

BMJ Open Design and methods of the research unit 5187 PREACT (towards precision psychotherapy for non-respondent patients: from signatures to predictions to clinical utility) – a study protocol for a multicentre observational study in outpatient clinics

Till Langhammer ¹, Chantal Unterfeld,¹ Felix Blankenburg,^{2,3} Susanne Erk,⁴ Lydia Fehm,¹ John-Dylan Haynes,⁴ Stephan Heinzel,⁵ Kevin Hilbert,⁶ Frank Jacobi,⁷ Norbert Kathmann,¹ Christine Knaevelsrud,^{8,9} Babette Renneberg,^{9,10} Kerstin Ritter,^{4,11} Nikola Stenzel,⁷ Henrik Walter,⁴ Ulrike Lueken ^{1,9}

To cite: Langhammer T, Unterfeld C, Blankenburg F, *et al.* Design and methods of the research unit 5187 PREACT (towards precision psychotherapy for non-respondent patients: from signatures to predictions to clinical utility) – a study protocol for a multicentre observational study in outpatient clinics. *BMJ Open* 2025;**15**:e094110. doi:10.1136/bmjopen-2024-094110

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-094110>).

Received 23 September 2024
Accepted 17 January 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Till Langhammer;
till.langhammer@hu-berlin.de

ABSTRACT

Introduction Cognitive-behavioural therapy (CBT) works—but not equally well for all patients. Less than 50% of patients with internalising disorders achieve clinically meaningful improvement, with negative consequences for patients and healthcare systems. The research unit (RU) 5187 seeks to improve this situation by an in-depth investigation of the phenomenon of treatment non-response (TNR) to CBT. We aim to identify bio-behavioural signatures associated with TNR, develop predictive models applicable to individual patients and enhance the utility of predictive analytics by collecting a naturalistic cohort with high ecological validity for the outpatient sector.

Methods and analysis The RU is composed of nine subprojects (SPs), spanning from clinical, machine learning and neuroimaging science and service projects to particular research questions on psychological, electrophysiological/autonomic, digital and neural signatures of TNR. The clinical study SP 1 comprises a four-centre, prospective-longitudinal observational trial where we recruit a cohort of 585 patients with a wide range of internalising disorders (specific phobia, social anxiety disorder, panic disorder, agoraphobia, generalised anxiety disorder, obsessive-compulsive disorder, post-traumatic stress disorder, and unipolar depressive disorders) using minimal exclusion criteria. Our experimental focus lies on emotion (dys)-regulation as a putative key mechanism of CBT and TNR. We use state-of-the-art machine learning methods to achieve single-patient predictions, incorporating pretrained convolutional neural networks for high-dimensional neuroimaging data and multiple kernel learning to integrate information from various modalities. The RU aims to advance precision psychotherapy by identifying emotion regulation-based biobehavioural markers of TNR, setting up a multilevel assessment for optimal predictors

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Prospective design specifically tailored for predicting psychotherapy outcomes.
- ⇒ Large, naturalistic sample from multiple outpatient clinics enhances ecological validity.
- ⇒ Multimodal data collection (clinical, ambulatory, neuroimaging and psychophysiological) and state-of-the-art algorithms provide comprehensive insights.
- ⇒ Extensive assessments may increase participant burden, potentially affecting adherence.
- ⇒ Exclusion of hard-to-reach patient groups that do not present at outpatient clinics may limit the generalisability of the results.

and using an ecologically valid sample to apply findings in diverse clinical settings, thereby addressing the needs of vulnerable patients.

Ethics and dissemination The study has received ethical approval from the Institutional Ethics Committee of the Department of Psychology at Humboldt-Universität zu Berlin (approval no. 2021-01) and the Ethics Committee of Charité-Universitätsmedizin Berlin (approval no. EA1/186/22).

Results will be disseminated through peer-reviewed journals and presentations at national and international conferences. Deidentified data and analysis scripts will be made available to researchers within the RU via a secure server, in line with ethical guidelines and participant consent. In compliance with European and German data protection regulations, patient data will not be publicly available through open science frameworks but may be shared with external researchers on reasonable request and under appropriate data protection agreements.

Trial registration number DRKS00030915.

INTRODUCTION

Internalising disorders—such as anxiety disorders (ADs), unipolar depression, obsessive-compulsive disorder (OCD) and post-traumatic stress disorder (PTSD)—are the most widespread and economically burdensome mental health conditions, often characterising the typical outpatient care recipient.^{1–3} Cognitive-behavioural therapy (CBT) stands as an effective first-line treatment for these disorders,⁴ yet less than 50% of the patients show clinically meaningful improvement.^{5–9} This gap highlights the urgent need for early identification of individuals likely to show treatment non-response (TNR) to CBT and subsequent stratification towards more personalised treatments, in other words: precision psychotherapy.^{10–16}

A thorough understanding of TNR is essential for designing personalised treatments. However, the limited insights into the moderators and mechanisms behind TNR call for an in-depth examination. Traditional group-based research sheds light on potential mechanisms but falls short of offering single patient predictions, which are crucial for an a priori patient stratification.

Predictive analytics, which is based on artificial intelligence (AI) and machine learning (ML) algorithms, offers a promising approach to bridge the translational gap

from group-based evidence ('what is effective?') to individual case predictions ('... for whom?'). Big data and AI are increasingly integrated into mental health research, but initial optimistic results overestimated their predictive accuracy.^{17 18} The reliance on small sample sizes from clinical trials, prioritising internal over external validity, provides a biased foundation for developing algorithms meant for broader healthcare application. In addition, efficacy studies do not necessarily translate into predictive information for treatment outcome. Moreover, the emerging field of predictive analytics is rapidly evolving, and earlier methodologies have been criticised for producing overly optimistic or inconsistent predictions, highlighting the need for caution and rigour in this area.^{19 20} Rather than using randomised controlled trial (RCT) data for secondary predictive analytics, we advocate for study designs that are designed from the outset for predictive purposes, including a representative, ecologically valid sample, an a priori defined predictive use case (eg, outcome prediction) and a hypotheses-based data collection focusing on the predictive question of interest. The research unit (RU) 5187 is a collaborative effort to better understand the signature of TNR based on a multimodal deep phenotyping approach and

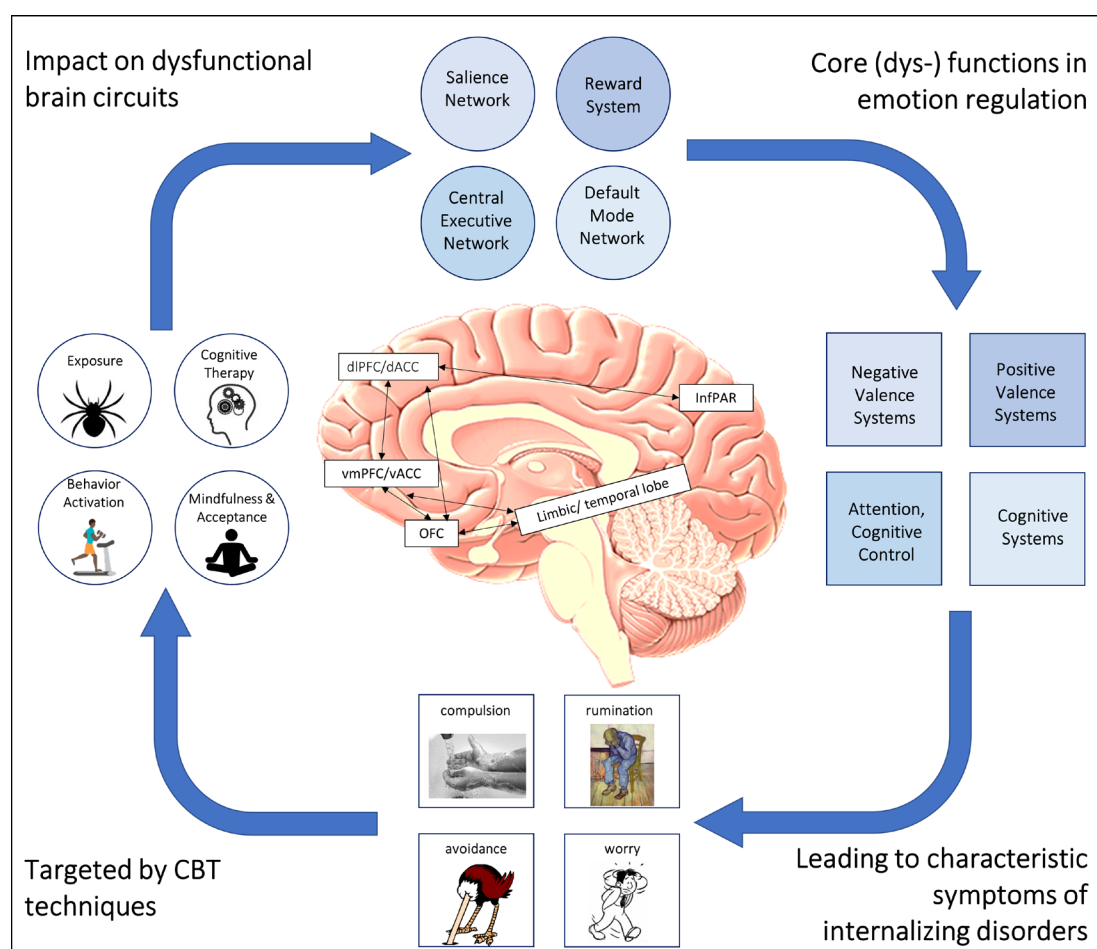


Figure 1 The neurodynamics of emotion regulation in internalising disorders, targeted by CBT techniques. CBT, cognitive-behavioural therapy; dACC, dorsal anterior cingulate cortex; dIPFC, dorsolateral prefrontal cortex; InfPAR, inferior parietal lobe; OFC, orbitofrontal cortex; vACC, ventral ACC; vmPFC, ventro-medial PFC.

