Moment-to-moment brain signal variability reliably predicts psychiatric treatment outcome

Kristoffer N T Månsson[a,b,c], PhD*

Leonhard Waschke[a,b], PhD

Amirhossain Manzouri[d], MSc

Tomas Furmark[e], PhD

Håkan Fischer[d,f], PhD

Douglas D Garrett[a,b], PhD

[a] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
[b] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin/London
[c] Centre of Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[d] Department of Psychology, Stockholm University, Stockholm, Sweden
[e] Department of Psychology, Uppsala University, Uppsala, Sweden
[f] Stockholm University Brain Imaging Centre, Stockholm, Sweden

***Address for correspondence:** kristoffer.mansson@ki.se; mansson@mpib-berlin.mpg.de

Author Contributions: K.M., T.F., H.F., and D.G. planned and designed the study. K.M. collected the data. K.M., A.H., L.W., and D.G. established and developed the methods, and ran statistical analyses. K.M., L.W., and D.G. drafted the manuscript, and all authors discussed the results, conclusions and edited the manuscript.

Short/running title: Brain variability predicts treatment outcome

Keywords: fMRI signal variability, resting-state, prediction, social anxiety disorder, CBT

This document includes:

- Main text
- Figures 1 to 5
- Table 1
- Supplementary Material 1

Abstract

BACKGROUND: Biomarkers of psychiatric treatment response remain elusive. Functional magnetic resonance imaging (fMRI) has shown promise, but low reliability has limited the utility of typical fMRI measures (e.g., average brain signal) as harbingers of treatment success. Notably, although historically considered a source of "noise", temporal brain signal variability continues to gain momentum as a sensitive and reliable indicator of individual differences in neural efficacy, yet has not been examined in relation to psychiatric treatment outcomes.

METHODS: Forty-five patients with social anxiety disorder were scanned twice (11 weeks apart) using simple task-based and resting-state fMRI to capture moment-to-moment neural variability. After fMRI test-retest, patients underwent a 9-week cognitive-behavioral therapy. Multivariate modeling and reliability-based cross-validation were utilized to perform brain-based prediction of treatment outcomes.

RESULTS: Task-based brain signal variability was the strongest contributor in a treatment outcome prediction model (total $r_{\text{ACTUAL,PREDICTED}}$=.77) - outperforming self-reports, resting-state neural variability, and standard mean-based measures of neural activity. Notably, task-based brain signal variability showed excellent test-retest reliability (intraclass correlation coefficient=.80), even with a task length less than 3 minutes long.

CONCLUSIONS: Rather than a source of undesirable "noise", moment-to-moment fMRI signal variability may instead serve as a highly reliable and efficient prognostic indicator of clinical outcome.

Moment-to-moment brain signal variability reliably predicts psychiatric treatment outcome

Biomarkers of psychiatric treatment response remain elusive. The search for such biomarkers is of particular importance given that subjective ratings of pre-treatment symptom severity often fail to predict treatment outcomes for a range of common psychiatric disorders (e.g., 1). Non-invasive functional magnetic resonance imaging (fMRI) serves as one theoretically viable alternative for prediction of treatment outcomes (e.g., 2). However, typical neuroimaging-based treatment outcome prediction models have been heavily critiqued under the argument that thousands of patients are needed to successfully establish treatment predictors (3). Further, recent meta-analyses demonstrate low overall reliability of both task and resting-state fMRI using standard measures (e.g., functional connectivity and average brain signals) (4,5). Inaccurate predictions of treatment outcomes will thus necessarily remain (despite large-scale, resource-intensive efforts) if the same unreliable measures continue to be utilized. We need a different approach.

Grossly underappreciated in the clinical domain, evidence continues to mount revealing that moment-to-moment fluctuations in brain activity (i.e., brain signal variability) can viably index the adaptability and effectiveness of neural systems. For example, cognitive performance has been repeatedly linked to brain signal variability (6–8), the level of which can also be boosted pharmacologically (9,10). Crucially, although the unique predictive power of signal variability can be more than five times that of conventional mean signal-based approaches (11) and initial evidence for measurement reliability is promising (12), no treatment outcome prediction studies to date have examined brain signal variability.

Here, we provide a first test of the predictive utility of brain signal variability in relation to cognitive-behavioral therapy (CBT) outcomes in social anxiety disorder (SAD) patients. CBT for SAD is an evidence-based treatment intended to limit the avoidance of social situations and reduce self-focused attention - hallmarks of the disorder (13). Although the average group-level effect of CBT can be strong (14), there is considerable variability across-patient response rates, with many SAD patients remaining symptomatic after treatment (15). At its core, CBT is intended to help patients adapt to momentary, social anxiety-provoking demands in the internal and external environment (13). This prompts the question of whether such socially relevant demands may be reflected in moment-to-moment fluctuations in brain signals (i.e., fMRI-variability), and whether brain signal variability could provide a novel predictive signature of CBT treatment outcome in SAD.

To this end, 45 SAD patients underwent fMRI twice during an 11-week test-retest period before enrollment in a 9-week CBT. We investigated the reliability and predictive power of moment-to-moment neural variability at rest and during a disorder-relevant socio-affective task, while further comparing the predictive accuracy of variability to conventional measures (mean neural responses and behavioral self-reports). Our results show that on-task moment-to-moment brain signal variability provides maximal reliability and treatment outcome prediction for SAD.

METHODS and MATERIALS

The study was registered at ClinicalTrials.gov (id: NCT02592564) and ethical approval was obtained from the regional committee at Umeå University, Umeå, Sweden. All participants gave written informed consent prior to participation.

General procedure and recruitment of patients

Individuals experiencing social anxiety (>18 years of age) and seeking treatment were targeted via media advertisements, provided self-reports, and participated in a diagnostic interview as part of the screening. Included participants underwent internet-delivered CBT for SAD for 9 weeks. Before CBT, patients underwent an 11-week test-retest period (pre-treatment) during which we assessed self-reported social anxiety symptoms and recorded two separate fMRI scans (i.e., baseline 1 and baseline 2; see Figure 1A). Multiple baseline measures of brain and behavior were included to control for standard confounds (e.g., regression to the mean and spontaneous remission) and to directly estimate test-retest reliability. Forty-six patients with a primary SAD diagnosis (as determined by structured clinical interviews) took part in the current study. Recorded fMRI data contained outliers from one patient (Mahalanobis distance = 20.2; see Supplementary Material Figure S1) which was

excluded from all analyses. Patients in the sample (*n* = 45) did not receive concurrent psychological treatment at any point during the present study, and if treated with a psychotropic medication (*n* = 4, 8.9%), agreed to maintain a stable dose at least 3 months prior to and during the current study. All patients remained throughout the intervention and took part in post-treatment assessments.

