Single-Subject Anxiety Treatment Outcome Prediction using Functional Neuroimaging

Tali M Ball*,[1,2], Murray B Stein[1,3,4], Holly J Ramsawh[5], Laura Campbell-Sills[1] and Martin P Paulus[1,3]

[1]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; [2]San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA; [3]Psychiatry Service, Veterans Affairs San Diego Healthcare System, San Diego, CA, USA; [4]Department of Family and Preventive Medicine, University of California San Diego, La Jolla, CA, USA; [5]Department of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

The possibility of individualized treatment prediction has profound implications for the development of personalized interventions for patients with anxiety disorders. Here we utilize random forest classification and pre-treatment functional magnetic resonance imaging (fMRI) data from individuals with generalized anxiety disorder (GAD) and panic disorder (PD) to generate individual subject treatment outcome predictions. Before cognitive behavioral therapy (CBT), 48 adults (25 GAD and 23 PD) reduced (via cognitive reappraisal) or maintained their emotional responses to negative images during fMRI scanning. CBT responder status was predicted using activations from 70 anatomically defined regions. The final random forest model included 10 predictors contributing most to classification accuracy. A similar analysis was conducted using the clinical and demographic variables. Activations in the hippocampus during maintenance and anterior insula, superior temporal, supramarginal, and superior frontal gyri during reappraisal were among the best predictors, with greater activation in responders than non-responders. The final fMRI-based model yielded 79% accuracy, with good sensitivity (0.86), specificity (0.68), and positive and negative likelihood ratios (2.73, 0.20). Clinical and demographic variables yielded poorer accuracy (69%), sensitivity (0.79), specificity (0.53), and likelihood ratios (1.67, 0.39). This is the first use of random forest models to predict treatment outcome from pre-treatment neuroimaging data in psychiatry. Together, random forest models and fMRI can provide single-subject predictions with good test characteristics. Moreover, activation patterns are consistent with the notion that greater activation in cortico-limbic circuitry predicts better CBT response in GAD and PD.

Neuropsychopharmacology (2014) **39,** 1254–1261; doi:10.1038/npp.2013.328; published online 8 January 2014

Keywords: anxiety disorders; prediction; fMRI; random forest; emotion regulation; cognitive behavioral therapy

INTRODUCTION

Anxiety disorders are highly prevalent (Kessler *et al*, 2005a) and significantly impairing (Sherbourne *et al*, 2010). Furthermore, anxiety disorders have increased in prevalence over the past several decades despite simultaneous increases in the availability of effective treatments (Kessler *et al*, 2005b). Although efficacious treatments exist, it is unclear who will respond best to what type of treatment. Thus, being able to match a patient to the best treatment for that individual is one of the most critical and important problems in psychiatry, yet few investigations have been conducted to gain a greater understanding of these individual differences.

One approach is to identify pre-treatment characteristics that predict treatment response. In addition to variables such as demographics and clinical characteristics, functional neuroimaging data may be particularly useful, because they are thought to quantify the underlying biological disease state. Previous neuroimaging studies of outcomes in psychiatric disorders have focused predominately on mood disorders and have identified the anterior cingulate as a potential predictor of antidepressant medication (Mayberg *et al*, 1997) or cognitive therapy (Siegle *et al*, 2012) response. In comparison, research in anxiety disorders has been limited to very small studies examining medication response and restricted to the amygdala and anterior cingulate (eg, Nitschke *et al*, 2009; Whalen *et al*, 2008). Results from these studies have been inconsistent and have not generated clinically useful single-subject predictions. There is therefore a profound need for research in anxiety disorders that (a) examines psychotherapy outcomes, (b) explores additional potential regions of interest (ROIs), and (c) moves towards clinically applicable single-subject prediction (as in Siegle *et al*, 2012).

In the present study, 48 patients with generalized anxiety disorder (GAD) or panic disorder (PD) were scanned while completing an emotion regulation task before cognitive

behavioral therapy (CBT). Individuals with anxiety disorders have been hypothesized to have difficulty with emotion regulation (Aldao et al, 2010) and have demonstrated hypo-activation in dorsolateral, dorsomedial, and ventrolateral prefrontal cortex (PFC) during emotion regulation (Ball et al, 2013; Goldin et al, 2009a; Goldin et al, 2009b). We therefore sought to examine emotion regulation-related brain activation as predictors of CBT outcome.