Table 1 Scientific subprojects

Projects	Institution	Open science framework preregistration
SP1: Single-case prediction of treatment (non-)response to cognitive-behavioural therapy (CBT) in the outpatient sector: a prospective-longitudinal observational study	HU Berlin FU Berlin PHB	https://osf.io/bcgax
SP2: Methods toolbox and infrastructure for predictive analytics	Charité	https://osf.io/yhvm5
SP3: Neuroimaging backbone and large-scale data harmonisation	Charité FU Berlin	https://osf.io/jv97r
SP4: Brain-electrical and cardiovascular indicators of emotion regulation as predictors of treatment (non-)response to CBT in internalising disorders	HU Berlin	https://osf.io/fbrvd/ https://osf.io/w9anr/
SP5: Transdiagnostic psychological factors as predictors of treatment non-response and cost-effectiveness measures related to predictive analytics in psychotherapy	PHB	https://osf.io/7e4qc (psychological factors), https://osf.io/a7f2b (cost-effectiveness)
SP6: Digital Phenotyping of emotion (dys-)regulation as transdiagnostic process and proxy for clinical and neurobiological markers of treatment (non-)response	FU Berlin	https://osf.io/253nb
SP7: ACC-based biomarkers for predicting treatment (non-)response to cognitive behavioural therapy in internalising disorders	HU Berlin Charité	https://osf.io/cef24 (retrospective), https://osf.io/zehpy (prospective)
SP8: Dynamic causal modelling of emotion regulation as predictors of treatment (non-)response to CBT in internalising disorders	FU Berlin Charité	https://osf.io/w327j (retrospective), https://osf.io/tk38p (prospective)
SP9: Generalising predictive patterns of treatment (non-)response: from specific phobia and obsessive-compulsive disorder to the anxiety spectrum	HU Berlin Charité	https://osf.io/6mxk5
ACC, anterior cingulate cortex; FU, Freie Universität; HU, Humboldt University; PHB, Psychologische Hochschule Berlin.		

develops predictive analytics within a naturalistic cohort of patients receiving outpatient care. In line with contemporary developments towards a neurobiologically based, dimensional classification of mental disorders, such as the Research Domain Criteria by Insel *et al*, and the Hierarchical Taxonomy of Psychopathology by Kotov *et al*, our approach is both transdiagnostic, multimodal and multileveled.^{21 22} We aim to improve current limitations in the field of precision psychotherapy by pursuing the following overarching research questions:

What is the biobehavioural signature of TNR towards CBT?

Our approach integrates a multimodal deep phenotyping method to identify the signature of TNR to CBT across various levels of analysis. We focus on various aspects and components of emotion regulation (ER), a crucial process targeted by CBT techniques, as depicted in figure 1. Given ER's relevance across a wide range of mental disorders within the internalising spectrum, understanding its role in TNR could lead to more customised therapeutic interventions.^{23 24} We hypothesise that initial configurations of ER are linked with TNR, anticipating that these signatures will—in line with our transdiagnostic approach—exhibit considerable overlap in different units of analysis across internalising disorders. To characterise ER, we employ a combination of methods including ecological momentary assessment (EMA), passive sensing, psychological

assessments, psychophysiological measures (EEG and cardiac measures—ECG) and neuroimaging (brain structure, intrinsic connectivity, task-based brain activation and connectivity). We intend to investigate a spectrum of ER strategies, from implicit to explicit and automatic to controlled.²³ These behavioural probes are embedded within tailored experimental paradigms, targeting the putative signature of TNR.^{25 26}

Can we predict TNR on the single patient level?

Building a translational bridge from group-based information to single-case predictions, we hypothesise that we can predict TNR on the single-case level with a prediction accuracy that is above chance so that expert decisions on the treatment of choice can be supported. To achieve this, we design our study specifically for the purpose of predictive analytics in the outpatient sector by recruiting a naturalistic cohort with minimal inclusion and exclusion criteria reflecting typical patient populations. We furthermore expect that prediction accuracies can be significantly improved by combining multiple data levels. State-of-the-art ML models will be employed to learn from both unimodal and multimodal data across various granularities (eg, imaging-derived phenotypes vs raw MRI data). Data will be integrated at different levels using techniques such as multimodal fusion. Employing appropriate cross-validation protocols will prevent data

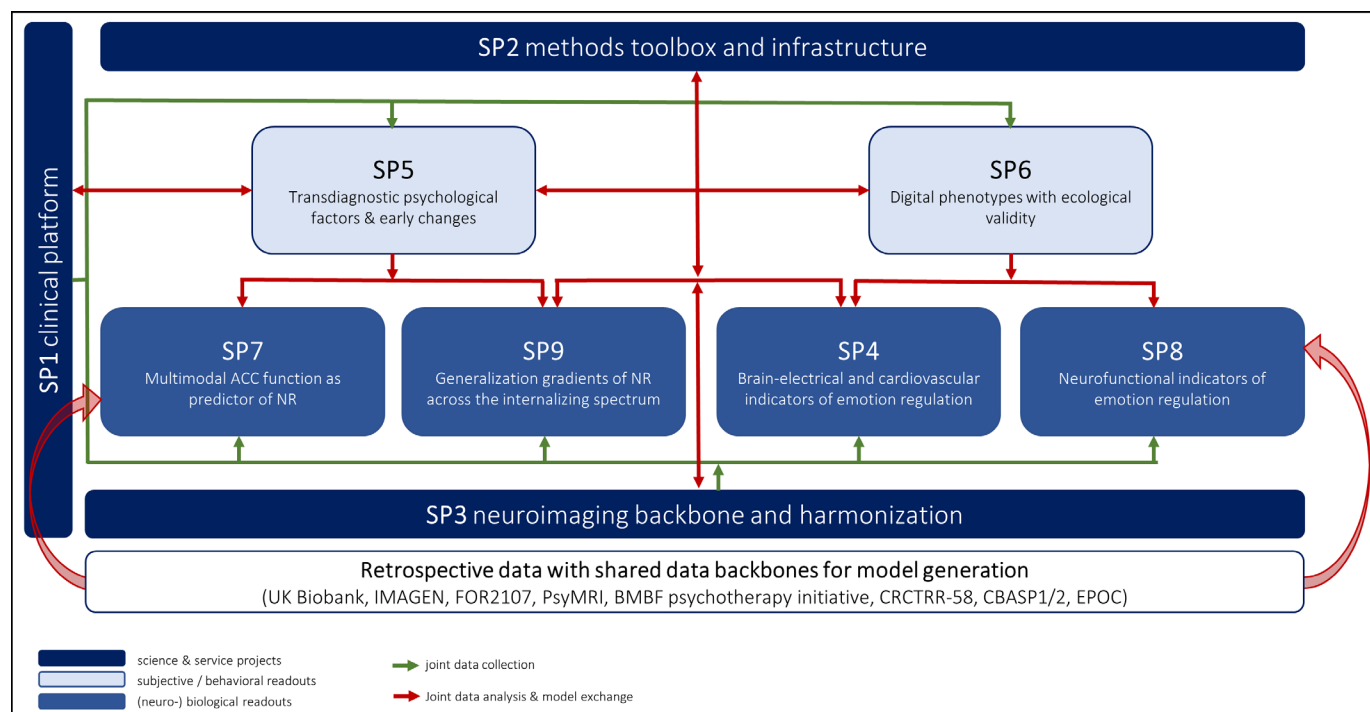


Figure 2 Structure and content of the RU and major collaborations among the subproject (SP). RU, research unit.

leakage and provide reliable estimates of prediction accuracies, thereby addressing the limitations of previous work.^{20 27} We will furthermore search for best combination of predictors. Finally, we hypothesise that we can identify a cost-efficient set of predictors (EMA, EEG, ECG) that may serve as proxy measures for cost-intensive (eg, neuroimaging) assessments. Health-economic considerations will guide our investigation to find cost-efficient combinations of markers or proxies thereof.

METHODS AND ANALYSIS

Composition of the RU

The RU 5187 PREACT ('Towards precision psychotherapy for non-respondent patients: from signatures to predictions to clinical utility') is a collaborative initiative grounded in the rich academic landscape of Berlin (Germany) and is carried out at Humboldt-Universität zu Berlin (HU Berlin), Freie Universität Berlin (FU Berlin), Charité—Universitätsmedizin Berlin (Charité) and Psychologische Hochschule Berlin (PHB). The RU brings together a diverse team of 13 principal investigators (PIs) alongside 36 additional members, including postdoctoral researchers, doctoral candidates and student assistants. Recruitment is based on four academic outpatient clinics (in German: 'Hochschulambulanz'—HSA). Those HSAs are based in three of the participating sites (HU Berlin: HSA-HU and ZPHU—'Zentrum für Psychotherapie der Humboldt-Universität zu Berlin/Center for psychotherapy of the Humboldt-Universität zu Berlin'; FU Berlin: HSA-FU, PHB: HSA-PHB). The RU 5187 is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, project number:

442075332) with a grant spanning 4 years, from 1 July 2022 to 30 June 2026. The RU is funded from 1 July 2022 to 30 June 2026. Recruitment started in May 2023 and is expected to continue until May 2025. The RU's structure is intricate, consisting of nine specialised subprojects (SPs), all orchestrated under a comprehensive coordination project (table 1 and figure 2).

The coordination project's mandate extends beyond administrative oversight, embedding within its core objectives the promotion of young researchers, the advancement of gender equality, the management of network funds and the enhancement of scientific communication. This multidimensional approach not only fosters a nurturing environment for research but also cultivates a framework for inclusive and interdisciplinary collaboration (table 2).

The RU encompasses three core science and service projects (SP1, SP2 and SP3). Additionally, six specific science projects target narrower research questions (as detailed in table 1). These include EEG/ECG assessment (SP4), psychological measures (SP5), ambulatory assessment (AA, SP6) and neuroimaging (SP7, SP8, SP9). These projects are tightly integrated to increase the added value and answer research questions reaching beyond particular SPs. The rationale, hypotheses and analytical plans for all projects are preregistered with the open science framework, ensuring transparency and rigour. For additional information about the projects, refer to the preregistrations listed in table 1.

Summary of SP

SP1 sets up the naturalistic, prospective-longitudinal observational clinical study dedicated to patient

Table 2 Talent promotion and public outreach

Promotion of young scientists	Public outreach
<ul style="list-style-type: none"> ▶ Start-up funding for new projects ▶ Career mentoring for female scientists to foster gender equality ▶ Financial support of conference contributions via travel funding ▶ Financial support of publication costs ▶ Organisation of advanced learning and soft skill workshops 	<ul style="list-style-type: none"> ▶ Scientific symposium with external speakers (2024) ▶ Hackathon ‘Innovations for mental health’ ▶ Citizen-science dialogue with relevant stakeholders (eg, patient representatives) (2025) ▶ Participation in ‘Girls’ Day’ to inspire female students to STEM (Science, Technology, Engineering and Mathematics) subjects ▶ Homepage to inform public about project, publications and news

treatment, recruitment, clinical characterisation and documentation of treatments and investigates the extent to which TNR can be predicted on basic clinical information. SP1 will be supported by our methods (SP2) and neuroimaging platform (SP3). As such, SP1 contains both a research and a service component. Research-wise, we will test the hypothesis if treatment (non-)response can be predicted with sufficient accuracy (75%) based on clinical routine data, when state-of-the-art ML methods are applied. Service-wise, SP1 will coordinate the clinical trial by serving as a recruitment hub for the other SPs, implementing diagnostic assessments, documenting and quality-controlling treatment contents, and calculating primary and secondary outcomes.