Cognitive behavioral therapy

Briefly, internet-delivered CBT for SAD is a guided self-help intervention. Each week, a module containing text and homework assignments based on CBT was provided. All patients had identical treatment materials, just as in previous randomized controlled trials (16,17). Further details have been described elsewhere (e.g., 18–20). Patients were in weekly contact with a clinical psychologist who provided written feedback and guidance via a secured internet platform. Treatment compliance measures showed that most patients completed every module of the 9-week treatment period (see Supplement, page 3).

Clinical outcome measures

We used the Liebowitz social anxiety scale, self-report version (LSAS-SR), a reliable and gold-standard questionnaire to assess treatment-related changes in social anxiety symptoms (21). Structured clinical interviews, as well as the Clinical Global Impression-Improvement (CGI-I) scale (22) were administrated at post-treatment. Secondary anxiety outcomes, as well as depressive and insomnia symptom assessments are presented in the Supplement, page 4.

Functional neuroimaging, preprocessing, and variability estimation

Pre-treatment fMRI was performed twice, separated by 11 weeks (first and second baseline). Neither the time of scanning nor patients' subjective sleepiness ratings differed between baselines (see also Supplement, page 4).

In each scanning session, we first recorded resting-state fMRI, followed by task-fMRI. Resting-state recordings lasted 340 sec and were performed with eyes open (fixation-cross present on screen). As displayed in Figure 1B, during the socio-affective face task, patients passively viewed emotional faces (happy/fearful male/female) across two blocks (23). In each block, alternations of a single face (200 ms) and fixation cross (300 ms) were presented for a period of 80 sec (160 sec of face stimulation in total). Stimulus order (happy/fearful; male/female) was counter-balanced across patients. Fixation blocks also occurred before (20 sec), in-between (30 sec), and after (20 sec) the face stimulus blocks (see Figure 1B).

Brain images were collected on a 3T General Electric Discovery MR 750 scanner with a 32-channel head coil at the Umeå Centre for Functional Brain Imaging (Umeå, Sweden). The preprocessing pipeline included manual denoising by examining all functional volumes for artifacts via independent component analysis (ICA) (24). To compute the temporal standard deviation of blood-oxygen-level-dependent (BOLD) signals per voxel (SD_{BOLD}) we first subtracted the block mean and concatenated signals across all blocks before computing voxel-wise SD_{BOLD} across this concatenated time series (25). We also compared SD_{BOLD} results to typical mean fMRI activity ($MEAN_{BOLD}$) during the socio-affective face task. All MRI parameters (including anatomical scans), preprocessing steps, and signal variability estimation are described in detail in the Supplement, pages 4-5.

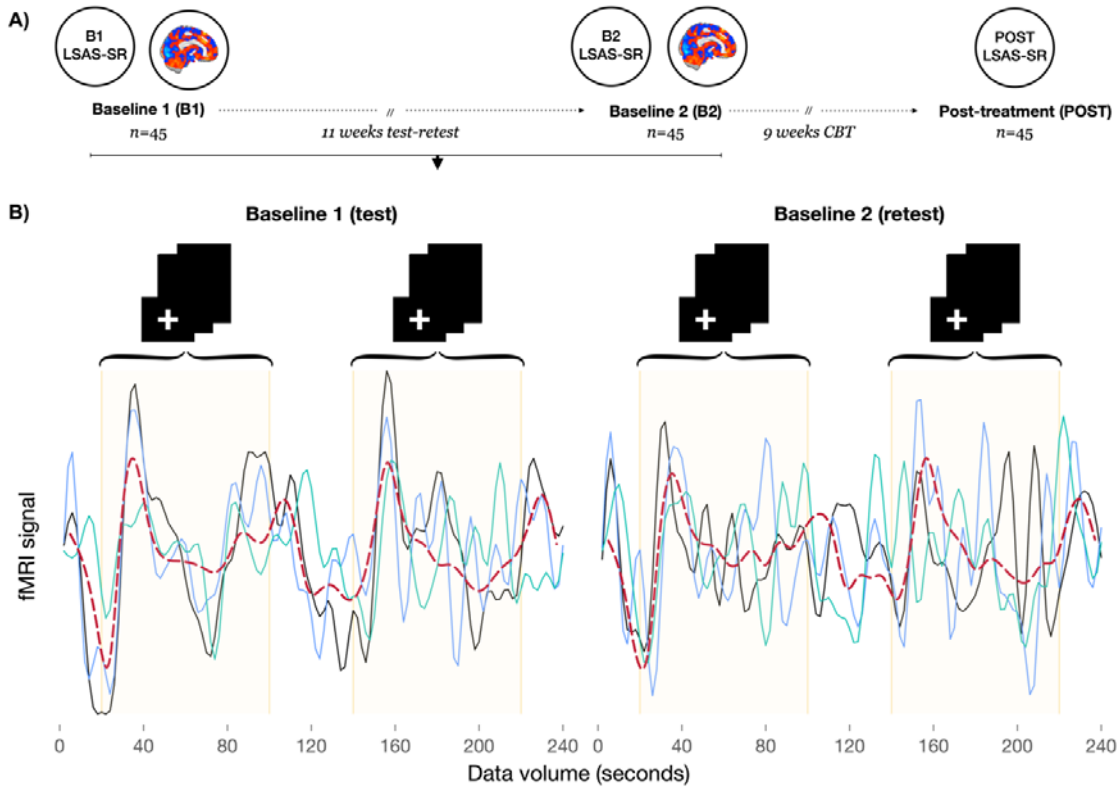

Figure 1. Study design and experimental task. A) Forty-five patients provided behavioral (e.g., LSAS-SR) and brain data (i.e., fMRI) at two time-points: baseline 1 (B1) and baseline 2 (B2) separated by 11 weeks. Further, post-treatment behavioral data after 9 weeks of Internet-delivered CBT were also collected. **B)** Example of visual cortex (mean-centered) fMRI time series (data volumes in sec) within each baseline session from three random patients (i.e., solid green/blue/black lines). The dashed red line represents the average (median cubic spline) signal across all patients in the current study (*n*=45). Vertical solid/yellow lines represent stimuli onsets: face 200 ms + 300 ms fixation, with 160 repetitions totalling 80 sec for each block. The non-shaded parts of the time-series represent fixation blocks (i.e., continuous presentation of a fixation cross). **Abbreviations:** CBT, cognitive-behavioral therapy; LSAS-SR, Liebowitz social anxiety scale, self-report version; fMRI, Functional magnetic resonance imaging; Post, Post-treatment;