A number of techniques have been proposed for classifying complex biological data, including support vector machine and logistic regression. Among these, random forest (Breiman, 2001) is one of the most consistently robust predictive techniques, yielding superior performance in independent replication (Qi et al, 2006). The random forest technique consists of a complex partitioning of the predictor variable space and is summarized in Figure 1 (see also Breiman, 2001; Genuer et al, 2010; Strobl et al, 2009). It is particularly appropriate when the number of predictor variables is much larger than the number of subjects (Bureau et al, 2005), which is typically the case with neuroimaging data sets. Moreover, random forest has a low tendency to over-fit (Strobl et al, 2009), and the stepwise partitioning of the feature space can yield high-order interactions among many predictor variables that cannot be identified using other classification procedures (Lunetta et al, 2004). Random forest models have successfully been used to detect Alzheimer's disease (O'Bryant et al, 2010), identify tumor cells (Shi et al, 2004), and predict in-hospital mortality (Ward et al, 2006) but have not been used to generate psychiatric outcome predictions based on functional neuroimaging.

In the present study, we used random forests to predict CBT outcomes in GAD and PD. We expected that key limbic (amygdala, insula, anterior cingulate) and prefrontal (dorsolateral, dorsomedial, ventrolateral PFC) regions previously implicated in the etiology of anxiety disorders (Shin and Liberzon, 2010) would have a role. In addition to identifying brain areas contributing most to outcome prediction in this sample, we also examined the utility of random forest modeling as a means to predict therapy outcome in anxiety. To that end, we calculated test characteristics of sensitivity, specificity, and positive and negative likelihood ratios.

MATERIALS AND METHODS

Participants

The University of California San Diego Human Research Protections Program approved this study. After providing written informed consent, 107 participants were screened by semi-structured diagnostic interview (Sheehan et al, 1998). Of these, 53 were eligible, meeting DSM-IV criteria for clinically predominant PD ($n = 26$: 22 with agoraphobia) or GAD ($n = 27$), did not meet criteria for lifetime psychosis, past-year substance dependence, or past-month substance abuse, and participated in an fMRI session before CBT. Other anxiety disorders, including co-occurring PD and GAD, were permitted. Five participants were excluded from analysis due to poor fMRI data quality leaving 23 PD and 25 GAD. All were free of psychotropic medications for 6 weeks (2 weeks for benzodiazepines) and met safety and eligibility
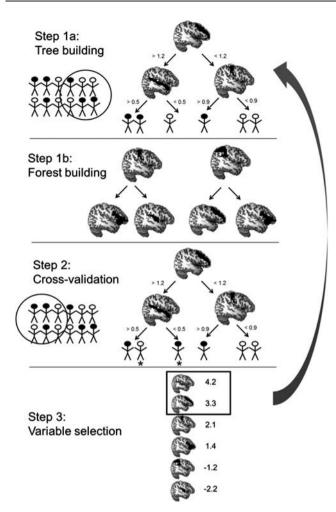

Figure 1 Random forest procedure. Step 1a is to build a decision tree based on a bootstrapped sample of participants (filled circles represent responders and open circles represent non-responders) and a random sample of predictor variables (eg, average activation in anatomically defined brain regions, arbitrary data shown). The random forest algorithm determines the optimal split point for each variable in order to correctly classify this subset of participants. Step 1b is to repeat this process hundreds or thousands of times to generate a forest of trees. In step 2, each tree classifies the participants that were not used in its original construction; each tree then 'votes' for the classification of these participants, and these votes are aggregated to provide the predicted status of each participant and thereby determine accuracy. Figures marked with an asterisk (*) indicate inaccurately classified participants in this example. Step 3 is the identification of the most important variables for prediction. Brain regions are ranked in terms of their variable importance scores: only those with greater importance than the absolute value of the most negative importance rating are selected for the final model (arbitrary data shown). The variables selected for inclusion are then used as the sole input variables for another iteration of steps 1 and 2, generating the final model.

criteria for fMRI. Data from 39 of the 48 participants were included in a previous report that did not include treatment data (Ball et al, 2013). Table 1 presents demographic and self-report measures.