SP2 also has a service and a science component. Within the service component, we will (1) securely store and organise retrospective and prospective data in a way that all participating PIs have access and (2) provide hardware for processing complex biomedical data with state-of-the-art ML algorithms including deep learning. Within the science component, we will develop a methods toolbox relying on predictive analytics and ML algorithms for identifying (non-)responders on a single-subject level. This will be done separately for the different disorders as well as across disorders to investigate if treatment response operates on an individual or cross-diagnostic scale. We will systematically compare the predictive value of different data domains, such as clinical data, AA, electrophysiological data and neuroimaging data and will evaluate different methods for data fusion. Specifically, we will experiment with sophisticated ML techniques such as convolutional neural networks for analysing neuroimaging data and will employ transfer learning for dealing with comparatively low sample sizes.

SP3 has three work packages (WP). In the first service WP, we will pilot and optimise our neuroimaging battery. It consists of three ER tasks, spanning the spectrum from automatic to voluntary ER, resting state fMRI as well as structural MRI. The second service WP is to guarantee the proper conduction of the main study by training of data acquisition, to perform continuous quality checks of acquired data, and to deliver preprocessed and basic first-level data for task-based and resting state fMRI. WP1 and WP2 are service components with no hypotheses to be tested. WP3 is the scientific component where we will use resting state fMRI for all and sometimes task-based fMRI.

Within the third WP, we will perform data-driven connectomics analyses using resting state fMRI for prediction of TNR using our newly developed network-based statistics (NBS)-predict tool. NBS-predict combines the data-driven construction of sparse networks that can be biologically interpreted and ML algorithms for prediction. In contrast to SP7, we do not focus on specific regions or hypotheses but use a data-driven method which will identify any network that proves to be useful for prediction. In contrast to SP9, we will not focus on specific disorders or the differences in prediction but on transdiagnostic prediction averaging over 4 and 3 disorders or 8 disorders, respectively.

SP4 aims to investigate whether previously established psychophysiological indicators of ER, such as the late positive potential (LPP), relative left frontal activity or ECG indices like HRV, can predict which patients are at risk of TNR.^{28 29} We hypothesise that low emotional reactivity as indicated by relatively small amplitudes of the LPP to pleasant or threatening pictures, respectively, as well as low capacity of ER as indicated by lower modulation of the LPP under ER instructions (ie, lower reduction of LPP under reappraisal or lower increase of LPP under savouring conditions) is predictive of TNR. Further, low relative left frontal activity under resting conditions and low increase under ER conditions, as well as low HRV generally and low increase under ER conditions can be used as indicators of aberrant emotional reactivity and deficient ER and are hypothesised to be predictive of TNR. Moreover, we hypothesise that compared with predictions based on single psychophysiological predictors, the combination of all available predictors, including their interactions, will yield models with higher overall prediction accuracy in identifying TNR on a single-subject level.

SP5 seeks to examining the relevance of specific psychological mechanisms and processes (so-called transdiagnostic factors) for TNR in internalising disorders. We will focus on ER strategies as a putative key mechanism of change underlying CBT and assess psychological ER processes that match and enrich—in terms of the targeted constructs—the prediction tools of the other SPs. Moreover, we assess further transdiagnostic skills and factors such as illness perceptions and treatment expectancies as well as therapeutic alliance. The relevance of these transdiagnostic factors will be examined above and beyond routinely available clinical diagnostic

and sociodemographic data. Further, an initial mediation model accounting for the interdependence of transdiagnostic factors will be examined. Moreover, this SP will focus on early changes in transdiagnostic factors in CBT, thereby acknowledging that early changes can predict treatment outcomes in internalising disorders. Lastly, we adopt a cost-effectiveness perspective and deliver estimates on the relation of costs of the RU assessments (ie, resources spent on the input for the prediction) to their contribution to increase prediction accuracy. Further, we model implications of a better prediction of TNR in ‘standard CBT’ in terms of cost-effectiveness when patients with a poor prognosis would be assigned to alternative or modified treatment options (which might address the psychological processes and mechanisms that have been identified to be associated with risk of TNR).

SP6 uses digital phenotyping to derive proxies that transfer the predictive value of neural TNR-signatures and related ER-(dys)functions of internalising disorders into clinical settings and can be assessed repeatedly with minimal effort for therapists and patients during different treatment periods (eg, before, during and after therapy). We use data actively and passively collected by personal electronic devices such as smartphones and wearables, to derive phenotypes including ER markers of individuals at risk of TNR based on indicators of ER as the overarching core construct. Research questions are: Can data actively and passively generated by personal electronic devices such as smartphones and wearables predict TNR for the individual patient at T20 (ie, after 20 sessions) and post/12 months? Which smartphone-collected data-modality predicts TNR for the individual patients at T20 and post/12 months best, and how does predictive power depend on the time lag? Do EMA and sensor-based ER markers link to ER’s neural and biobehavioural markers assessed in the other SPs?

SP7 focuses on the anterior cingulate cortex (ACC) a crucial hub region for ER given its integration in several core brain networks, which has been demonstrated to be central to both psychopathology and CBT. We propose that a dedicated systems-driven perspective will enable us to develop neuroimaging-derived biomarkers capable of predicting TNR to CBT for internalising disorders. In particular, SP6 will optimise ACC predictor information by employing three sophisticated methods: (1) an efficient gradient of brain connectivity encoding alterations in the neural basis of explicit-controlled and implicit-automatic ER, (2) a graph-theoretical characterisation of ACC networks and (3) a Bayesian hierarchical modelling approach to networks focusing on the ACC hub. Additionally, like in all other SP results from this project will be returned into the RU, by providing an optimal composition of ACC-centred system characteristics to SP2, by comparing our predictive performance with other neuroimaging-based predictors from SP8 and SP9 and by investigating whether our predictive model is associated with markers based on ER strategies and psychophysiological markers of ER from SP4 and SP5.

SP8 aims to develop and validate neurofunctional markers of ER as predictors of TNR. We will develop effective connectivity models of fronto-limbic modulation during three different ER strategies in order to establish a set of models with the best predictive values for TNR in retrospective data sets of patients with depressive, obsessive-compulsive and social AD (SAD). We will then validate our models in the transdiagnostic prospective sample of the proposed RU cohort, also testing for generalisation and specificity across tasks and diagnoses. We will refine our models by including behavioural markers of ER from clinical measures (SP5) and digital phenotyping (SP6), exchange results with (SP4) to develop psychometric measurement models of ER using BOLD, EEG and ECG data, and compare task-based effective connectivity measures with resting state connectivity (SP7). Finally, we will share our data with SP2 for single case prediction.

SP9 aims to investigate the hypothesis to which extent neural structure and function related to treatment response can be generalised from specific phobia (SPH) as a ‘model’ disorder for pathological fear to other, more complex ADs and related conditions (SAD, panic disorder (PD), agoraphobia (AG), generalised AD (GAD) and PTSD). Second, empirical evidence for the separate classification of OCD will be tested by comparing signatures and predictions between OCD and the anxiety spectrum. SP9 will use two retrospective datasets in order to train predictive models which in turn will be replicated/cross-validated in the prospective cohort from the RU. All datasets share a common neuroimaging backbone comprising neuroanatomic data, resting-state networks (EEG, fMRI) and task-based activity and connectivity which will be used to investigate generalisation gradients.

Naturalistic prospective-longitudinal observational clinical trial: study design, participants, eligibility criteria and current status

We include patients with a primary diagnosis of an internalising disorder receiving CBT at one of the four participating academic outpatient clinics. These university-based clinics are integral to Germany’s psychotherapy landscape, serving dual roles in research and education within the German healthcare system, hence ensuring the high ecological validity and representativeness of our sample. The inclusion criteria of the study are as follows: (1) Age of 18 years or older, (2) a primary diagnosis based on the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 criteria for either SPH, SAD, PD, AG, GAD, OCD, PTSD or unipolar depressive disorder (major depression or dysthymia, DD), (3) an indication for outpatient CBT treatment, (4) a minimum treatment plan of 12 sessions (excluding mandatory probatory sessions required in Germany before treatment), (5) a symptom severity indicated by a General Severity Index of the Brief Symptom Inventory (BSI-GSI³⁰) greater than 0.56 (except for SPH³¹), (6) a Clinical Global Impression-Severity Scale (CGI-S³²) score of 3 or higher (indicating at least mild illness) and (7) a willingness to

participate in the study. We abandoned the BSI-GSI inclusion criteria for SPH because too many patients had to be excluded at the beginning of recruitment due to this criterion, making it impossible to reach our recruitment goal. Exclusion criteria were held to a minimum in order to foster ecological validity of the sample. They mainly encompass contraindications for outpatient treatment, including (1) a current secondary diagnosis of moderate to severe substance use disorder (including regular use of benzodiazepines), (2) current psychosis, (3) current bipolar disorder, (4) more than moderate suicidality and (5) medical conditions that contraindicate CBT. Patients who require inpatient treatment during the course of the outpatient treatment will be excluded. However, the presence of other comorbid disorders will not lead to exclusion. Concurrent pharmacological treatment is permitted and will be thoroughly documented.

Patients opting for the experimental assessments (EMA, fMRI and EEG/ECG) will receive detailed information about possible risks and additional eligibility criteria including specific safety-related exclusion criteria will be applied.

In our sample (28 November 2024), the distribution of diagnoses was as follows: ADs 35.5%, OCD 18.6%, DD 41.0% and PTSD 4.9%. Compared with a nationwide sample from the KODAP consortium with 6098 patients (ADs 29.29%, OCD 5.76%, DD 58.53% and PTSD 6.42%), the distributions are largely similar.^{33 34} However, the proportion of patients with OCD is notably higher in our sample. This can be attributed to the specialisation of one of our outpatient clinics in OCD treatment, which led to an intentional enrichment of the sample.

For current study progress, see [figure 3](#).

Sample size calculation

We estimate effect size from two perspectives. First, we aim to achieve or surpass the best ML model prediction accuracy reported in existing literature, which is around 75%. We expect that improving prediction accuracy by 25% from the chance level (50%) will have clinical and health-economic significance.

To accurately estimate the necessary sample size, two key factors are considered: (1) prediction accuracy is expected to increase with a larger training dataset, up to the maximal achievable accuracy based on the underlying effect and (2) fluctuations in detected prediction accuracy around the true value will decrease as the test set size increases. The exact training set size needed to achieve a 75% prediction accuracy is uncertain and depends on both predictor characteristics and the chosen algorithm. Our simulations, consistent with previous research,³⁵ suggest that prediction accuracies plateau at around 300 samples in the training set across various models ([figure 4](#)). Further simulations indicate that a test set of about 100 patients will likely yield a measured prediction accuracy of 75%–85%.