Statistical modeling

Estimating fMRI correlates of clinical outcome

Clinical outcomes were defined as continuous LSAS-SR delta scores (i.e., each questionnaire's total score at post-treatment minus pre-treatment), capturing the overall change in symptom severity. To examine the association between BOLD activity and treatment success, we utilized a partial least squares (PLS) analysis (27,28) in MATLAB (Ver 9.6.0.1072779, R2019a), where estimation of neuro-behavioral correlations are performed in latent space. See detailed benefits of PLS (relative to standard GLM) in the Supplement, page 7). In brief, PLS models were based on a correlation matrix capturing the between-subject correlation (Pearson's *r*) of brain activity (e.g., SD_{BOLD}) in each voxel (51609 per subject) and subject-wise delta total LSAS-SR score (post-treatment minus baseline 1, B1). Next, this correlation matrix was decomposed using singular value decomposition (SVD), which yielded voxel-based saliences (weights) proportionate to the correlation strength between BOLD activity (e.g., SD_{BOLD}) and delta LSAS-SR scores. For every subject, so-called "brain scores" were then calculated via the dot product of these saliences with voxel-wise BOLD values. To estimate the robustness of voxel saliencies, 1000 bootstraps with replacement were utilized, and the division of each voxel-wise salience by its corresponding bootstrap standard error yielded pseudo-z estimates of robustness typically referred to as "bootstrap ratios (BSRs)". For all other regression models aimed at predicting treatment outcome, we performed 1000 bootstraps (with replacement) to estimate bootstrapped confidence intervals. Taken together, bootstrapping was used to estimate strength and

confidence intervals of effects, while permutations were used to test the significance of the found effects. PLS brain maps are found in Figures S2-S4 and peak coordinates, bootstrap ratios, and cluster sizes are noted in Tables S2-S4.

To compare relative predictive utility and reliability, this PLS approach was utilized separately to test how different brain measures (i.e., socio-affective face task-based SD_{BOLD}, face task-based $MEAN_{BOLD}$, and resting-state SD_{BOLD}) linked to treatment-related LSAS-SR changes.

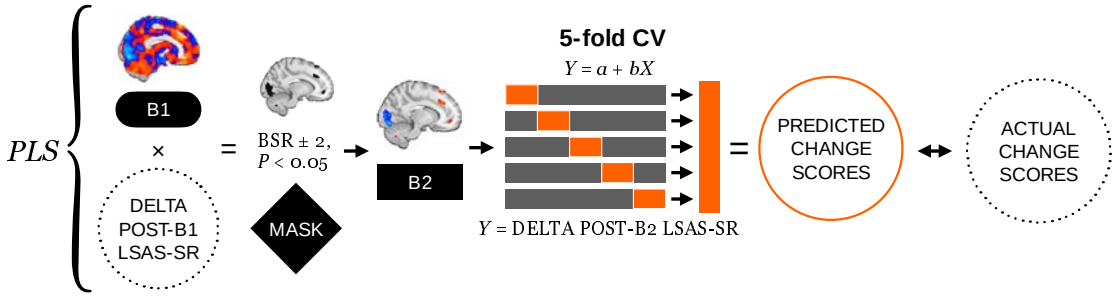

Figure 2. A general overview of the analytic framework. The analysis of neuro-behavioral correlations between BOLD signals at B1 and the treatment outcome (i.e., delta LSAS-R score post-treatment minus B1) was computed for all 45 subjects via behavioral PLS. A bootstrap-based mask was created (BSR ± 2) based on B1 data. To compute "reliability-based cross validation", weights from this mask were applied to corresponding voxels in B2 brain data to permit extraction of subject-specific brain scores at B2 (i.e., no additional PLS model was run on B2 data). Prediction of treatment outcome-based change scores was then performed using 5-fold cross-validation, from which empirical treatment-based change scores were correlated with predicted scores. **Abbreviations:** LSAS-SR, Liebowitz social anxiety scale, self-report version; fMRI, Functional magnetic resonance imaging; CBT, Cognitive behavior therapy; CV, Cross-validation; BSR, Bootstrapped ratio; PLS, Partial least squares; B1, Baseline 1; B2, Baseline 2; Post, Post-treatment;

Reliability-based cross-validation framework for brain and behavioral prediction of treatment outcome

One key goal of the present work is to establish and compare the relative strength and reliability of different brain and behavioral predictors of treatment outcome. To further characterize the associations identified by PLS, we employed a unique two-step reliability-based cross-validation framework (see Figure 2).

First, as described above, we computed PLS models linking brain activity (separately for each brain measure) and reductions in total LSAS-SR scores of all participants based on the first baseline (delta LSAS-SR Post-B1) MRI recording. Second, the resulting voxel-wise BSRs for each model were thresholded at ±2.0 while excluding all clusters smaller than 20 voxels (Figures S2-S4 and Tables S2-S4). As control analyses, a similar PLS model was computed based on the second baseline, and demonstrates that neuro-behavioral relationships as captured by PLS-based brain scores (separated by 11 weeks) are highly correlated (r = .77; Figure S5). Third, we applied the corresponding weights to fMRI data recorded during the second baseline measurement (B2; 11 weeks after B1) to extract subject-specific brain scores without re-estimating PLS models. We also generated brain scores from the B1 and the B2 measurements (i.e., weights from B1 applied to either of these), and they were all normally distributed (Figure S6). Applying B1 weights to B2 data permits a form of "metric invariance" (29) (here, the fixing of model weights between the two baseline periods), allowing an assessment of stability between these measurement points. Finally, these B2 brain scores were used to estimate linear relationships between BOLD activity and changes in total LSAS-SR scores (delta LSAS-SR Post-B2) within a 5-fold cross validation framework. For each fold, linear coefficients were estimated on a subset of subjects (training set, n=36) before being applied to another, non-overlapping subset of subjects (test set, n=9), based on which predictions were generated. We then calculated Pearson correlations between predicted and observed LSAS-SR changes, and the mean absolute scaled error (MASE) (26) was used as a metric for the relative comparison of predictors (see Supplement, page 12 for details).

While our reliability-based modeling approach does not, by definition, seek out-of-sample prediction *per se* due to the common topographical weight map estimated at B1 and fixed at B2 (see above), we nevertheless test for "generalizability" of our models on two intertwined levels. First,

because PLS weights from the B1 fMRI measurement are applied to B2 fMRI data without re-running PLS, any statistically meaningful prediction of treatment success can *only* emerge if the link between BOLD activity and treatment outcome is similar in both magnitude and topographical distribution across two completely separate fMRI recordings 11 weeks apart. Second, our 5-fold cross validation approach further limits model overfitting (in the presence of our limited sample size). Building on the direct comparisons of strength and reliability of different brain measures for treatment prediction, future work can target larger-scale out of sample prediction of treatment outcomes.

Standard intraclass (ICC) and Pearson's *r* correlation coefficients (30) were also calculated to determine test-retest reliability between the two baseline measurements (see Supplement, page 15).

Data and code availability

All code and statistical software commands will be available online (https://github.com/LNDG/). Due to ethics constraints, we cannot at present make the raw patient data openly available; please contact the first author (K.M.) to discuss potential routes to data access.