Procedure

Eligible participants were screened with a medical examination consisting of a laboratory evaluation, EKG, and drug and pregnancy screen. Participants also completed

Table 1 Clinical and Demographic Measures

	All (n = 48)	Responders (n = 29)	Non-responders (n = 19)	Statistic
Gender (% female)	81%	90%	68%	$\chi^2(1) = 3.40$
Age in years	31.7 (9.7)	31.1 (9.7)	32.5 (10.1)	$t(46) = 0.47$
Education in years	15.0 (1.9)	14.9 (1.8)	15.1 (2.1)	$t(46) = 0.22$
Pre-treatment				
OASIS[a]	9.3 (3.8)	7.8 (2.9)	11.6 (3.9)	$t(46) = 3.83$[b]
ASI[a]	32.4 (11.3)	28.8 (9.5)	37.9 (11.7)	$t(46) = 2.95$[c]
PSWQ-A[a]	25.1 (3.3)	24.1 (3.0)	26.8 (3.2)	$t(46) = 2.74$[c]
QIDS	8.0 (4.4)	6.6 (4.1)	10.1 (4.2)	$t(46) = 2.84$[c]
Post-treatment				
OASIS	5.4 (3.6)	2.9 (1.5)	9.2 (2.4)	$t(46) = 11.19$[b]
ASI	16.5 (11.5)	11.4 (9.3)	24.4 (10.1)	$t(46) = 4.58$[b]
PSWQ-A	23.8 (2.9)	23.0 (2.4)	25.1 (3.1)	$t(46) = 2.60$[c]
QIDS	5.4 (4.5)	2.7 (2.1)	9.6 (4.1)	$t(45) = 7.68$[b]

Values presented as mean (SD). Statistic refers to comparison between responders and non-responders. QIDS at post-treatment was missing 1 data point (from a responder).
[a]Indicates variable was selected in the final clinical/demographic and combined models.
[b]Indicates $p < 0.004$ (Bonferroni corrected for 11 comparisons).
[c]Indicates $p < 0.05$.

self-report measures: the Overall Anxiety Severity and Impairment Scale (OASIS; Norman *et al*, 2006), an abbreviated version of the Penn State Worry Questionnaire (PSWQ-A; Hopko *et al*, 2003), the Anxiety Sensitivity Index (ASI; Reiss *et al*, 1986), the Intolerance of Uncertainty Scale (IUS; Buhr and Dugas, 2002), the Quick Inventory of Depressive Symptomology (QIDS; Rush *et al*, 2003), and the NEO Personality Inventory (Costa and McCrae, 1992). Finally, all participants completed an fMRI scan before 10 sessions of open-label weekly individual CBT (Craske *et al*, 2009) with a clinical psychologist or Masters-level therapist supervised by a clinical psychologist with CBT expertise.

The OASIS was the primary outcome measure due to its clinical relevance and applicability to both disorders. Responders were classified based on OASIS scores of ≤5 at the end of therapy (Roy-Byrne *et al*, 2010), with 60% of participants classified as responders (14 GAD, 15 PD). OASIS scores were collected every 2 weeks: participants who dropped out of treatment after completing at least four sessions ($n = 7$) were included in the analysis using their most recent score.

fMRI Task

Two processes were examined with the emotion regulation task: in each trial, individuals were asked to either 'Keep Up' (hereafter maintain) or 'Reduce' (hereafter reappraise) their emotional responses to negative images. Twelve negative images for each condition were selected from the International Affective Picture System (Lang *et al*, 2008). No neutral images were used, instead a pixel-wise scrambled version of each image was presented as a baseline at the

start of each trial. Participants were also prompted to rate their emotional state during each baseline and image presentation period. More details about this task have previously been described (Campbell-Sills *et al*, 2011).

Image Acquisition

One 9 min 40 s BOLD fMRI run was acquired, using a Signa EXCITE 3.0 Tesla-GE scanner (T2*-weighted echo planar imaging, TR = 2000 ms, TE = 32 ms, FOV = 240 × 240 mm, 64 × 64 matrix, 30 2.6-mm axial slices with a 1.4-mm gap, 290 scans). For anatomical reference, a high-resolution T1-weighted image (SPGR, TI = 450, TR = 8 ms, TE = 3 ms, FOV = 250 × 250 mm, flip angle = 12°, 172 sagittally acquired slices with 1-mm thickness) was obtained during the same session.