Based on these insights, a fivefold cross-validation method appears suitable. This involves assigning 400

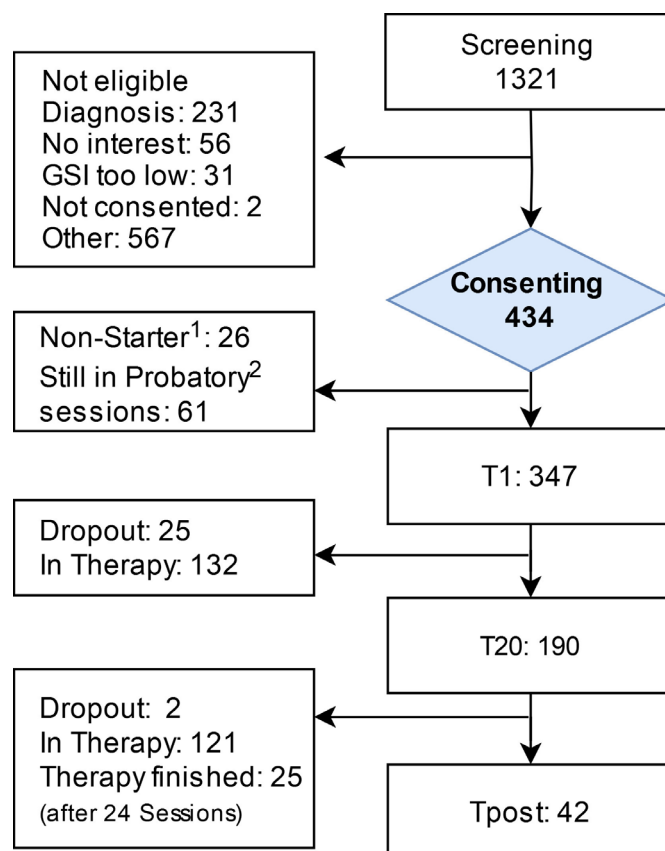


Figure 3 Study CONSORT flow chart, 28 November 2024. ¹Termination during probatory sessions. ²Mandatory preliminary sessions (up to four) required in the German healthcare system before initiating treatment. CONSORT, Consolidated Standards of Reporting Trials.

patients to the training set and 100 to the test set, with the analysis repeated five times to estimate variability.

Additionally, for comparative purposes with traditional studies, we also present an effect size calculation for group-based linear multiple regression analyses, despite not using this method in our study. Considering an alpha level of 5%, a power (β) of 80% and 19 predictors, small effects ($R^2=0.07$) can be detected with a sample size of 500 patients. This calculation is based on the power analysis conducted using G*Power V.3.1.9.2 for linear multiple regression. The recruitment plan for the study is designed to ensure a steady and sufficient number of participants throughout the trial. Initially, at baseline, 585 patients will be recruited. By the fifth session (S5), it is anticipated that the patient number will decrease to 555 due to an expected attrition rate of 5%. As the study progresses to the 20th session (S20), the number of participants is projected to be 527, accounting for a cumulative attrition rate of 10%. Finally, at the postassessment, the study expects to have 500 patients, reflecting a cumulative attrition rate of 15%. This recruitment strategy considers potential dropouts and aims to maintain a robust sample size for the study's duration.

Treatment setting, contents and procedure

The treatment follows the same basic CBT approach at all sites and is based on established manuals,^{36–39} but not

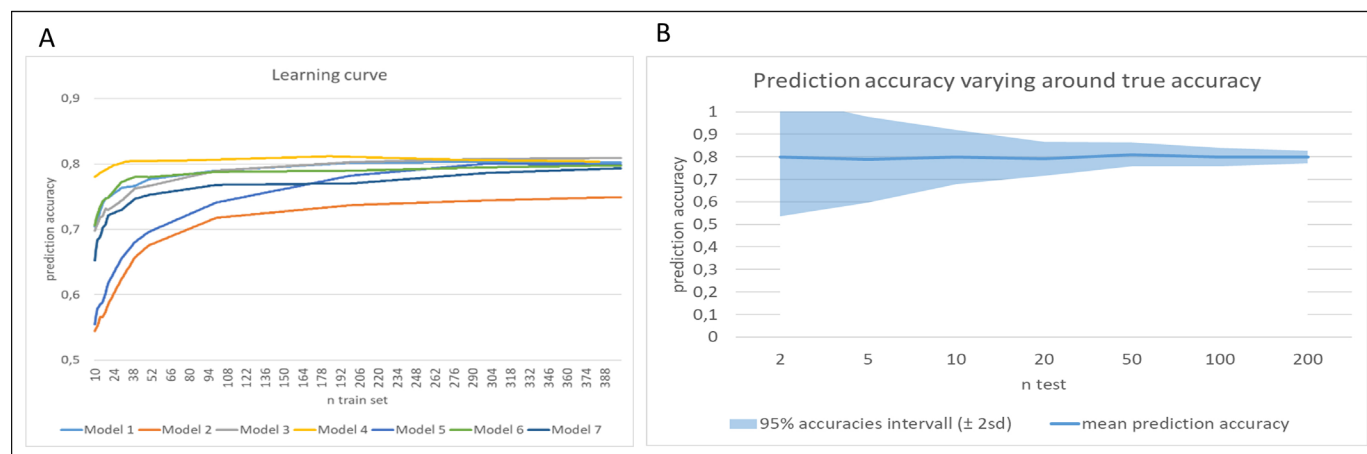


Figure 4 Simulation of association between sample size and prediction accuracy. Association of train and test set sizes with prediction accuracy across simulations. (A) Across increasing size of n train, prediction accuracy is increasing until reaching a plateau (ie, the maximum prediction accuracy possible based on the underlying effect). Starting point and gain in prediction accuracy depend on data and model parameters. (B) Across increasing size of n test, the measured prediction accuracy varies less and less around the true prediction accuracy for the underlying effect.

strictly manualised to mirror the ecological approach. We aim to predict TNR in a naturalistic setting, where CBT is usually tailored (based on clinician expertise, ie, ‘eminence-based personalisation’) to the specific requirements of each patient. However, despite this form of personalisation, TNR remains a frequent outcome. Therefore, this naturalistic CBT-based treatment addresses individual variation in clinical phenomenology, including comorbid disorders. These elements are not only addressed through individualised therapy plans but are also analysed as potential predictors of TNR.

Standard manuals^{36–39} typically encompass psychoeducational components, various intervention forms like exposure techniques, behavioural activation or cognitive restructuring and relapse prevention strategies. Research^{40–41} indicates that despite their differences, these techniques collectively enhance basic ER skills. The regimen and duration of treatment adhere to the specific regulations of the German public healthcare system, markedly distinct from those in other countries such as the UK or USA. The legal framework permits up to 80 sessions of 50 min each. Insurance companies approve varying numbers of sessions, categorised into ‘short-term therapy part 1’ for sessions S1–12, short-term therapy part 2 for sessions S13–24, and extensions for long-term treatment usually up to 60 or in some cases up to 80 sessions. Typically, treatments involve one session per week, with longer sessions for exposure therapy.

The number and length of sessions at our outpatient clinics vary between the type and severity of the disorder, averaging 40–50 sessions over 1–1.5 years. Nonetheless, treatment can deviate to as few as 10–20 sessions (short-term treatment) over 4 months or extend to 80 sessions (long-term treatment) over 2 years. Considering these variations and trial time constraints, we decided to evaluate treatment outcomes after a predetermined number of sessions (‘treatment dose’ at S20) and at the end of

treatment with a maximum duration of 12 months (‘treatment time’; although the naturalistic treatment may go beyond, but not within the study regimen). Additionally, we will monitor early psychological changes after the fifth session (S5).

Patients seeking treatment will first undergo routine screening, receive a detailed explanation of the study protocol and provide informed consent before participating. The initial five preliminary/probatory sessions, designated as the diagnostic interval, serve as the baseline period for eligibility assessment based on inclusion and exclusion criteria, and for collecting multimodal predictors such as clinical data, AA including EMA and passive sensing, fMRI and EEG/ECG data. All clinical data are electronically compiled in a central database via ‘Research Electronic Data Capture’ (REDCap), which will be integrated into a shared and secured server alongside our multimodal datasets.^{42–43}

For an in-depth overview of the study flow, see figure 5.

The study was approved by the Institutional Ethics Committee of the Department of Psychology at Humboldt Universität zu Berlin (approval no. 2021-01) and the Ethics Committee of Charité-Universitätsmedizin Berlin (approval no. EA1/186/22). Participants in the experimental sessions, including AA, EEG/ECG and neuroimaging studies, will receive financial compensation. The trial has been registered with the German Clinical Trials Register (DRKS00030915) and on the open science framework platform of the Center for Open Science, available at <https://osf.io/bcgax>.

Measures and outcomes

Clinical assessments

The clinical interview, demographics and the questionnaires outlined in table 3 will be conducted by trained clinical psychologists who hold at least a bachelor’s