RESULTS

Treatment outcomes

The primary social anxiety outcome measure (LSAS-SR) decreased 33.46 points on average from screening to post-treatment (Figure 3A). Large within-group Cohen's d effect sizes (>1.49) were observed for both LSAS-SR and secondary outcomes, all permuted *P*s<0.001. The clinician-administered CGI-I interviews found 32 of 45 patients' mental health to be much (48.9%) or very much (22.2%) improved at post-treatment. See Tables S5-S6 and Figures S7-S9 for a detailed presentation of all clinical outcomes.

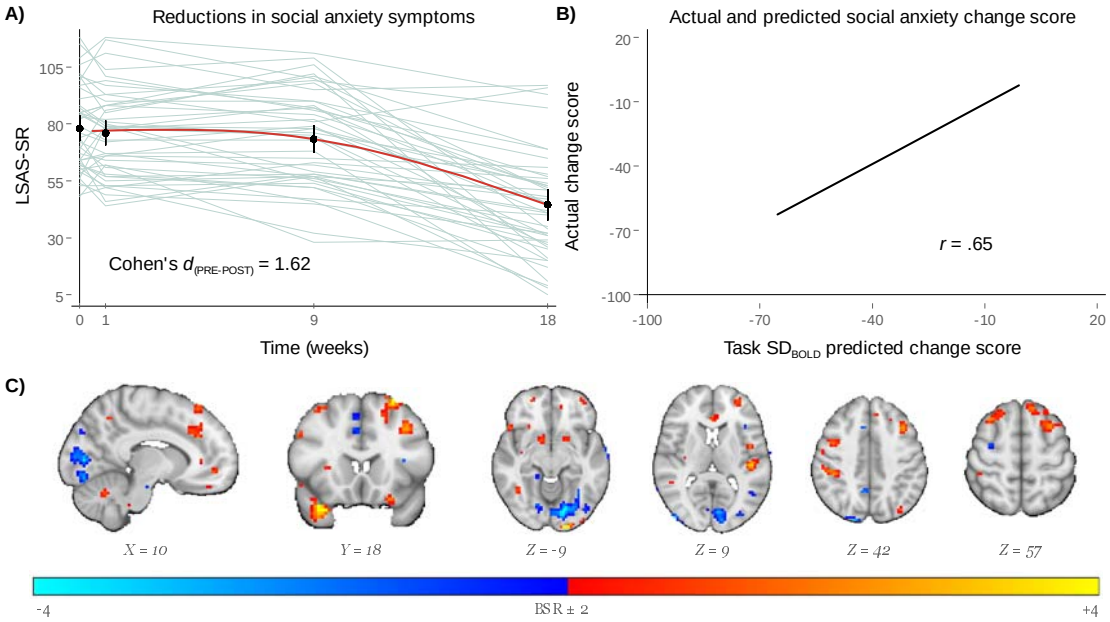

Figure 3. Reductions in social anxiety symptoms and task-related brain signal variability as a predictor of treatment outcome. A) Change in the primary social anxiety outcome LSAS-SR from screening, baseline 1, baseline 2 to post-treatment. The solid line represents the median cubic spline. **B)** Task-based SD_{BOLD} predicted treatment change score is strongly related to empirical change scores. **C)** Task-based SD_{BOLD} spatial pattern reflecting treatment outcome. Blue regions=lower SD_{BOLD} associated with better treatment outcome; yellow/red regions: higher SD_{BOLD} associated with better treatment outcome. *X Y Z* below the brains represent MNI coordinates. Further, unthresholded statistical maps are available at NeuroVault.org (https://identifiers.org/neurovault.collection:9030). **Abbreviations:** LSAS-SR, Liebowitz social anxiety scale, self-report version; SD_{BOLD}, standard deviation of BOLD; BOLD, Blood-oxygen-level-dependent imaging;

Task-related brain signal variability strongly predicts treatment outcome

Moment-to-moment brain signal variability during emotional face processing robustly predicted social anxiety change scores (post-pre CBT; 5-fold cross-validated; $r_{ACTUAL,PREDICTED}$=.65, MASE=.54, permuted P<0.001; Figure 3B). Specifically, low signal variability in the right visual cortex, and high variability in the anterior cingulate, medial prefrontal, and temporal cortices predicted larger reductions in social anxiety symptoms (see Figure 3C, Figure S2 and Table S2 for a complete presentation of neural activations and data density plots). The predictive power of task-related SD_{BOLD} remained nearly identical even when the data volume was reduced by either 50% (80 sec; $r_{ACT,PRED}$=.65, MASE=.53, permuted P<0.001) or 75% (40 sec; $r_{ACT,PRED}$=.62, MASE=.52, permuted P<0.001) - see also Figure 4 and Tables S7 for details of various multiple regression models spanning brain measures and data volumes.