Image Processing

All structural and functional image processing was done with Analysis of Functional NeuroImages (AFNI) software (Cox, 1996). Time points with >2SDs more outlier voxels than the subject's mean were excluded from analysis, as determined by the AFNI function *3dToutcount*. Voxel time series were interpolated to correct for non-simultaneous slice acquisition and corrected for three-dimensional motion. Anatomical and functional volumes were co-registered algorithmically (Saad *et al*, 2009).

Individual participant time series data were analyzed with AFNI's *3dDeconvolve* program. Orthogonal regressors of interest modeled the maintain and reappraise conditions. Additional regressors of non-interest modeled the emotion rating periods, motion regressors, and linear and quadratic trends in the time series. Regressors were convolved with a modified gamma variate function to account for the hemodynamic response. Following deconvolution, data were converted to percentage of signal change by dividing the coefficient by the zero-order regressor within each voxel. Data were normalized to Talairach coordinates (Talairach and Tournoux, 1988) and subjected to 4 mm Gaussian spatial smoothing.

To obtain an easily replicable data set, we obtained average voxel-wise activation estimates for 70 anatomical ROIs during each of the two conditions (reappraise and maintain) to yield 140 independent variables. These ROIs were constructed as described in a previous report (Fonzo *et al*, 2012). Briefly, the Talairach atlas (Talairach and Tournoux, 1988) was combined with grey matter probabilities based on high-resolution T1 images from a group of 43 healthy adults. Grey matter probabilities were determined by applying grey matter segmentation with SPM5 (Statistical Parametric Mapping software; http://www.fil.ion.ucl.ac.uk/spm) for each subject, yielding voxel-wise probabilities of assignment to grey matter across all subjects. The probability maps were transformed to Talairach coordinates, and ROIs were defined using the Talairach atlas. Insula ROIs were divided into anterior and posterior at $y = 0$, and the medial frontal gyrus was divided into ventromedial and dorsomedial at $z = 14$. A total of 70 anatomical ROIs were created (Supplementary Table S1), and average activations were extracted for each subject, in each ROI, for each of the two task conditions, yielding 140 variables.

Statistical Analysis

Random forest classification was implemented in R statistics (http://cran.r-project.org; *randomForest* library). The goal of the procedure was to predict responder status using activations extracted from the anatomical ROIs during the two emotion regulation conditions. The random forest procedure involves several steps, as summarized in Figure 1 and described in detail elsewhere (Breiman, 2001; Genuer *et al*, 2010; Strobl *et al*, 2009). The first step is to construct a large number of classification trees (2000 trees were used, as recommended in Genuer *et al*, 2010), each using a bootstrapped subsample of participants and a randomly selected subset of the independent variables (Figure 1, Step 1). The number of independent variables used to create each classification tree was the square root of the total number of independent variables, as recommended by Díaz-Uriarte and De Andres (2006). At each node of the tree, a decision variable is selected randomly from the subset of available independent variables, and an optimal split point for that node is determined. The tree ends with each participant classified as a predicted responder or non-responder. In the second step, data from the out-of-bag sample (ie, for each classification tree, data from participants not used to create that tree) are used to evaluate how well the trees, in aggregate, classify subjects. Specifically, classification accuracy, sensitivity, specificity, and positive and negative likelihood ratios were evaluated (Figure 1, Step 2).

Third, data from the out-of-bag sample are used to evaluate variable importance, indicating how much each variable contributes to classification accuracy. Here, permutation importance scores were used to select the subset of best-performing variables for inclusion in a more parsimonious final model (Genuer *et al*, 2010). The permutation importance score of a predictor variable is defined as the decrease in overall accuracy when values of that variable are randomly permuted (Breiman, 2001). Variables that relate strongly to the outcome and contribute substantially to classification have large permutation importance, whereas those that do not contribute to classification have small or even negative scores (Strobl *et al*, 2009). Because negative scores are due to random variation around zero of the poor predictor variables (Strobl *et al*, 2009), only variables with an importance score greater than the magnitude of the most negative score were selected for inclusion (Figure 1, Step 3). Following Nicodemus *et al* (2010), the number of variables retained in the final model was limited to 10 and was based on median permutation importance scores from 500 repetitions of the random forest analysis to ensure stability of importance score estimates. The removal of poor performing variables can increase overall accuracy by increasing the selection of relevant variables for the decision trees and therefore increasing the relevance of included data to outcome prediction. A final, reduced model was built using the selected subset of predictor variables.