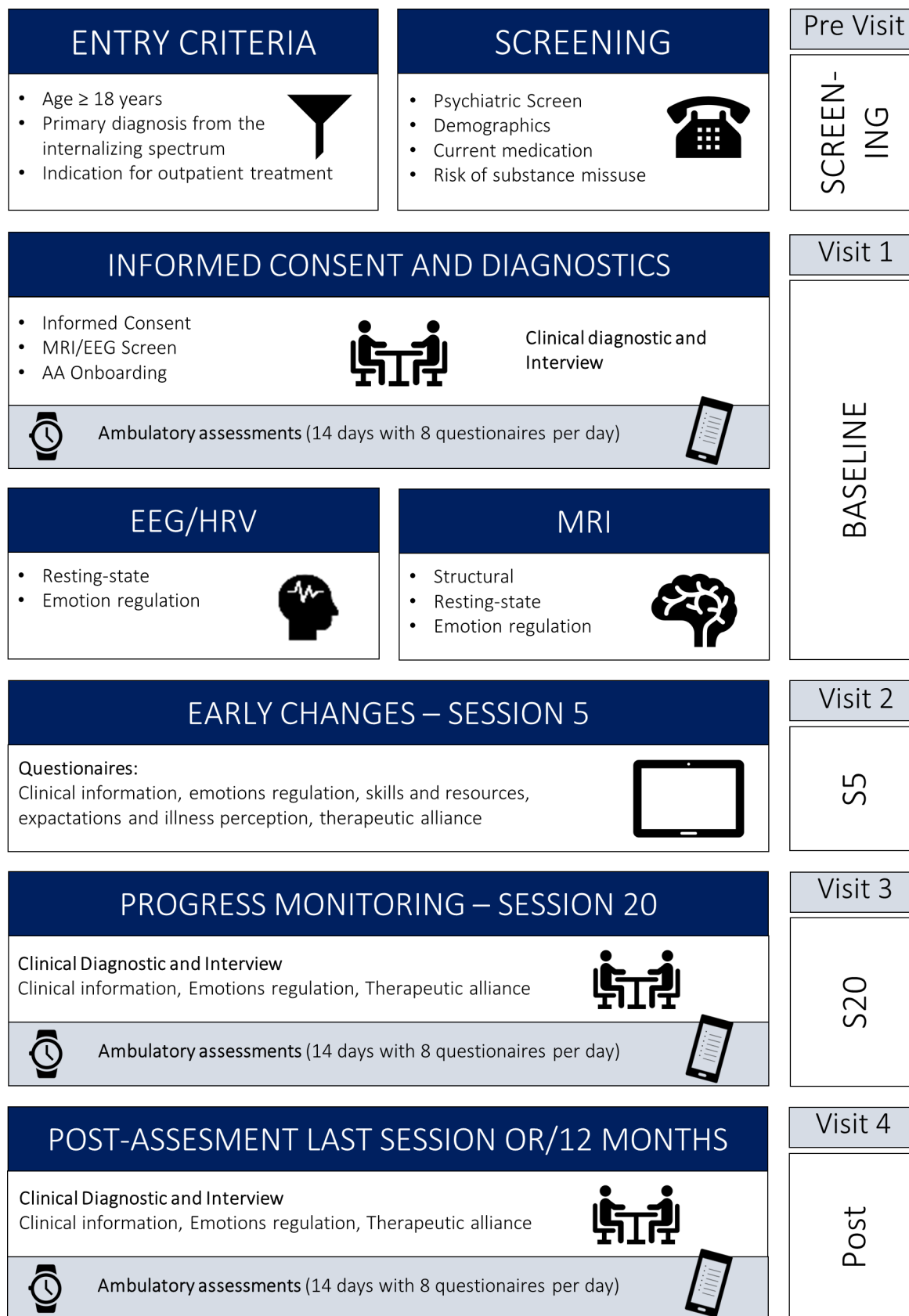


Figure 5 Study workflow. AA, ambulatory assessments; EEG, electroencephalogram; HRV, heart rate variability; S, treatment session.

Table 3 Study measures

Domain	Measure	Rater	Function	Baseline	S5	S20	Post
Demographics	Self-constructed	Patient	Predictor	x			
Clinical information	CTQ	Patient	Predictor	x			
	BDI-II	Patient	Predictor	x	x	x	x
	ASI	Patient	Predictor	x		x	x
	BSI	Patient	Outcome	x		x	x
	HAM-A	Clinician/therapist	Outcome	x		x	x
	MADRS	Clinician/therapist	Outcome	x		x	x
	Y-BOCS	Clinician/therapist	Outcome	x		x	x
	CAPS-5	Clinician/therapist	Outcome	x		x	x
	CGI-S	Clinician/therapist	Outcome	x		x	x
	CGI-S prospective*	Clinician/therapist	Outcome	x			
	SCID-5-CV	Clinician/therapist	Diagnostics	x			
	SCID-5-SPQ	Patient	Diagnostics	x			
Emotion regulation	HFERST	Patient	Predictor	x	x	x	x
	PASS	Patient	Predictor	x	x	x	x
	CERQ	Patient	Predictor	x	x	x	x
	BERQ	Patient	Predictor	x	x	x	x
Skills and Resources	ICQ	Patient	Predictor	x	x		
	MOS social support	Patient	Predictor	x	x		
Expectations and illness perception	IPQ-MH	Patient	Predictor	x	x		
	TEX-Q	Patient	Predictor	x	x		
Therapeutic alliance	WAI	Patient	Predictor		x		
	WAI	Clinician/therapist	Predictor		x		
	TAI	Patient	Predictor		x	x	x
Therapy content	Study derived		Adherence			x	x

All other-rating instruments are administered by certified assessors.

*Clinicians are requested to provide their anticipated ratings for the CGI-S following therapy.

ASI, Anxiety Sensitivity Inventory; BDI-II, Beck Depression Inventory II; BERQ, Behavioural Emotion Regulation Questionnaire; BSI, Brief symptom inventory; CAPS-5, Clinician-Administered PTSD Scale DSM-5; CERQ, Cognitive Emotion Regulation Questionnaire; CGI-S, Clinical Global Impression–Severity of illness; CTQ, Childhood Trauma Questionnaire; HAM-A, Hamilton Anxiety Rating Scale; HFERST, Heidelberg Form for Emotion Regulation Strategies; ICQ, Interpersonal Competence Questionnaire; IPQ-MH, Illness Perception Questionnaire Mental Health; MADRS, Montgomery–Åsberg Depression Rating Scale; MOS, Social Support Survey; PASS, Positive Appraisal Style Scale; S5, session 5; S20, session 20; SCID-5-CV, Structured Interview for DSM-5 Disorders, Clinician Version; SCID-5-SPQ, SCID Screening Personality Questionnaire; TAI, Therapeutic Agency Inventory; TEX-Q, Treatment Expectation Questionnaire; WAI, Working Alliance Inventory; Y-BOCS, Yale-Brown Obsessive Compulsive Scale.

degree. These professionals will operate under the supervision of board-certified psychotherapists.

Primary outcome

In defining treatment response, we adhere to the concept of clinically significant change as introduced by Jacobson & Truax (1991).⁴⁴ In line with the guidelines for the German academic outpatient sector,³¹ the shift from a dysfunctional to a functional mental condition is operationalised by achieving a BSI-GSI score of less than 0.56 at T20/post-treatment assessment. Additionally, the change observed from baseline to T20/post-treatment must be reliable, as defined by the Reliable Change Index. Considering the reasons mentioned above, the primary outcome does not apply to SPH. Here, we will apply the secondary

outcome as primary (50% reduction of symptom load in the Hamilton Anxiety Rating Scale (HAM-A)).

Secondary outcomes

Secondary outcomes will be assessed using disorder-specific severity measures derived from clinical interviews. In line with the approach of other multicentre clinical studies,⁴⁵ a 50% reduction in symptoms on the HAM-A⁴⁶ will signify treatment response for PD, AG, SAD, SPH and GAD. For DD, a 50% symptom reduction on the Montgomery–Åsberg Depression Rating Scale⁴⁷ is required. For OCD, a 35% symptom reduction is needed, as measured by the Yale-Brown Obsessive Compulsive Scale.^{48 49} For PTSD, a 30% symptom reduction is needed, as measured by the Clinician-Administered PTSD Scale for DSM-5.^{50 51}

Box 1 Ambulatory assessment

Ecological momentary assessment (EMA)

- ⇒ Affect (8×per day; positive and negative affect scale).
- ⇒ Emotion regulation (8×per day; adapted from RESS-EMA).
- ⇒ Situation/context (8×per day; self-constructed).
- ⇒ Social contact (8×per day; self-constructed).
- ⇒ Significant events (8×per day; self-constructed).
- ⇒ Physical fitness (1×per day; self-constructed).
- ⇒ Therapeutic agency (1×per day; adapted from TAI).
- ⇒ Reminder manual ECG (2×per day; self-constructed).

Passive sensing

- ⇒ Physical activity.
- ⇒ Sleep.
- ⇒ Heart rate.
- ⇒ Location (pseudonymised; event-based trigger).

RESS, Regulation of Emotion Systems Survey; TAI, Therapeutic Agency Inventory.

Additionally, baseline ER capacities and their change processes, which are key mechanisms of interest, will be thoroughly examined pretreatment and after session 5. In the event of a patient withdrawing from the treatment, the reasons for dropout will be documented, and reasons for dropout as well as outcome measure proxies will be gathered post-dropout where feasible.

For a detailed overview of further clinical characteristics and questionnaires concerning ER, skills and resources, expectations/illness perception, and therapeutic alliance, refer to [table 3](#).

Ambulatory assessment

Participants registering for this SP are provided with a wearable (Withings France SA Scanwatch; <https://www.withings.com/de/de/scanwatch>) and the respective app for passive data collection, as well as a study app for EMA and location tracking. Participants can decide if they want to take part in (1) three active EMA phases at baseline, S20 and post and continuous passive data sharing for up to 365 days or (2) one active EMA phase at baseline and passive data sharing for 15 days. Each EMA phase lasts 14 days with 8 questionnaires per day. The full set of assessed features is provided in [box 1](#).

Biobehavioural assessments

Psychophysiology

The EEG/ECG measurement employs three consecutive experimental paradigms: A resting state paradigm, a picture viewing task (including passive viewing and ER subtasks), and a feedback processing paradigm (including a guessing ‘doors task’ phase, followed by a probabilistic reversal learning phase). Detailed specifications of the measurement protocol are provided in [table 4](#). More information on the experimental tasks is given in the preregistrations (see [table 1](#)).

Table 4 EEG/ECG session

Task	Domain/measured variables	Duration (min)
Resting state	Alpha asymmetry, HRV	4:00
Picture viewing task	Alpha asymmetry, LPP, HRV	25:00
Feedback processing	RewP, P300, HRV	22:00

EEG, electroencephalogram; HRV, heart rate variability; LPP, late positive potential; RewP, reward positivity.

Neuroimaging

MRI assessments in this study are conducted using two 3-Tesla Siemens Prisma MRI scanners, located at two sites (Center for Cognitive Neuroimaging Berlin, at FU Berlin and Berlin Center for Advanced Neuroimaging, BCAN, at Charité–Universitätsmedizin Berlin). The MRI protocol comprises an anatomical scan, a resting-state functional MRI, and three distinct ER tasks span a spectrum from automatic to voluntary ER mechanisms. Owing to the bicentric data collection and associated requirements for harmonising quality control measures, we strictly adhere to standard operating procedures that are clearly defined. We continuously cooperate and provide feedback between sites, conduct regular site visits and perform quality checks on the acquired data as soon as they are available. Standardised preprocessing protocols from the ENIGMA fMRI group will be applied.⁵² Detailed specifications of the scanning session are provided in [table 5](#). Details on the experimental tasks and scanning protocols can be found in the preregistration (see [table 1](#)).

Data analysis: ML preprocessing and validation pathway

Here, we focus on ML models for the predictive analytics. For detailed analyses of single SP, see preregistrations in [table 1](#).

Table 5 Neuroimaging session

32 Channel head coil: structural and fMRI	Duration (min)
Head scout	0:14
MRI anatomical scan (T1-W)	5:39
fMRI Field Map	0:16
fMRI Resting State with eyes open	8:00
fMRI Field Map	0:54
fMRI Emotional Face Matching Task (EMO-FM) ⁸⁷	4:20
fMRI Voluntary Emotion Regulation Task (EMO-REG) ^{88 89}	4:48×4 runs
fMRI Field Map	0:54
fMRI Emotional Delayed-Match-to-Sample Working Memory Task ⁹⁰	12:05
MRI T2-W Clinical Flair	5:44
MRI anatomical scan (T2-W)	3:23

For predicting TNR, we will initially develop separate models for each data modality—including clinical, behavioural, psychophysiological and neuroimaging data. Subsequently, we will integrate these modalities to enhance the predictive accuracy.