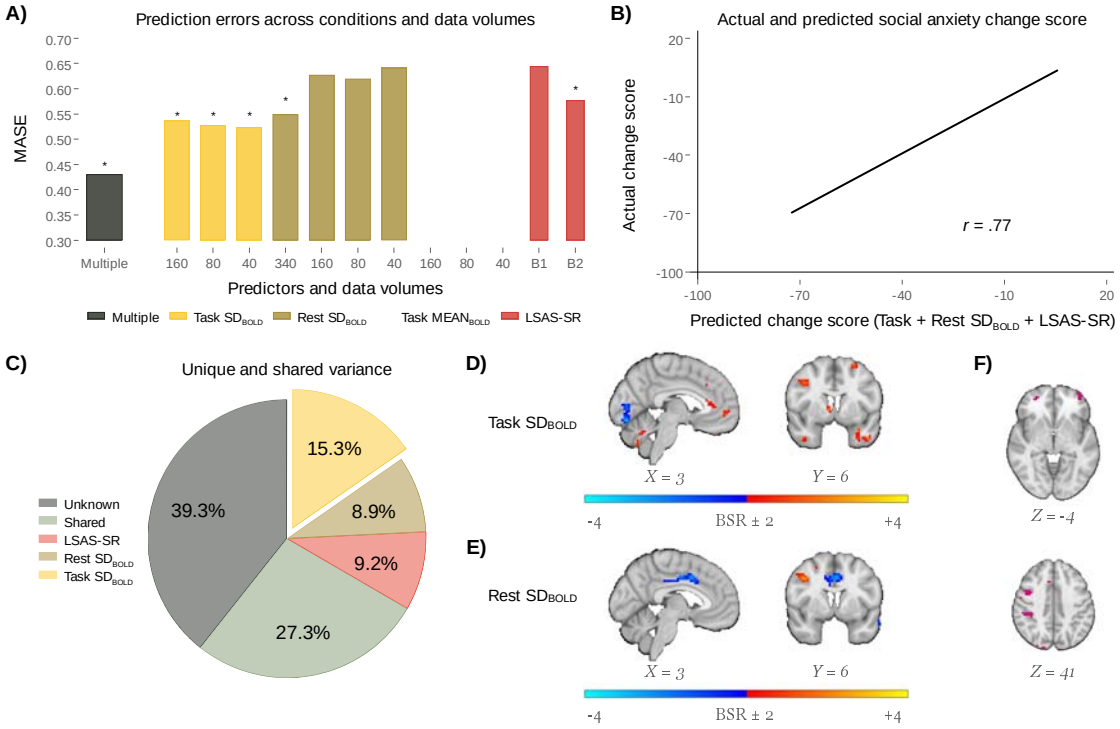

Figure 4. Treatment outcome prediction accuracies and brain signal variability. A) Prediction accuracy (i.e., MASE) for each condition (i.e., task SD_{BOLD}, resting-state SD_{BOLD}, task $MEAN_{BOLD}$, LSAS-SR at B1 and B2) and across data volumes (i.e., 40, 80, 160 and 340 sec). * denotes significant (permuted P<0.001) zero-order prediction models noted in Table S7. Lower values indicate better model performance. **B)** 5-fold cross-validated correlation between empirical and predicted treatment change scores using all zero-order significant conditions (i.e., 160 sec task SD_{BOLD}, 340 sec resting-state SD_{BOLD}, and LSAS-SR at B2). **C)** Unique and shared variance (R^2) between model predictors and treatment outcome. **D)** Task-related SD_{BOLD} and **E)** resting-state SD_{BOLD} spatial pattern reflecting treatment outcome. Blue regions represent less variability predicting better outcome, whereas yellow/red regions represent higher fMRI signal variability predicting better outcome. **F)** Displays overlapping activations between task SD_{BOLD} (160 sec) and resting-state SD_{BOLD} (340 sec). In figure F, the spatial pattern for each condition represents a binary mask. *X Y Z* below the brains represent MNI coordinates. Unthresholded statistical maps are available at NeuroVault.org (https://identifiers.org/neurovault.collection:9030). **Abbreviations:** MASE, Mean absolute scaled error; LSAS-SR, Liebowitz social anxiety scale, self-report version; B1, Baseline 1; B2, Baseline 2; SD_{BOLD}, Standard deviation of BOLD; $MEAN_{BOLD}$, Average BOLD activity; BSR, Bootstrap ratio; BOLD, Blood-oxygen-level-dependent imaging;

Gauging the relative predictive utility of task-based BOLD variability

In a multiple regression model including all potential behavioral and brain-based predicted social anxiety change scores (and equal data volumes of 160 sec for all brain measures), task-related SD_{BOLD} (β=.61, permuted P<0.001) dominantly outperformed (A) resting-state SD_{BOLD} (β=.26, permuted P=0.090), (B) task-related $MEAN_{BOLD}$ (β=-.07, permuted P=0.621), and (C) pre-treatment

social anxiety severity (second baseline LSAS-SR, β=.22, permuted P=0.186). The model accounted for 54% of the variance in the social anxiety change score (Table S8). Although self-reported social anxiety at pre-treatment did not predict treatment outcome within our full model, a moderate zero-order correlation indicated that patients suffering from a more severe SAD exhibited greater reductions in social anxiety ($r_{ACT,PRED}$=.45, permuted P<0.001) - see also Table 1 and Figure 4A for a complete presentation of statistical results. In a second model including the same predictors, but instead using the brain score from the full available resting-state SD_{BOLD} data volume (340 sec). Here, we found that resting-state SD_{BOLD} also uniquely predicted treatment outcome (β=.34, permuted P=0.039), but task-related SD_{BOLD} remained the strongest treatment outcome predictor (β=.41, permuted P=0.018) despite relying on less than half the data volume (160 sec) compared to resting-state (Table S9). Furthermore, we demonstrate good (although slightly reduced) model performance without thresholding weights based on B1 data (i.e., no feature selection; Figure S11D), however as Figure S11A-C demonstrates, a BSR of 2 remains optimal for single and multiple predictor model performance. All peak neural activations from the initial PLS models (i.e., task SD_{BOLD}, resting-state SD_{BOLD}, task $MEAN_{BOLD}$) and corresponding coordinates are reported in Tables S2-S4 and Figures S2-S4. Unthresholded statistical maps are available at NeuroVault.org (https://identifiers.org/neurovault.collection:9030).

A final cross-validated model including only the univariate significant predictors of treatment outcome (i.e., task SD_{BOLD} (160 sec), resting-state SD_{BOLD} (340 sec), and second baseline LSAS-SR), improved the predictive accuracy beyond any single predictor ($r_{ACT,PRED}$=.77, MASE=.43, permuted P<0.001; see Figure 4B and Table 1). However, the unique variance associated with task-based SD_{BOLD} was notably higher than all other predictors (Figure 4C). Furthermore, the spatial patterns capturing neuro-behavioral correlations of task-based and resting-state SD_{BOLD} differed considerably (Figures 4F and S12), suggesting that the contribution of each SD_{BOLD} estimate to treatment outcome prediction is complementary both statistically (with regard to effect size and MASE values) and with regard to brain regions involved.

Neither depression nor insomnia severity at pre-treatment predicted social anxiety treatment outcomes (all permuted Ps>0.188). Further, a single principal component analysis (PCA) including all secondary social anxiety outcomes correlated strongly with the task SD_{BOLD}-predicted social anxiety change score noted in Figure 3B (R^2=34%, permuted P<0.001), but did not correlate with reductions in depressive or insomnia symptoms (Supplement, page 21).

Table 1. Univariate (zero-order) and multiple predictor models of treatment outcome across different within-patient data volumes. See Supplementary Material Table S7 for a complete presentation of model results across all data volumes.

				95% CI		
Pre-treatment predictors	**Data volume**	**MASE**	***r***	**Lower**	**Higher**	**Permuted *P***
Behavioral						
LSAS-SR at B1	48 items	0.65	.27	-.01	.56	0.071
LSAS-SR at B2	48 items	0.58	.45	.20	.70	<0.001*
Brain						
Task SD_{BOLD}	160 sec	0.54	.65	.51	.79	<0.001*
Rest SD_{BOLD}	160 sec	0.63	.19	-.08	.46	0.085
	340 sec	0.55	.55	.35	.75	<0.001*
Task $MEAN_{BOLD}$	160 sec	0.67	-.18	-.49	.12	0.843
Brain SD_{BOLD} and behavioral self-reports combined						
Multiple predictors**		0.43	.77	.66	.89	<0.001

*Significant zero-order prediction. **The multiple regression model includes all significant zero-order predictors (i.e., 160 sec task SD_{BOLD}, 340 sec resting-state SD_{BOLD}, and LSAS-SR at B2).