A similar analysis was undertaken using baseline clinical measures (OASIS, ASI, PSWQ-A, IUS, QIDS, NEO-Neuroticism, number of comorbid diagnoses) and demographics (age, gender, years of education) for comparison with the fMRI-based model. In addition, a third model was built using both brain imaging and clinical and demographic measures. For each of these, the best performing predictor variables were selected as described above.

Because the response rate was 60% (95% CI: 47–74%), the simple model that predicts all cases will respond to therapy will achieve 60% accuracy. Therefore, model accuracy for each random forest model was compared with the base response rate and its confidence interval. Models that can generate prediction accuracy outside the base rate confidence interval (ie, >74%) therefore perform statistically significantly better than predictions based only on the estimated responder rate and exceed the accuracy expected by chance.

To statistically compare the three models to each other, McNemar's test was used. In order to aid in the interpretation of the selected ROIs, correlations were performed between the ROIs in the final model and selected self-report measures (OASIS, ASI, PSWQ, and QIDS). Bonferroni correction ($\alpha = 0.005$) was used; however, uncorrected significance tests ($\alpha = 0.05$) are also reported for exploratory purposes.

We used all participants ($n = 48$) to create the random forest models for several reasons. First, this task yielded few differences at baseline between GAD and PD participants (Ball *et al*, 2013). Second, emotion regulation is considered a trans-diagnostic construct underlying affective disorders (Barlow *et al*, 2004), and the CBT protocol was designed such that it can be flexibly implemented across diagnoses (Craske *et al*, 2009). Finally, to best assess our ability to create predictive models using neuroimaging data, the increased power afforded by combining the groups was a distinct advantage. However, because the set of participants includes individuals with either primary GAD or PD, analyses were also conducted in each group separately. Individuals with comorbid PD and GAD were excluded from this supplemental analysis.

RESULTS

Table 1 shows baseline clinical and demographic measures. FMRI task effects, group effects, and brain–behavior relationships in the majority of these participants have been reported elsewhere (Ball *et al*, 2013). For each set of predictors, two models are described: an initial model built with all variables, and a final model built with the subset of predictors contributing most to classification accuracy. Results of each model in GAD and PD separately are in the Supplementary Results (Supplementary Table S2).

Clinical and Demographic Model

The initial model was comprised of baseline clinical and demographic variables as predictors of responder status and yielded accuracy of 63%. Three variables met criteria for inclusion in the final model: OASIS, ASI, and PSWQ-A, representing overall severity and impairment, anxiety sensitivity, and worry severity, respectively. The final model with these variables yielded accuracy of 69%, which is not statistically significantly different from the model based on response rate alone. Sensitivity was 0.79, and specificity was 0.53. The positive likelihood ratio of 1.67 (95% CI: 1.01, 2.79) and the negative likelihood ratio of

Table 2 Test Characteristics of All Three Final Models

Model	Accuracy	Sensitivity	Specificity	Positive LR (95% CI)	Negative LR (95% CI)
Clinical and demographic	69%	0.79	0.53	1.67 (1.01, 2.79)	0.39 (0.17, 0.90)
fMRI	79%	0.86	0.68	2.73 (1.39, 5.38)	0.20 (0.08, 0.53)
Combined	73%	0.83	0.58	1.97 (1.13, 3.42)	0.30 (0.12, 0.72)

CI, confidence interval; fMRI, functional magnetic resonance imaging; LR, likelihood ratio.

0.39 (95% CI: 0.17, 0.90) were statistically significantly different from each other and from 1.0 ($p < 0.05$).

fMRI Model

The initial model contained average reappraise- and maintain-related activation before treatment in each of the 70 anatomical ROIs. These 140 variables were entered as predictors of responder status yielding accuracy of 65%. Ten variables met criteria for inclusion in the final model: right hippocampus and left uncus activation during maintenance, as well as left transverse temporal gyrus, left anterior insula, right and left superior temporal gyrus, left supramarginal gyrus, left precentral gyrus, left superior frontal gyrus, and right substantia nigra activation during reappraisal. The final model with these variables yielded accuracy of 79%, which is statistically significantly greater than the model based on response rate alone ($p < 0.05$). Sensitivity was 0.86, and specificity was 0.68. The positive likelihood ratio of 2.73 (95% confidence interval: 1.39, 5.38) and negative likelihood ratio of 0.20 (95% confidence interval: 0.08, 0.53) were statistically significantly different from each other and from 1.0 ($p < 0.05$).

