, Lewis SJ, Fazel S, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Mol Psychiatry*. 2022;27(6):2700-2708. doi:10.1038/s41380-022-01528-4

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*. 2014;9(5):e97303. doi:10.1371/journal.pone.0097303

16. Salomonsson S, Santoft F, Lindsäter E, et al. Predictors of outcome in guided self-help cognitive behavioural therapy for common mental disorders in primary care. *Cogn Behav Ther*. 2020;49(6):455-474. doi:10.1080/16506073.2019.1669701

17. Wallert J, Boberg J, Kaldo V, et al. Predicting remission after internet-delivered psychotherapy in patients with depression using machine learning and multi-modal data. *Transl Psychiatry*. 2022;12(1):357. doi:10.1038/s41398-022-02133-3

18. Prasad N, Chien I, Regan T, et al. Deep learning for the prediction of clinical outcomes in

internet-delivered CBT for depression and anxiety. *PLOS ONE*. 2023;18(11):e0272685. doi:10.1371/journal.pone.0272685

19. Groves SJ, Douglas KM, Porter RJ. A Systematic Review of Cognitive Predictors of Treatment Outcome in Major Depression. *Front Psychiatry*. 2018;9:382. doi:10.3389/fpsyt.2018.00382

20. 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*. 2020;274:949-954. doi:10.1016/j.jad.2020.05.139

21. Clason van de Leur J, Johansson F, McCracken LM, Åhs F, Brodda Jansen G, Buhrman M. Predictors and sub-groups in the treatment of stress-induced exhaustion disorder. *Cogn Behav Ther*. 2023;0(0):1-22. doi:10.1080/16506073.2023.2197148

22. Chekroud AM, Bondar J, Delgadillo J, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*. 2021;20(2):154-170. doi:10.1002/wps.20882

23. 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*. 2022;97:102193. doi:10.1016/j.cpr.2022.102193

24. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol*. 2018;14(1):91-118. doi:10.1146/annurev-clinpsy-032816-045037

25. Bjurner P, Isacsson NH, Abdesslem FB, Boman M, Forsell E, Kaldo 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. Published online April 17, 2024. Accessed July 8, 2024. https://osf.io/cs4bx/

26. Forsell E, Jernelöv S, Blom K, 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*. 2019;176(4):315-323. doi:10.1176/appi.ajp.2018.18060699

27. Forsell E, Jernelöv S, Blom K, Kaldo 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*. 2022;29:100554. doi:10.1016/j.invent.2022.100554

28. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Calster BV. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004

29. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Published online April 16, 2024:e078378. doi:10.1136/bmj-2023-078378

30. Lecrubier Y, Sheehan DV, Weiller E, et al. Mini International Neuropsychiatric Interview. Published online July 8, 2013. doi:10.1037/t18597-000

31. Glise K, Hadzibajramovic E, Jonsdottir IH, Ahlborg 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*. 2010;83(5):511-520. doi:10.1007/s00420-009-0490-x

32. 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. doi:10.1159/000484415

33. Cohen S. Perceived stress in a probability sample of the United States. In: *The Social Psychology of Health*. The Claremont Symposium on Applied Social Psychology. Sage Publications, Inc; 1988:31-67.

34. Lee EH. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res*. 2012;6(4):121-127. doi:10.1016/j.anr.2012.08.004

35. Nordin M, Nordin S. Psychometric evaluation and normative data of the Swedish version of the 10-item perceived stress scale. *Scand J Psychol*. 2013;54(6):502-507. doi:10.1111/sjop.12071

36. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12-19. doi:10.1037//0022-006x.59.1.12

37. 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. *Addict Abingdon Engl*. 1993;88(6):791-804. doi:10.1111/j.1360-0443.1993.tb02093.x

38. Bergman H, Källmén H. Alcohol use among Swedes and a psychometric evaluation of the alcohol use disorders identification test. *Alcohol Alcohol Oxf Oxfs*. 2002;37(3):245-251. doi:10.1093/alcalc/37.3.245

39. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092

40. Lundgren-Nilsson Å, Jonsdottir IH, Pallant J, Ahlborg G. Internal construct validity of the Shirom-Melamed Burnout Questionnaire (SMBQ). *BMC Public Health*. 2012;12(1):1. doi:10.1186/1471-2458-12-1

41. Melamed S, Kushnir T, Shirom A. Burnout and Risk Factors for Cardiovascular Diseases. *Behav Med*. 1992;18(2):53-60. doi:10.1080/08964289.1992.9935172

42. Montgomery SA, Åsberg M. A New Depression Scale Designed to be Sensitive to Change. *Br J Psychiatry*. 1979;134(4):382-389. doi:10.1192/bjp.134.4.382

43. 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*. 2014;55(1):72-82. doi:10.1111/sjop.12088

44. Üstün TB, Chatterji S, Kostanjsek N, et al. Developing the World Health Organization Disability Assessment Schedule 2.0. *Bull World Health Organ*. 2010;88(11):815-823. doi:10.2471/BLT.09.067231

45. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337-343. doi:10.3109/07853890109002087

46. 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*. 2022;20(1):167. doi:10.1186/s12955-022-02083-w

47. Lindner P, Frykheden O, Forsström D, et al. The Brunnsviken Brief Quality of Life Scale (BBQ): Development and Psychometric Evaluation. *Cogn Behav Ther*. 2016;45(3):182-195. doi:10.1080/16506073.2016.1143526

48. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med*. 2001;2(4):297-307. doi:10.1016/s1389-9457(00)00065-4