Abbreviations: MASE, Mean absolute scaled error; LSAS-SR, Liebowitz social anxiety scale, self-report version; B1, Baseline 1; B2, Baseline 2; SD_{BOLD}, neural variability; $MEAN_{BOLD}$, average neural response; Sec, Seconds; BOLD, Blood-oxygen-level-dependent imaging;

Eleven-week test-retest reliability

Eleven-week test-retest reliability (B1 *versus* B2) was excellent both for the primary social anxiety measure (LSAS-SR; $ICC_{B1,B2}$=0.84, CI=[0.75,0.90]) and our task-related SD_{BOLD} measure (reliability-based brain scores; $ICC_{B1,B2}$=0.80, CI=[0.70,0.90]; see Figure 5 and Tables S10-11). The task SD_{BOLD} ICC value was nearly identical after reducing data volume to 80 sec ($ICC_{B1,B2}$=0.78) and decreased slightly when reducing data volume to 40 sec ($ICC_{B1,B2}$=0.62). Further details and voxel-wise whole-brain calculations and plots are displayed in Table S11 and Figures S13-S15. In contrast to estimates of signal variability, task-related $MEAN_{BOLD}$ showed very poor reliability (all ICC's$_{B1,B2}$~0). Interestingly, control analyses revealed that this is not due to data quality issues *per se*, as $MEAN_{BOLD}$ response to faces replicated common topographical brain activity of face viewing (e.g., visual cortex, amygdala; Table S12 and Figure S16).

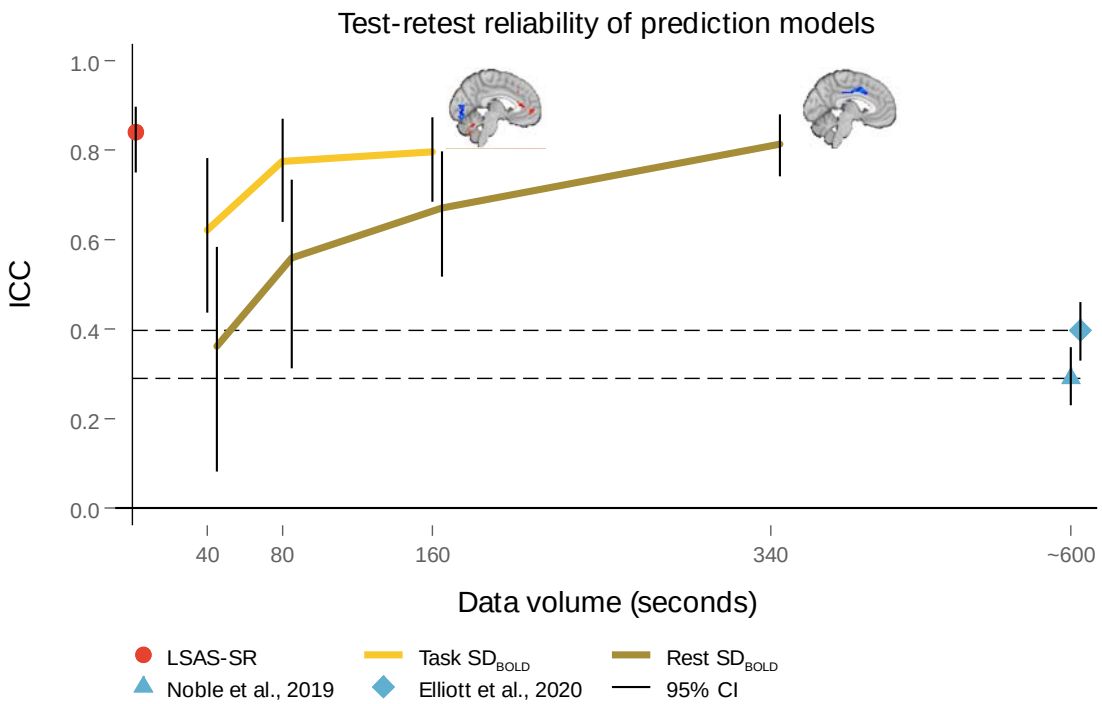

Figure 5. Test-retest reliability. ICCs were computed for LSAS-SR scores as well as fMRI BOLD across conditions (i.e., task SD_{BOLD} and resting-state SD_{BOLD}) and data volumes (i.e., 40, 80, 160 and 340 sec). Error bars represent bootstrapped, 95% confidence intervals. For reference, two meta-analyses on test-retest reliability using conventional analytics (i.e., functional connectivity and average brain signals) are presented. Noble et al., 2019 (ref 4) and Elliott et al., 2020 (ref 5) are meta-analyses on standard measures of task- and resting-state fMRI. **Abbreviations:** ICC, Intraclass correlation coefficient; SD_{BOLD}, neural variability; BOLD, blood-oxygen-level-dependent imaging; fMRI, functional magnetic resonance imaging; LSAS-SR, Liebowitz Social Anxiety Scale. Self-report version;

DISCUSSION

In the present study, we found that internet-delivered CBT successfully reduced SAD patients suffering and that pre-treatment brain signal variability was an accurate and reliable predictor of treatment outcome. A multiple predictor model that included task-based SD_{BOLD}, resting-state SD_{BOLD}, and pre-treatment social anxiety severity showed excellent prediction accuracy. Task-related SD_{BOLD} was the strongest predictor and exhibited excellent reliability. This relatively short (160 sec) estimate of task-based BOLD signal variability outperformed resting-state SD_{BOLD}, standard $MEAN_{BOLD}$, and pre-treatment self-reported social anxiety. Resting-state variability also uniquely predicted treatment outcome in our full model, but accounted for ~50% less unique explained variance than task-based SD_{BOLD} and required more than double the data volume to achieve comparable reliability and treatment outcome prediction accuracy.

Estimating brain signal variability during simple, disorder-relevant tasks may help optimize treatment prediction

Clinical neuroscientists often argue that resting-state neuroimaging protocols are preferable for ease of implementation and minimization of demands on patients. Here, superior treatment outcome prediction was achieved using a disorder-relevant task (socio-affective visual processing in SAD patients) with extremely low cognitive requirements (passive viewing, no behavioral responses required) and absolutely minimal scan time (2 minutes 40 sec, far shorter than typical resting-state scans). Further, the spatial patterns linking SD_{BOLD} to treatment outcome were largely distinct for task and resting-state SD_{BOLD}. Thus, while both task and resting-state variability contributed to CBT outcome prediction, the two measurements represent different neural signatures. If simple, demand-

minimal fMRI remains a primary goal for biomarker development in psychiatry, then passive, disorder-relevant tasks should be included in future large-scale studies of treatment outcome, particularly when BOLD fluctuations can be examined. Here, we employed a simple and straightforward calculation of each patient's brain signal variability, for which code is freely available and deployable for use in the majority of already collected patient fMRI data in the field (see https://github.com/LNDG/).