Combined Model

The initial model was comprised of all baseline clinical, demographic, and fMRI variables as predictors of responder status and yielded accuracy of 60%. Ten variables met the criteria for inclusion in the final model: OASIS, ASI, PSWQ-A, as well as right hippocampus and left uncus activation during maintenance, and left transverse temporal gyrus, left supramarginal gyrus, left precentral gyrus, left superior frontal gyrus, and right substantia nigra activation during reappraisal. The final model with these variables yielded accuracy of 73%, which is not statistically significantly different from the model based on response rate alone. Sensitivity was 0.83, and specificity was 0.58. The positive likelihood ratio of 1.97 (95% confidence interval: 1.13, 3.42) and negative likelihood ratio of 0.30 (95% confidence interval: 0.12, 0.72) were statistically significantly different from each other and from 1.0 ($p < 0.05$).

Model Comparisons

Table 2 shows the test characteristics of all three final models. Figure 2 illustrates the predictive information gained by each model over and above knowing the

treatment response rate. In addition, McNemar's chi-squared test was computed for all pairwise model comparisons. None of the comparisons were significant ($p > 0.2$) indicating that the models did not statistically significantly differ from each other.

Further Investigation of the fMRI Model

The precise contribution of each variable to the outcome prediction is complex, due to the high-order interactions critical to the success of random forests. However, main effects can be investigated straightforwardly. As shown in Figure 3, responders had greater activation than non-responders in all 10 ROIs comprising the final fMRI model. Baseline anxiety severity and impairment (OASIS) was inversely associated with activation in the left uncus during maintenance, as well as in the right superior temporal gyrus, left precentral gyrus, left superior frontal gyrus, and right substantia nigra during reappraisal ($p < 0.05$). None of the regions were significantly associated with baseline anxiety sensitivity (ASI). Baseline worry severity (PSWQ-A) was inversely associated with activation in the right hippocampus and left uncus during the maintain condition ($p < 0.05$). Finally, activation in the left transverse temporal gyrus and right superior temporal gyrus during reappraisal was associated with baseline depressive symptoms (QIDS; $p < 0.05$). However, the only associations significant with Bonferroni correction were with the left uncus during maintenance, such that greater activation was associated with less overall anxiety and worry severity.

DISCUSSION

To our knowledge, this is the first study to use random forest models and functional neuroimaging data to address a fundamental problem in psychiatry: predicting who will respond to a treatment. The results provide proof-of-principle that functional neuroimaging data can be used to generate predictions with good test characteristics. The likelihood ratios indicate that relative to treatment response odds, predicted responders based on the fMRI model are almost three times more likely and predicted non-responders five times less likely to respond to treatment. Although none of the model comparisons were statistically significant, the fMRI model demonstrated numerically higher classification accuracy than the model based on clinical and demographic variables and even had slightly greater accuracy than the combined model that used both types of data. Furthermore, only the fMRI model generated predictions that were significantly better than chance (ie, the model based on the response rate alone). Therefore, functional neuroimaging together with sophisticated classification procedures may ultimately make it possible to predict who will respond to what type of treatment and inform clinical decision-making.

We also identified potential predictors of CBT outcome in individuals with GAD and PD. Although these results are exploratory and require replication in an independent sample, they expand on previous work in anxiety disorders (eg, Nitschke et al, 2009; Whalen et al, 2008) by including more than one disorder group, by investigating psychother-
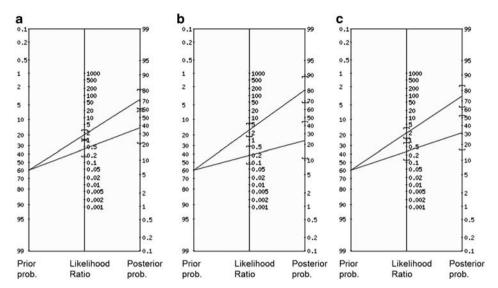
Figure 2 Model comparison of positive and negative likelihood ratios and posterior probabilities for (a) the clinical and demographic predictive model, (b) the fMRI predictive model, and (c) the combined model. Brackets indicate 95% confidence intervals. Upper lines indicate positive test result (ie, predicted responder) and lower lines indicate negative test result (ie, predicted non-responder).
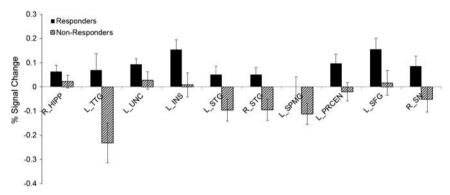