49. Eriksson I, Undén AL, Elofsson S. Self-rated health. Comparisons between three different measures. Results from a population study. *Int J Epidemiol*. 2001;30(2):326-333. doi:10.1093/ije/30.2.326

50. 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*. 2013;54(2):112-117. doi:10.1111/sjop.12029

51. Stigsdotter Neely. Everyday Memory Problems Questionnaire and stress related disorders: Personal Communication.

52. 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*. 2007;33(3):223-232. doi:10.5271/sjweh.1131

53. Österberg K, Skogsliden S, Karlson B. Neuropsychological sequelae of work-stress-related exhaustion. *Stress*. 2014;17(1):59-69. doi:10.3109/10253890.2013.862615

54. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. 2000;31(2):73-86. doi:10.1016/s0005-7916(00)00012-4

55. Smith A. Symbol Digit Modalities Test (SDMT). Manual (Revised). Published online 1982.

56. Smith A. Symbol digit modalities test. *Clin Neuropsychol*. Published online 1973.

57. Strober L, DeLuca J, Benedict RH, et al. Symbol Digit Modalities Test: A valid clinical trial endpoint for measuring cognition in multiple sclerosis. *Mult Scler J*. 2019;25(13):1781-1790. doi:10.1177/1352458518808204

58. 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*. 2018;21(2):97-107. doi:10.1093/ijnp/pyx070

59. 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*. 2023;23(1):565. doi:10.1186/s12888-023-05048-5

60. Spreen O, Benton AL. *Neurosensory Center Comprehensive Examination for Aphasia*. Victoria BC: University of Victoria, Neuropsychology Laboratory; 1977.

61. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935;18(6):643-662. doi:10.1037/h0054651

62. Jensen AR, Rohwer WD. The stroop color-word test: A review. *Acta Psychol (Amst)*. 1966;25:36-93. doi:10.1016/0001-6918(66)90004-7

63. 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. *Cognit Psychol*. 2000;41(1):49-100. doi:10.1006/cogp.1999.0734

64. Ellbin S, Engen N, Jonsdottir IH, Nordlund AIK. Assessment of cognitive function in patients with stress-related exhaustion using the Cognitive Assessment Battery (CAB). *J Clin Exp Neuropsychol*. 2018;40(6):567-575. doi:10.1080/13803395.2017.1388359

65. Jonsdottir IH, Nordlund A, Ellbin S, et al. Cognitive impairment in patients with stress-related exhaustion. *Stress Amst Neth*. 2013;16(2):181-190. doi:10.3109/10253890.2012.708950

66. Nelson A, Gavelin HM, Boraxbekk CJ, et al. Subjective cognitive complaints in patients with stress-related exhaustion disorder: a cross sectional study. *BMC Psychol*. 2021;9(1):84. doi:10.1186/s40359-021-00576-9

67. Rossetti HC, Munro Cullum C, Hynan LS, Lacritz L. The CERAD Neuropsychological Battery Total Score and the Progression of Alzheimer’s Disease. *Alzheimer Dis Assoc Disord*. 2010;24(2):138-142. doi:10.1097/WAD.0b013e3181b76415

68. Rossum G van, Drake FL. *The Python Language Reference*. Release 3.0.1 [Repr.]. Python Software Foundation; 2010.

69. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2

70. McKinney W. Data Structures for Statistical Computing in Python. In: Walt S van der, Millman J, eds. *Proceedings of the 9th Python in Science Conference, Austin*. ; 2010:56-61. doi:10.25080/Majora-92bf1922-00a

71. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.

72. Mindmore. Mindmore användarmanual Version 1.1.25. Published online 2022. Accessed December 11, 2021. https://s3.sto2.safedc.net/mindmore-user-manual/Mindmore%20Anvandarmanual.pdf

73. 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*. Published online May 24, 2021:1-15. doi:10.1017/S135561772100045X

Abbreviations

AD	Adjustment Disorder
ADNM-8	Adjustment Disorder New Module-8
AI	Artificial Intelligence
AUC	Area Under the Curve
AUDIT	Alcohol Use Disorder Identification Test
BACC	Balanced Accuracy
BBQ	Brunnsviken Brief Quality of Life Scale
CBT	Cognitive Behavioral Therapy
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
ED	Exhaustion Disorder
EQ-5D-5L	EuroQol 5-Dimension 5-Level
FAS	Verbal Fluency Test
GAD-7	General Anxiety Disorder-7
GHP	General Health-Promoting advice
ICBT	Internet-delivered Cognitive Behavioral Therapy
ICD	International Classification of Diseases
IQR	Interquartile Range
ISI	Insomnia Severity Index
KEDS	Karolinska Exhaustion Disorder Scale
KNN	K-Nearest Neighbors
LogReg	Logistic Regression
MADRS-S	Montgomery-Åsberg Depression Rating Scale - Self report
MINI	Mini International Neuropsychiatric Interview
ML	Machine Learning
PHQ-15	Patient Health Questionnaire-15
PSS-10	Perceived Stress Scale-10
QEMP	Questionnaire of Everyday Memory Problems
RCI	Reliable Change Index
RCT	Randomized Controlled Trial
RF	Random Forest
SDMT	Symbol Digit Modality Test
SEM	Standard Error of Measurement
SMBQ-18	Shirom-Melamed Burnout Questionnaire-18
SRH5	Self-Rated Health 5
SVM	Support Vector Machine
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
WHODAS 2.0	World Health Organization Disability Assessment Schedule 2.0