Moving beyond average neural signals for reliable treatment prediction

Our task-based SD_{BOLD} prediction model also dominated a more conventional analytic approach using mean brain signals ($MEAN_{BOLD}$; i.e., the average fMRI signal across time) to estimate treatment outcomes. Why might $MEAN_{BOLD}$ perform so poorly? Recently, alarming meta-analyses demonstrate that the average ICC may be as low as .40 (CI=[.33,.46]) for common experiments using conventional mean-based analyses in task-based fMRI (5). Crucially, the test-retest reliability of such standard (and alternative) fMRI measures in the treatment outcome prediction literature remains largely unknown. Our task-based SD_{BOLD} treatment prediction model demonstrated excellent 11-week test-retest reliability, and even with minimal data (40 sec), task SD_{BOLD} was far more reliable than $MEAN_{BOLD}$ when all available data (160 sec) were used. Note that we did replicate common topographical brain activity patterns of face viewing based on the conventional $MEAN_{BOLD}$ data (Table S12 and Figure S16); hence, the low reliability of the $MEAN_{BOLD}$ is unlikely to trace back to data quality *per se*. Beyond the poor performance of $MEAN_{BOLD}$ here, another meta-analysis also revealed very low reliability (ICC=.29; CI=[.23,.36]) for connectivity-based analyses of resting-state fMRI data (4). Utilizing brain measures with such poor reliability may continue to contribute to reduced probability of replication, and we argue that moment-to-moment brain signal variability computations are therefore strong candidates for future smaller- and larger-scale investigations of biomarkers in psychiatric research and treatment outcome prediction.

What could BOLD variability reveal about social anxiety treatment outcomes?

Researchers often conceive signal variability as unreliable and unwanted "noise." However, moment-to-moment brain signal variability continues to exhibit a host of behaviorally- and group-relevant effects in cognitive neuroscience (e.g., for a review see (8,31)), yet remains grossly underutilized in clinical research. To our knowledge, SAD has not previously been linked to BOLD variability. It has been argued that an individual's brain signal variability may (i) reflect available neural dynamic range for the more accurate processing of incoming stimuli (32) and (ii) index a more cognitively effective system overall (31). One characteristic SAD symptom is self-focused attention; as a result of an external socio-affective "trigger", SAD patients become self-attentive and biased towards internal cognitive and emotional processes, leading to deficits in the ability to disengage from internally-focused modes (13). As such, SAD patients may indeed "filter" external socio-affective stimuli through their own internal biases, showing a heightened focus on socio-affective content at the cost of an incomplete representation of such stimuli. CBT includes cognitive and behavioral interventions for dealing with excessive anxiety, such as shifting one's attention from internally referenced processing toward a more faithful representation of external input. Previous work has shown that lower signal variability in visual cortex should be expected when individuals *do not* fully process the complexity of visual input (32). It is plausible that treatment-responsive SAD patients express more limited visuo-cortical brain signal variability due to a relative inability to fully process external socio-affective stimuli, a function that may be directly improved via CBT. Complementarily, we also found that patients who displayed *higher* variability in the prefrontal cortex profited more from CBT. Past work consistently shows that greater BOLD variability in the frontal lobes typifies healthy, higher performing adults across a host of different cognitive domains, such as attentional capacity, working memory, and verbal abilities (6,9,11,12,33,34) (for a review see (31)). Accordingly, prefrontal signal variability may be required to respond to internet-delivered CBT, a treatment process that requires self-motivated learning, working memory, and verbal capacity. Although these interpretations remain speculative, novel questions related to intra-individual variability could be key for future directions in neuroscience-based psychiatric research. To best do so, future longitudinal studies are needed that mechanistically investigate the links between joint changes in neural variability and psychiatric treatment outcomes.

Limitations and future directions

We provide first evidence for within-sample reliability-based mapping of a series of different fMRI-based measures and experimental conditions in relation to treatment outcome. However, ultimately, the utility of any prediction model is determined by its ability to generalize to new, unseen patients. Poldrack and colleagues (3) recently criticized current prediction practices in the neuroimaging literature (e.g., 35–38), claiming that hundreds or even thousands of patients are needed for out of sample prediction. Even when such high data volume is available, prediction using conventional brain measures can work (*N*=1188) (2), but replication is not necessarily guaranteed in completely independent patient samples (39). We provide reliability-based evidence for the importance of brain signal variability in treatment prediction and the use of such reliable tools may markedly reduce the need for such massive, resource-intensive samples for treatment prediction.

It is also well-established that moment-to-moment brain signal variability is linked to a wide variety of state- and trait-related functions in different samples (9,31), rendering it possible that BOLD variability may be similarly sensitive to a variety of treatment outcomes in other common psychiatric disorders. However, it remains to be tested whether moment-to-moment neural variability estimated during other task types (e.g., memory and/or motor tasks; tasks that would not be judged *a priori* as "disorder relevant") also predicts treatment outcomes. Such future comparisons could then address whether variability-based prediction reflects clinically-relevant state- or trait-like neural signatures. Here, our socio-affective face-based task SD_{BOLD} treatment outcome prediction model was related to social anxiety, but not depression or insomnia, suggesting that the specific type of task might indeed determine the specificity of brain-based predictions.

Summary

Neural variability has the potential to offer unique insights into factors that affect patient responses to psychiatric treatments. Here, we demonstrate that intra-individual neural response variability is a reliable and accurate predictive biomarker of treatment success, even when using a simple passive task administered in under 3 minutes. Ultimately, our findings may help improve precision medicine and clinical decision-making in psychiatric populations.

Acknowledgements

This study was performed at the Centre for Functional Brain Imaging (UFBI), Umeå University and the University Hospital of Umeå, and we kindly thank Rebeca de Peredo Axelsson, Hans-Olov Karlsson, Kerstin Englund, Mikael Stiernstedt, Lars Nyberg and Carl-Johan Boraxbekk for support. Gerhard Andersson, Martin Kraepelien, Lise Bergman Nordgren, Erik Hedman-Lagerlöf, Samir El Alaoui and Jens Högström provided state-of-the-art clinical expertise. Cecilia Svanborg and Josef Isung contributed with clinical interviews. The MRI data collection was supported by Nils Hentati Isacsson and Örn Kolbeinsson - you are the best! Thanks also to the Max Planck Institute for Human Development, Lifespan Neural Dynamic Group team members for support in adapting and reviewing code: Steffen Wiegert, Marija Tochadse, Alexander Skowron. Last but not least, a warm thanks to all participants in this study!

We kindly thank the generous research grants K.M. and T.F. received from the Swedish Research Council (2018-06729, and 2016-02228), and the Swedish Brain Foundation (FO-2016-0106). D.G. and K.M. were supported partially by an Emmy Noether Programme grant from the German Research Foundation to D.G. and by the Max Planck UCL Centre for Computational Psychiatry and Ageing Research in Berlin.