Figure 3 Average activation in responders and non-responder in regions selected for the final fMRI model. Error bars = SEM. Y axis is the perentage of signal change. Hippocampus and uncus activations were from the maintain condition, all other regions were from the reappraise condition. HIPP, hippocampus; INS, anterior insula; L, left; PRCEN, precentral gyrus; R, right; SFG, superior frontal gyrus; SN, substantia nigra; SPMG, supramarginal gyrus; STG, superior temporal gyrus; TTG, transverse temporal gyrus.

apy rather than medication, and by not constricting analyses to specific ROIs.

Consistent with our expectations, greater pre-treatment activation in cortico-limbic circuitry (ie, superior frontal gyrus, anterior insula) during emotion regulation was associated with CBT response. The superior frontal gyrus, particularly the dorsolateral PFC, has been implicated in top-down control and emotion regulation in healthy adults (Ochsner and Gross, 2005). However, individuals with anxiety disorders, including a large subsample of these GAD and PD patients (Ball et al, 2013), have demonstrated reduced engagement of this region during emotion regulation (Goldin et al, 2009a; Goldin et al, 2009b). The anterior insula is critical for interoception (Craig, 2009) and integration of emotional information (Lamm and Singer, 2010). Recent theoretical models have highlighted the role of the insula in maintaining problematic anxiety (Paulus and Stein, 2010). The identification of these regions as predictive by the random forest model provides face validity given their role in emotion regulation and anxiety pathology.

The uncus was also identified as a region contributing significantly to prediction accuracy. Furthermore, uncus activation was significantly correlated with worry and overall anxiety severity. Although the uncus has not been highlighted in previous studies of anxiety disorders, it has previously been associated with anxiety levels in response to emotional stimuli (Ewbank et al, 2009; Klumpp et al, 2011). Given the dearth of previous research examining neuroimaging predictors of therapy outcome in anxiety, the identification by the random forest model of non-hypothesized regions, including the uncus, underscores the benefit of an approach that is not constrained to known ROIs.

Perhaps the most studied region in anxiety is the amygdala, which has been robustly linked to the acquisition, experience, and expression of fear (Davis, 1992), and is hyperactive in anxiety disorders (Etkin and Wager, 2007). However, previous studies have not found amygdala activation to consistently predict anxiety treatment outcome (Nitschke et al, 2009; Whalen et al, 2008). The amygdala was not identified as a region contributing to treatment

response prediction in the present study. It is possible that although the amygdala's important role in the anxiety pathology its variability is less important for predicting who will respond to a particular treatment and who will not.

Limitations

One limitation of the present findings is the relatively small sample. Through the use of bootstrapping, random forest models generate reliable predictions even with small sample sizes and are not prone to over-fitting. However, though larger than previous investigations (eg, Nitschke *et al*, 2009; Whalen *et al*, 2008), the current sample may still be too small to statistically compare models with traditional hypothesis testing, and therefore strict conclusions about model comparisons are limited.

Another potential limitation is the combination of the two diagnostic groups. Whether mechanisms predicting treatment outcomes are consistent within anxiety disorders broadly or specific to each diagnosis remains an open question. Our sample included anxious adults with either clinically predominant GAD or PD. The rationale for this was partially theoretical, based on the trans-diagnostic approach of both the treatment (Craske *et al*, 2009) and the task used during brain imaging (Ball *et al*, 2013), and partially practical, to increase statistical power. We also report results from random forest models built using GAD or PD participants separately (Supplementary Material). These models also show strong test characteristics.