For tabular data, which includes demographic, clinical and psychological data as well as features derived from EMA data, we will primarily use classical ML algorithms like gradient boosting and random forests, known for their efficacy with structured data.⁵³ We will also explore the use of elastic nets, support vector machines and experimental deep learning techniques,⁵⁴ although cautiously due to their unproven benefits for non-hierarchical data and small sample sizes. For the interpretation and assessing the importance of the individual features, we will use model-agnostic shapley values (SHAP;⁵⁵). This will also allow for comparing the relative importance of the different data sources (eg, demographic vs EMA data).

For neuroimaging data, we will systematically compare classical ML methods relying on extracted features (ie, image-derived phenotypes) with deep learning methods directly applied to minimally processed (ie, harmonised and linearly registered to MNI space) MRI data. Starting point will be a comparably simple CNN model that we originally trained for differentiating between patients with Alzheimer's disease and healthy controls based on T1-weighted (MPRAGE) data,⁵⁶ but used as a baseline model for other use cases.⁵⁷ Based on this baseline model, we will experiment with different settings of the network based on the training data to optimally adapt to the acquired data while still allowing to generalise (eg, numbers of layers or filters, different loss functions, different regularisation schemes, including short-cuts or dense connections, etc).

To manage small sample sizes in deep learning models based on MRI data, we will implement two key strategies. First, data augmentation will be employed, where modified versions of the input data are incorporated into the dataset. This approach not only serves as regularisation but has also proven to enhance performance in MRI studies.⁵⁸ Second, we will use transfer learning.⁵⁹ Here, models are initially trained on a source dataset (eg, UK Biobank or other retrospective datasets) for a specific task (like disease classification) and subsequently fine-tuned on a target dataset for a different task (such as outcome prediction). This method leverages pretrained weights to refine the model for the target task, a technique demonstrated to significantly boost performance in our previous studies, including across different tasks and MRI sequences.⁵⁷

Additionally, we will explore various scenarios for effective transfer learning, such as the adaptation of models from general disease classification to specific outcome prediction within patient groups. This process not only helps in identifying key areas for differentiation between patient groups but also in assessing the transferability and efficacy of models across different tasks. Our approach will include a comparative analysis of models with and

without pretraining to determine the most effective settings for transfer learning. Successful model transfer to clinical practice will be a key consideration. To elucidate the decision-making process of these neural networks, we will use advanced visualisation techniques such as SHAP and layer-wise relevance propagation.⁶⁰ These post hoc methods will help in understanding the neurobiological correlates of TNR as identified by neuroimaging techniques, and in assessing the impact of different transfer learning models.

After analysing each data modality separately to evaluate its stand-alone predictive performance, we will focus on integrating diverse data types, including neuroimaging data, psychophysiological, clinical data, psychological test scores and EMA data. Combining these varied data sources presents unique challenges, especially given the unknown interactions between them and the uncertainty regarding the informativeness of specific features.⁶¹ We will explore both intermediate and late data fusion techniques. In intermediate fusion, features from different modalities are combined into a single model after separate processing and extraction. In contrast, late fusion uses these features as inputs for different classifiers in an ensemble approach. For MRI data, we will use either image-derived phenotypes or CNN-based representations. Our aim is to determine the relevance of different data domains, including emerging putative biomarker candidates from various sources. We will also pursue a cost-sensitive strategy, focusing on minimal yet effective feature sets, applying dimensionality reduction and feature selection techniques like recursive feature elimination. Data fusion methods will range from simple concatenation to more complex techniques like ensemble learning, boosting, attention-based methods and multiple kernel learning, the latter of which allows for consideration of interdependencies between data sources.^{61 62} For deep learning-based multimodal data integration, we will rely on the recently published toolbox fusilli (<https://iopscience.iop.org/article/10.1088/2516-1091/acc2fe>),

To maintain comparability between methods and data types (tabular or neuroimaging, unimodal vs multimodal analysis), we will apply a consistent validation strategy. Data will be randomly divided into training, validation and test sets. To account for the relatively small sample size, we will perform these splits multiple times (based on computational feasibility and prediction variance) and calculate average performance metrics as well as variability indices to estimate the reliability of prediction. In cases of data imbalance to the respective label (age, sex or TNR), stratified splitting will be used. The effectiveness of the models will be assessed based on their ability to predict treatment outcomes for individual patients and across various disorders. This evaluation will also consider the interpretability of the models by clinical or domain experts. Further, we will examine models for age-specific or sex-specific effects by separately analysing performance across different subgroups, such as women vs men or younger versus older individuals, and our developed

deep representation visualiser allowing to estimate the effects of different confounding variables in a trained ML model.⁶³ The significance of each model's performance will be verified using permutation tests, where labels are randomly shuffled. This comprehensive approach ensures rigorous evaluation and comparability of the models across different data types and research scenarios.

We are committed to using Python for model development, leveraging libraries like sklearn⁶⁴ and PyTorch, and making our code publicly available for broader use and collaboration. ML experiments will be conducted at the computing facility of Charité–Universitätsmedizin Berlin.

Data handling and dissemination

In compliance with the stringent requirements of the General Data Protection Regulation in Germany ('Datenschutz-Grundverordnung'—DSGVO), we have set up a data protection concept for prospective clinical studies and signed a Joint Controller Agreement by all participating institutions for data handling. These documents have been approved by all respective data protection officers. Given the sensitivity of these health data, they are subject to strict protection protocols and are not shared with third parties as long as they remain identifiable. To ensure the privacy of the patients is preserved, all data are safeguarded by appropriate technical (audit-trails and user rights management in access group) and organisational measures (standard operating procedures and standardised personal training) throughout the study phase. One year after the completion of the trial, all data will be irreversibly anonymised. MRI will be defaced to remove identifiable features. Only after this anonymisation process will the data be shared with other consortia such as ENIGMA (Enhancing neuro imaging genetics through meta-analysis; <https://enigma.ini.usc.edu/>).

Patient and public involvement

Implementing a participatory research approach right from the outset together with the DZPG (Deutsches Zentrum für psychische Gesundheit—German Research Center for Mental Health) PPI (patient participation involvement), we developed the research objectives and study design of the renewal proposal conjointly with a patient advisory board encompassing five experts with lived experience. Those were invited to join our consortium and consulted us in several meetings. Importantly, principal investigators wrote layman's sketches of each study plan which were uploaded in a joint document that was reviewed and commented by those five experts.

DISCUSSION AND OUTLOOK

TNR to CBT poses a substantial challenge for affected individuals, mental healthcare resources and societies. The research consortium RU 5187 aims to enhance the understanding of this phenomenon by revealing its biobehavioural signature, ranging from deep to digital phenotypes. Transcending the mechanistic understanding, we

aim to provide predictive models of TNR that can be applied to single patients. Both goals are pursued within a clinical sample that exerts a high ecological validity for the outpatient sector, thus offering potential to close the translational gap to clinical utility. As such, we primarily focus on internalising disorders, which represent the majority of patients in this treatment setting. By employing innovative ML techniques, we intend to develop theranostic markers tailored to the individual needs of patients. If successful, these predictors will provide scientifically informed markers of TNR, potentially aiding future decision-making on personalised treatments in clinical practice.

Methods-wise, our approach also emphasises improving predictive analytics by creating a comprehensive dataset that was a priori set up for this purpose and which has a high degree of ecological validity, hence overcoming current shortcomings in the field.¹⁷ This dataset will enable testing of various models and algorithms to determine which are most suitable for our data. Health-economic considerations are integral to our study, guiding us towards identifying cost-efficient combinations of markers or their proxies. The deeply phenotyped and ecologically valid sample will complement large-scale initiatives for external validation, contribute to the development of a data model warehouse for public exchange and align with national and international consortia such as ENIGMA and others.

However, in order to alleviate high rates of TNR, early identification of vulnerable patients by the use of predictive analytics is not enough. Rather, we need to know what works for whom and how we can augment current (psychotherapeutic) first-line treatments. Let us imagine that in the future we could implement an algorithm that is able to predict the risk of TNR to a first-line treatment with high precision. How can we optimise existing treatments for the respective patient in a next step? While clinicians often personalise psychotherapy based on their clinical judgement and experience, there is a growing need to enhance this process through data-driven insights to optimise treatment outcomes for patients at risk of TNR. Clinicians' predictions are poor and outperformed by ML algorithms.^{65–67} Our approach aims to move beyond traditional clinician-based personalisation by incorporating predictive analytics to identify biobehavioural markers associated with TNR. This allows for the tailored adaptation of treatment strategies based on individual patient profiles.

While pharmacological treatment augmentation (eg, adding medication from a different class) has already made its way into current treatment guidelines, non-pharmacological augmentation strategies are an emerging field of research, presenting a rich body of strategies—from neuroscientifically informed approaches such as neurofeedback or non-invasive brain stimulation^{68–74} to behavioural approaches such as exercise, mind-body practices and high-ventilation breathwork^{75–76} to more psychologically based add-on interventions.^{77–80}

The long-term aim of RU 5187 for the next funding phase is to implement non-pharmacological augmentation of psychotherapy and to enrich clinical study designs with predictive modelling as laid out by individualised treatment rule models such as the Personalised Advantage Index (PAI^{78 79}).

For example, neurofeedback interventions can modulate neural circuits involved in ER, thereby enhancing the effectiveness of CBT for patients with deficits in this area.⁷¹ Similarly, non-invasive brain stimulation techniques, such as transcranial magnetic stimulation, transcranial direct current stimulation and transcutaneous auricular vagus nerve stimulation, have been shown to facilitate neuroplasticity and improve cognitive functions essential for therapeutic progress.⁷² Behavioural interventions like aerobic exercise can increase neurogenesis and release neurotrophic factors, which may augment the therapeutic effects of psychotherapy.⁷⁶

By identifying specific deficits through predictive markers—such as impaired ER or cognitive control—we can select appropriate augmentation strategies to target these processes. This precision augmentation aims to enhance the patient's responsiveness to psychotherapy, potentially leading to better outcomes.

Future research, integrating clinical augmentation studies with predictive analytics, should investigate the optimal timing, type and combination of augmentation strategies. Understanding when to introduce augmentation (eg, before, during or after psychotherapy sessions) and which interventions are most effective for specific patient profiles will be crucial for integrating these approaches into clinical practice. Moving towards data-driven personalisation and incorporating enriched treatment protocols holds promise for improving treatment outcomes for patients at risk of TNR. By enhancing our understanding of the mechanisms underlying TNR and treatment augmentation on the other side, we can develop targeted interventions that complement psychotherapy and address the specific needs of each patient.