Disclosures

The authors declare no conflict of interest.

REFERENCES

1. Forsell E, Isacsson N, Blom K, Jernelöv S, Ben Abdesslem F, Lindefors N, *et al.* (2020): Predicting treatment failure in regular care Internet-Delivered Cognitive Behavior Therapy for depression and anxiety using only weekly symptom measures. *J Consult Clin Psychol* 88: 311–321.
2. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, *et al.* (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23: 28–38.
3. Poldrack RA, Huckins G, Varoquaux G (2020): Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry* 77: 534–540.
4. Noble S, Scheinost D, Constable RT (2019): A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* 203: 116157.
5. Elliott ML, Knodt AR, Ireland D, Morris ML, Poulton R, Ramrakha S, *et al.* (2020): What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychol Sci* 31: 792–806.
6. Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2010): Blood oxygen level-dependent signal variability is more than just noise. *J Neurosci* 30: 4914–4921.
7. Garrett DD, McIntosh AR, Grady CL (2014): Brain signal variability is parametrically modifiable. *Cereb Cortex* 24: 2931–2940.
8. Waschke L, Kloosterman NA, Obleser J, Garrett DD (2021): Behavior needs neural variability. *Neuron* 109: 751–766.
9. Garrett DD, Nagel IE, Preuschhof C, Burzynska AZ, Marchner J, Wiegert S, *et al.* (2015): Amphetamine modulates brain signal variability and working memory in younger and older adults. *Proc Natl Acad Sci U S A* 112: 7593–7598.
10. Alavash M, Lim S-J, Thiel C, Sehm B, Deserno L, Obleser J (2018): Dopaminergic modulation of hemodynamic signal variability and the functional connectome during cognitive performance. *Neuroimage* 172: 341–356.
11. Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2011): The importance of being variable. *J Neurosci* 31: 4496–4503.
12. Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2013): The modulation of BOLD variability between cognitive states varies by age and processing speed. *Cereb Cortex* 23: 684–693.
13. Clark DM, Wells A (1995): A cognitive model of social phobia. In: R. Heimberg, M. Liebowitz, D. A. Hope, & F. R. Schneier, editor. *Social Phobia: Diagnosis, Assessment and Treatment*. New York: Guilford Press, pp 69–93.
14. Mayo-Wilson E, Dias S, Mavranezouli I, Kew K, Clark DM, Ades AE, Pilling S (2014): Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. *Lancet Psychiatry* 1: 368–376.
15. El Alaoui S, Hedman E, Kaldo V, Hesser H, Kraepelien M, Andersson E, *et al.* (2015): Effectiveness of Internet-based cognitive-behavior therapy for social anxiety disorder in clinical psychiatry. *J Consult Clin Psychol* 83: 902–914.
16. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, *et al.* (2009): Guided and unguided self-help for social anxiety disorder: randomised controlled trial. *Br J Psychiatry* 195: 440–447.
17. Andersson G, Carlbring P, Holmström A, Sparthan E, Furmark T, Nilsson-Ihrfelt E, *et al.* (2006): Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: a randomized controlled trial. *J Consult Clin Psychol* 74: 677–686.
18. Andersson G (2016): Internet-Delivered Psychological Treatments. *Annu Rev Clin Psychol* 12: 157–179.
19. Månsson KNT, Lindqvist D, Yang LL, Svanborg C, Isung J, Nilsonne G, *et al.* (2019): Improvement in indices of cellular protection after psychological treatment for social anxiety disorder. *Transl Psychiatry* 9: 340.
20. Månsson KNT, Salami A, Frick A, Carlbring P, Andersson G, Furmark T, Boraxbekk C-J (2016): Neuroplasticity in response to cognitive behavior therapy for social anxiety disorder. *Transl Psychiatry* 6: e727.
21. Baker SL, Heinrichs N, Kim H-J, Hofmann SG (2002): The Liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. *Behav Res Ther* 40: 701–715.
22. Zaider TI, Heimberg RG, Fresco DM, Schneier FR, Liebowitz MR (2003): Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychol Med* 33: 611–622.

23. Fischer H, Wright CI, Whalen PJ, McInerney SC, Shin LM, Rauch SL (2003): Brain habituation during repeated exposure to fearful and neutral faces: a functional MRI study. *Brain Res Bull* 59: 387–392.
24. Beckmann CF, Smith SM (2004): Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* 23: 137–152.
25. Garrett DD, Grady CL, Hasher L (2010): Everyday memory compensation: the impact of cognitive reserve, subjective memory, and stress. *Psychol Aging* 25: 74–83.
26. Hyndman RJ, Koehler AB (2006): Another look at measures of forecast accuracy. *Int J Forecast* 22: 679–688.
27. Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011): Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56: 455–475.
28. McIntosh AR, Bookstein FL, Haxby JV, Grady CL (1996): Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* 3: 143–157.
29. Putnick DL, Bornstein MH (2016): Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Dev Rev* 41: 71–90.
30. McGraw KO, Wong SP (1996): Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1: 30–46.
31. Garrett DD, Samanez-Larkin GR, MacDonald SWS, Lindenberger U, McIntosh AR, Grady CL (2013): Moment-to-moment brain signal variability: a next frontier in human brain mapping? *Neurosci Biobehav Rev* 37: 610–624.
32. Garrett DD, Epp SM, Kleemeyer M, Lindenberger U, Polk TA (2020): Higher performers upregulate brain signal variability in response to more feature-rich visual input. *Neuroimage* 217: 116836.
33. Guitart-Masip M, Salami A, Garrett D, Rieckmann A, Lindenberger U, Bäckman L (2016): BOLD Variability is Related to Dopaminergic Neurotransmission and Cognitive Aging. *Cereb Cortex* 26: 2074–2083.
34. Grady CL, Garrett DD (2018): Brain signal variability is modulated as a function of internal and external demand in younger and older adults. *Neuroimage* 169: 510–523.
35. Doehrmann O, Ghosh SS, Polli FE, Reynolds GO, Horn F, Keshavan A, *et al.* (2013): Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry* 70: 87–97.
36. Månsson KNT, Frick A, Boraxbekk C-J, Marquand AF, Williams SCR, Carlbring P, *et al.* (2015): Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl Psychiatry* 5: e530.
37. Frick A, Engman J, Alaie I, Björkstrand J, Gingnell M, Larsson E-M, *et al.* (2020): Neuroimaging, genetic, clinical, and demographic predictors of treatment response in patients with social anxiety disorder. *J Affect Disord* 261: 230–237.
38. Lueken U, Zierhut KC, Hahn T, Straube B, Kircher T, Reif A, *et al.* (2016): Neurobiological markers predicting treatment response in anxiety disorders: A systematic review and implications for clinical application. *Neurosci Biobehav Rev* 66: 143–162.
39. Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, *et al.* (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *Neuroimage Clin* 22: 101796.