The sophisticated design and methods developed in this study for predicting TNR also offer promising avenues for clinical application. One viable implementation strategy is the integration of these models into digital therapy monitoring and feedback systems. The Trier Treatment Navigator serves as an excellent example of best practice in this domain.⁸¹ The Trier Treatment Navigator uses data-driven individual case predictions to estimate the risk of a patient being 'not on track' for therapy success, providing clinicians with actionable feedback. This includes tailored recommendations to adapt therapeutic strategies, focusing on overarching change mechanisms. The prospective validation of the Trier Treatment Navigator in a RCT demonstrated its utility, with patients monitored by the Trier Treatment Navigator achieving significantly greater symptom reduction (effect size $d=0.3$) when therapists adhered to the feedback within the first 10 sessions.⁸¹ This underscores the potential of feedback systems to enhance treatment outcomes,

contingent on clinician engagement and adherence to recommended adjustments.

Another compelling methodological approach for personalised treatment selection is the PAI, an interesting methodological approach for selecting the right treatment from the multitude of possible treatment options.⁸² It is based on an algorithm that is trained on a data set of patients who have been assigned to one of at least two (or more) treatment alternatives. In a first step, two predictive models are developed that predict response to treatment A or B. In the second step, the treatment outcome for the treatment received (so-called factual treatment) and the potential outcome of the treatment alternative (so-called counterfactual treatment) are then predicted in the test data set for each individual patient. This is possible when in both treatment groups the same potential predictors have been measured. If both models come to a very similar result, then treatment options A and B appear to work comparably well for this patient. However, the greater the difference between the two models, the more this patient would benefit from treatment stratification. Initial studies support the usefulness of the PAI for the treatment stratification of patients with a depressive disorder⁸³ or PTSD.⁸⁴ Interestingly, data-driven stratification using the PAI appears to be superior to clinical decision-making.⁶⁷ Nevertheless, a recent study with an external validation on two independent samples shows that the generalisability of PAI-based prediction is still limited⁸⁵ and thus points to a general problem of generalisability in predictive analytics.¹⁷ In addition, early PAI analysis pipelines appeared to be affected by data leakage and limited cross-validation procedures.⁸⁶ Nevertheless, the PAI offers an interesting approach to solving the problem of 'What works for whom?' in order to arrive at evidence-based stratification recommendations for individual patients. Thus, the PAI might be an interesting approach to select an optimal add-on treatment for those who have been predicted to be at risk for TNR.

In conclusion, TNR to psychotherapy—despite its well-documented general effectiveness—is a substantial problem in mental healthcare.⁶ This situation can be improved by (1) better understanding the phenomenon of TNR in order to inspire novel treatment approaches or augmentation strategies, (2) early identification of patients at high risk for TNR and (3) evidence-based recommendations to (non-pharmacological) augmentation strategies. Predictive analytics offers suitable methodological approaches to arrive at a data-based prediction or stratification recommendation. Big data and AI have undoubtedly arrived in psychotherapy research. However, limitations such as insufficient generalisability need to be met before predictive analytics models can be put into practice. In addition to aspects of data protection for sensitive health data, the technical implementation, manageability and acceptance of corresponding digital assistance systems must be considered.¹⁰ Despite all the possibilities that digital innovations in the field of predictive analytics can offer to support prognosis, treatment

selection and process monitoring, the clinical decision will always lie with the practitioner. However, with a balanced consideration of the risks and opportunities of AI-supported predictive analytics, it has the potential to optimise the effectiveness of psychotherapy for the benefit of our patients. The RU aims to make a significant contribution to further developing these areas.

ETHICS AND DISSEMINATION

The study has received ethical approval from the Institutional Ethics Committee of the Department of Psychology at Humboldt Universität zu Berlin (approval no. 2021-01) and the Ethics Committee of Charité-Universitätsmedizin Berlin (approval no. EA1/186/22).

Results will be disseminated through peer-reviewed journals and presentations at national and international conferences. Deidentified data and analysis scripts will be made available to researchers within the RU via a secure server, in line with ethical guidelines and participant consent. In compliance with European and German data protection regulations, patient data will not be publicly available through open science frameworks but may be shared with external researchers upon reasonable request and under appropriate data protection agreements.

All participants provide written informed consent after being fully informed about the study's aims, procedures, potential risks and benefits and are aware of their right to withdraw at any time without consequence.

To ensure confidentiality and data protection, participant data are pseudonymised, with personal information securely stored on encrypted servers accessible only to authorised personnel. We have protocols in place to monitor data integrity, participant safety and to address any adverse events throughout the study.

One year after the last patient out, data will be fully anonymised to allow for open science practices and data sharing, while at the same time securing patients' data protection rights. Summaries of findings will also be shared with participants and the public through institutional websites and, where appropriate, public events.

Author affiliations

¹Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

²Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, Berlin, Germany

³Berlin School of Mind and Brain, Berlin, Germany

⁴Charité - Universitätsmedizin Berlin, Berlin, Germany

⁵Department of Educational Sciences and Psychology, TU Dortmund University, Dortmund, Germany

⁶Department of Psychology, HMU Health and Medical University Erfurt GmbH, Erfurt, Germany

⁷Psychologische Hochschule Berlin, Berlin, Germany

⁸Clinical Psychology Intervention, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

⁹German Center for Mental Health (DZPG), Berlin-Potsdam Partner Site, Berlin, Germany

¹⁰Clinical Psychology and Psychotherapy, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

¹¹Hertie Institute for AI in Brain Health, University of Tübingen, Tübingen, Germany

Collaborators Humboldt-Universität zu Berlin (HU Berlin), Freie Universität Berlin (FU Berlin), Charité-Universitätsmedizin Berlin (Charité) and Psychologische Hochschule Berlin (PHB).

Contributors All authors made significant contributions to the conception and design of the study protocol. The protocol was written by TL, CU and UL and critically reviewed by FB, SE, LF, J-DH, SH, KH, FJ, NK, CK, BR, KR, NS and HW. Guarantor is TL. AI technology used: We used OpenAI's ChatGPT, an advanced AI language model, to assist in the revision of our manuscript. Reason for use: The primary purpose of employing ChatGPT was to enhance the linguistic quality of the text and to identify and correct grammatical, typographical, and stylistic errors, thereby improving overall clarity and readability. How AI was used: Specific sections of the manuscript were input into ChatGPT, which provided revised versions aimed at refining sentence structure, coherence and flow. The AI was solely used for language editing and did not influence the study's design, analysis, interpretation of results or conclusions. Summary of input, output and review: Input: Portions of the manuscript text were submitted to ChatGPT for revision. Output: Revised text with improved grammar, style and clarity. Review process: All AI-generated revisions and suggestions were meticulously reviewed and approved by the authors to ensure accuracy and maintain the integrity of the content. The authors retained full control over all substantive aspects of the manuscript, ensuring that the original research and findings remained unaffected by the AI-assisted edits. The authors confirm that the final manuscript reflects their own work and that all AI-assisted modifications have been thoroughly vetted for correctness and appropriateness. Any substantive changes to the content, analysis or conclusions were made solely by the authors without AI intervention.

Funding This study is funded by German Research Foundation (DFG project number: 442075332).

Competing interests KH is a scientific advisor and received virtual stock options of Mental Tech, which develops an AI-based chatbot providing mental health support. All other authors have no competing interest to declare.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Till Langhammer <http://orcid.org/0000-0001-6729-956X>

Ulrike Lueken <http://orcid.org/0000-0003-1564-4012>

REFERENCES

- 1 GBD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* 2022;9:137–50.
- 2 Gustavsson A, Svensson M, Jacobi F, et al. Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* 2011;21:718–79.
- 3 Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* 2013;382:1575–86.
- 4 Cuijpers P, Miguel C, Harrer M, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry* 2023;22:105–15.
- 5 Cuijpers P, Karyotaki E, Ciharova M, et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatr Scand* 2021;144:288–99.
- 6 Cuijpers P, Miguel C, Ciharova M, et al. Absolute and relative outcomes of psychotherapies for eight mental disorders: a systematic review and meta-analysis. *World Psychiatry* 2024;23:267–75.
- 7 Loerinc AG, Meuret AE, Twohig MP, et al. Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clin Psychol Rev* 2015;42:72–82.

- 8 Öst L-G, Havnen A, Hansen B, *et al.* Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clin Psychol Rev* 2015;40:156–69.
- 9 Santoft F, Axelsson E, Öst L-G, *et al.* Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. *Psychol Med* 2019;49:1266–74.
- 10 Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 2017;22:37–43.
- 11 Hammelrath L, Hilbert K, Heinrich M, *et al.* Select or adjust? How information from early treatment stages boosts the prediction of non-response in internet-based depression treatment. *Psychol Med* 2024;54:1641–50.
- 12 Hilbert K, Kunas SL, Lueken U, *et al.* Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: A machine learning approach. *Behav Res Ther* 2020;124:103530.
- 13 Hilbert K, Jacobi T, Kunas SL, *et al.* Identifying CBT non-response among OCD outpatients: A machine-learning approach. *Psychother Res* 2021;31:52–62.
- 14 Hornstein S, Forman-Hoffman V, Nazander A, *et al.* Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach. *Digit Health* 2021;7:20552076211060659.
- 15 Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. *Science* 2015;348:499–500.
- 16 Leehr EJ, Roesmann K, Böhnlein J, *et al.* Clinical predictors of treatment response towards exposure therapy in virtue in spider phobia: A machine learning and external cross-validation approach. *J Anxiety Disord* 2021;83:102448.
- 17 Chekroud AM, Hawrilenko M, Loho H, *et al.* Illusory generalizability of clinical prediction models. *Science* 2024;383:164–7.
- 18 Vieira S, Liang X, Guiomar R, *et al.* 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.
- 19 Cearnas M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. *Transl Psychiatry* 2019;9:271.
- 20 Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry* 2020;77:534–40.
- 21 Insel T, Cuthbert B, Garvey M, *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *AJP* 2010;167:748–51.
- 22 Kotov R, Krueger RF, Watson D, *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* 2017;126:454–77.
- 23 Braunstein LM, Gross JJ, Ochsner KN. Explicit and implicit emotion regulation: a multi-level framework. *Soc Cogn Affect Neurosci* 2017;12:1545–57.
- 24 Lincoln TM, Schulze L, Renneberg B. The role of emotion regulation in the characterization, development and treatment of psychopathology. *Nat Rev Psychol* 2022;1:272–86.
- 25 Paul S, Simon D, Endrass T, *et al.* Altered emotion regulation in obsessive-compulsive disorder as evidenced by the late positive potential. *Psychol Med* 2016;46:137–47.
- 26 Wackerhagen C, Veer IM, Erk S, *et al.* Amygdala functional connectivity in major depression - disentangling markers of pathology, risk and resilience. *Psychol Med* 2020;50:2740–50.
- 27 Chekroud AM, Zotti RJ, Shehzad Z, *et al.* Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016;3:243–50.
- 28 MacNamara A, Joyner K, Klawohn J. Event-related potential studies of emotion regulation: A review of recent progress and future directions. *Int J Psychophysiol* 2022;176:73–88.
- 29 Paul S, Kathmann N, Elsner B, *et al.* Neural correlates of emotional reactivity predict response to cognitive-behavioral therapy in obsessive-compulsive disorder. *J Affect Disord* 2022;308:398–406.
- 30 Derogatis LR. The Brief Symptom Inventory (BSI): Administration, Scoring & Procedures Manual-II. Clinical Psychometric Research, 1992.
- 31 Hiller W, Schindler A, Andor T, *et al.* Vorschläge zur Evaluation regulärer Psychotherapien an Hochschulambulanzen im Sinne der Phase-IV-Therapieforschung. *Zeitschrift Für Klinische Psychologie Und Psychotherapie* 2011;40:22–32.
- 32 Guy W. ECDEU Assessment Manual for Psychopharmacology. US Department of Health, Education, and Welfare, Public Health Service, 1976.
- 33 Velten J, Margraf J, Benecke C, *et al.* Methodenpapier zur Koordination der Datenerhebung und -auswertung an Hochschul- und Ausbildungsambulanzen für Psychotherapie (KODAP). *Zeitschrift Für Klinische Psychologie Und Psychotherapie* 2017;46:169–75.
- 34 Teismann T, Forkmann T, Glaesmer H, *et al.* Prevalence of suicidal ideation in German psychotherapy outpatients: A large multicenter assessment. *J Affect Disord* 2024;351:971–6.
- 35 Figueroa RL, Zeng-Treitler Q, Kandula S, *et al.* Predicting sample size required for classification performance. *BMC Med Inform Decis Mak* 2012;12:8.
- 36 Hamm A. Spezifische Phobien. Göttingen: Hogrefe, 2006.
- 37 Hautzinger M. Kognitive Verhaltenstherapie Bei Depressionen. Weinheim: Beltz, 2013.
- 38 Lang T, Helbig-Lang S, Westphal D. Expositions-basierte Therapie Der Panikstörung Mit Agoraphobie. Goettingen: Hogrefe, 2011.
- 39 Stangier U, Heidenreich T, Phobien PMS. Ein Kognitiv-Verhaltenstherapeutisches Behandlungsmanual. Weinheim: Beltz, 2009.
- 40 Joormann J, Stanton CH. Examining emotion regulation in depression: A review and future directions. *Behav Res Ther* 2016;86:35–49.
- 41 Sheppes G, Suri G, Gross JJ. Emotion regulation and psychopathology. *Annu Rev Clin Psychol* 2015;11:379–405.
- 42 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
- 43 Harris PA, Taylor R, Minor BL, *et al.* The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
- 44 Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12–9.
- 45 Heinig I, Pittig A, Richter J, *et al.* Optimizing exposure-based CBT for anxiety disorders via enhanced extinction: Design and methods of a multicentre randomized clinical trial. *Int J Methods Psychiatr Res* 2017;26:e1560.
- 46 Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959;32:50–5.
- 47 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–9.
- 48 Mataix-Cols D, Fernández de la Cruz L, Nordsletten AE, *et al.* Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. *World Psychiatry* 2016;15:80–1.
- 49 Storch EA, Rasmussen SA, Price LH, *et al.* Development and psychometric evaluation of the Yale-Brown Obsessive-Compulsive Scale—Second Edition. *Psychol Assess* 2010;22:223–32.
- 50 Varker T, Kartal D, Watson L, *et al.* Defining response and nonresponse to posttraumatic stress disorder treatments: A systematic review. *Clin Psychol: Sci Pract* 2020;27.
- 51 Weathers FW, Bovin MJ, Lee DJ, *et al.* The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychol Assess* 2018;30:383–95.
- 52 Waller L, Erk S, Pozzi E, *et al.* ENIGMA HALPipe: Interactive, reproducible, and efficient analysis for resting-state and task-based fMRI data. *Hum Brain Mapp* 2022;43:2727–42.
- 53 Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? 2022.
- 54 Arik SO, Pfister T. TabNet: Attentive Interpretable Tabular Learning. *AAAI* 2020;35:6679–87.
- 55 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, *et al.*, eds. *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc, 2017: 4765–74.
- 56 Böhle M, Eitel F, Weygandt M, *et al.* Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. *Front Aging Neurosci* 2019;11:194.
- 57 Eitel F, Soehler E, Bellmann-Strobl J, *et al.* Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *Neuroimage Clin* 2019;24:102003.
- 58 Schulz M-A, Koch A, Guarino VE, *et al.* Data augmentation via partial nonlinear registration for brain-age prediction. In: Abdulkadir A, Bathula DR, Dvornek NC, eds. *Machine Learning in Clinical Neuroimaging*. Cham: Springer Nature Switzerland, 2022: 169–78.
- 59 Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data* 2016;3:9.

- 60 Montavon G, Binder A, Lapuschkin S, *et al.* Layer-wise relevance propagation: an overview. In: Samek W, Montavon G, Vedaldi A, *et al.*, eds. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham: Springer International Publishing 2019:193–209.
- 61 Cui C, Yang H, Wang Y, *et al.* Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review. *Prog Biomed Eng* 2023;5:022001.
- 62 Lahat D, Adali T, Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proc IEEE* 2015;103:1449–77.
- 63 Rane RP, Kim J, Umesha A, *et al.* DeepRepViz: Identifying Confounders in Deep Learning Model Predictions. *arXiv* 2023.
- 64 Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- 65 Symons M, Feeney GFX, Gallagher MR, *et al.* Machine learning vs addiction therapists: A pilot study predicting alcohol dependence treatment outcome from patient data in behavior therapy with adjunctive medication. *J Subst Abuse Treat* 2019;99:156–62.
- 66 Symons M, Feeney GFX, Gallagher MR, *et al.* Predicting alcohol dependence treatment outcomes: a prospective comparative study of clinical psychologists versus “trained” machine learning models. *Addiction* 2020;115:2164–75.
- 67 van Bronswijk SC, Lemmens L, Huibers MJH, *et al.* Selecting the optimal treatment for a depressed individual: Clinical judgment or statistical prediction? *J Affect Disord* 2021;279:149–57.
- 68 Kim Y-K, ed. *Treatment Resistance in Psychiatry: Risk Factors, Biology, and Management*. Singapore: Springer, 2019.
- 69 Balzus L, Klawohn J, Elsner B, *et al.* Non-invasive brain stimulation modulates neural correlates of performance monitoring in patients with obsessive-compulsive disorder. *Neuroimage Clin* 2022;35:103113.
- 70 Cobb AR, O'Connor P, Zaizar E, *et al.* tDCS-Augmented in vivo exposure therapy for specific fears: A randomized clinical trial. *J Anxiety Disord* 2021;78:102344.
- 71 Compère L, Siegle GJ, Riley E, *et al.* Enhanced efficacy of CBT following augmentation with amygdala rtfMRI neurofeedback in depression. *J Affect Disord* 2023;339:495–501.
- 72 De Smet S, Baeken C, Semmink N, *et al.* Non-invasive vagal nerve stimulation enhances cognitive emotion regulation. *Behav Res Ther* 2021;145:103933.
- 73 Hertenstein E, Trinca E, Schneider CL, *et al.* Augmentation of Psychotherapy with Neurobiological Methods: Current State and Future Directions. *Neuropsychobiology* 2021;80:437–53.
- 74 Kozel FA, Motes MA, Didehban N, *et al.* Repetitive TMS to augment cognitive processing therapy in combat veterans of recent conflicts with PTSD: A randomized clinical trial. *J Affect Disord* 2018;229:506–14.
- 75 Fincham GW, Kartar A, Uthaug MV, *et al.* High ventilation breathwork practices: An overview of their effects, mechanisms, and considerations for clinical applications. *Neuroscience & Biobehavioral Reviews* 2023;155:105453.
- 76 Gaudlitz K, Plag J, Dimeo F, *et al.* Aerobic exercise training facilitates the effectiveness of cognitive behavioral therapy in panic disorder. *Depress Anxiety* 2015;32:221–8.
- 77 Guzik AG, Cooke DL, Gage N, *et al.* CBT-Plus: A meta-analysis of cognitive behavioral therapy augmentation strategies for obsessive-compulsive disorder. *J Obsessive Compuls Relat Disord* 2018;19:6–14.
- 78 Hang Y, Xu L, Wang C, *et al.* Can attention bias modification augment the effect of CBT for anxiety disorders? A systematic review and meta-analysis. *Psychiatry Res* 2021;299:113892.
- 79 Meyer E, Souza F, Heldt E, *et al.* A randomized clinical trial to examine enhancing cognitive-behavioral group therapy for obsessive-compulsive disorder with motivational interviewing and thought mapping. *Behav Cogn Psychother* 2010;38:319–36.
- 80 Michael T, Schanz CG, Mattheus HK, *et al.* Do adjuvant interventions improve treatment outcome in adult patients with posttraumatic stress disorder receiving trauma-focused psychotherapy? A systematic review. *Eur J Psychotraumatol* 2019;10:1634938.
- 81 Lutz W, Deisenhofer A-K, Rubel J, *et al.* Prospective evaluation of a clinical decision support system in psychological therapy. *J Consult Clin Psychol* 2022;90:90–106.
- 82 DeRubeis RJ, Cohen ZD, Forand NR, *et al.* The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One* 2014;9:e83875.
- 83 Huibers MJH, Cohen ZD, Lemmens LHJM, *et al.* Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. *PLoS One* 2015;10:e0140771.
- 84 Deisenhofer A-K, Delgadillo J, Rubel JA, *et al.* Individual treatment selection for patients with posttraumatic stress disorder. *Depress Anxiety* 2018;35:541–50.
- 85 van Bronswijk SC, Bruijniks SJE, Lorenzo-Luaces L, *et al.* Cross-trial prediction in psychotherapy: External validation of the Personalized Advantage Index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression. *Psychother Res* 2021;31:78–91.
- 86 Meinke C, Hornstein S, Schmidt J, *et al.* Advancing the personalized advantage index (PAI): a systematic review and application in two large multi-site samples in anxiety disorders. *Psychol Med* 2024;54:4843–55.
- 87 Hariri AR, Mattay VS, Tessitore A, *et al.* Serotonin transporter genetic variation and the response of the human amygdala. *Science* 2002;297:400–3.
- 88 Walter H, von Kalckreuth A, Schardt D, *et al.* The temporal dynamics of voluntary emotion regulation. *PLoS ONE* 2009;4:e6726.
- 89 Erk S, Mikschl A, Stier S, *et al.* Acute and sustained effects of cognitive emotion regulation in major depression. *J Neurosci* 2010;30:15726–34.
- 90 Erk S, Kleczar A, Walter H. Valence-specific regulation effects in a working memory task with emotional context. *Neuroimage* 2007;37:623–32.