The present study examined response to a CBT protocol with established efficacy (Craske *et al*, 2009) and therefore did not include a waitlist or control group. The findings therefore may reflect a nonspecific tendency to improve with any intervention. Future research should utilize random forests in randomized clinical trials in order to establish whether the findings are specific to improvement with CBT. Examining differential treatment prediction will be especially important to assist clinicians in choosing among evidence-based treatments.

Finally, it is important to note that the regions identified in the present sample are exploratory and require replication. Although the use of bootstrapping in random forest is a significant strength, it does not remove the necessity for confirmatory testing of the identified regions in independent samples.

CONCLUSIONS

In conclusion, random forest models built with fMRI can provide single-subject predictions with good test characteristics. These preliminary data suggest that fMRI may have a role in predicting treatment outcomes. Future research should build on these findings by testing them in a larger sample and by continuing to use random forest methodology to develop predictive models with clinical relevance.

REFERENCES

Aldao A, Nolen-Hoeksema S, Schweizer S (2010). Emotion-regulation strategies across psychopathology: a meta-analytic review. *Clin Psychol Rev* **30**: 217–237.

Ball TM, Ramsawh HJ, Campbell-Sills L, Paulus MP, Stein MB (2013). Prefrontal dysfunction during emotion regulation in generalized anxiety and panic disorder. *Psychol Med* **43**: 1475–1486.

Barlow DH, Allen LB, Choate ML (2004). Toward a unified treatment for emotional disorders. *Behav Ther* **35**: 205–230.

Breiman L (2001). Random forests. *Machine Learn* **45**: 5–32.

Buhr K, Dugas MJ (2002). The Intolerance of Uncertainty Scale: psychometric properties of the English version. *Behav Res Ther* **40**: 931–945.

Bureau A, Dupuis J, Falls K, Lunetta KL, Hayward B, Keith TP *et al* (2005). Identifying SNPs predictive of phenotype using random forests. *Genet Epidemiol* **28**: 171–182.

Campbell-Sills L, Simmons AN, Lovero KL, Rochlin AA, Paulus MP, Stein MB (2011). Functioning of neural systems supporting emotion regulation in anxiety-prone individuals. *Neuroimage* **54**: 689–696.

Costa PT, McCrae RR (1992). Normal personality assessment in clinical practice: the NEO Personality Inventory. *Psychol Assess* **4**: 5–13.

Cox R (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**: 162–173.

Craig A (2009). How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* **10**: 59–70.

Craske MG, Rose RD, Lang A, Welch SS, Campbell-Sills L, Sullivan G *et al* (2009). Computer-assisted delivery of cognitive behavioral therapy for anxiety disorders in primary-care settings. *Depress Anxiety* **26**: 235–242.

Davis M (1992). The role of the amygdala in fear and anxiety. *Annu Rev Neurosci* **15**: 353–375.

Díaz-Uriarte R, De Andres SA (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* **7**: 3.

Etkin A, Wager T (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry* **164**: 1476–1488.

Ewbank MP, Lawrence AD, Passamonti L, Keane J, Peers PV, Calder AJ *et al* (2009). Anxiety predicts a differential neural response to attended and unattended facial signals of anger and fear. *Neuroimage* **44**: 1144–1141.

Fonzo GA, Flagan TM, Sullivan S, Allard CB, Grimes EM, Simmons AN *et al* (2012). Neural functional and structural correlates of childhood maltreatment in women with intimate-partner violence-related posttraumatic stress disorder. *Psychiatry Res Neuroimaging* **211**: 93–103.

Genuer R, Poggi JM, Tuleau-Malot C (2010). Variable selection using random forests. *Pattern Recognit Lett* **31**: 2225–2236.

Goldin PR, Manber T, Hakimi S, Canli T, Gross JJ (2009a). Neural bases of social anxiety disorder: emotional reactivity and

cognitive regulation during social and physical threat. *Arch Gen Psychiatry* 66: 170–180.

Goldin PR, Manber-Ball T, Werner K, Heimberg R, Gross JJ (2009b). Neural mechanisms of cognitive reappraisal of negative self-beliefs in social anxiety disorder. *Biol Psychiatry* 66: 1091–1099.

Hopko DR, Reas DL, Beck JG, Stanley MA, Wetherell JL, Novy DM *et al* (2003). Assessing worry in older adults: confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychol Assess* 15: 173–183.

Kessler RC, Chiu WT, Demler O, Walters EE (2005a). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 62: 617–627.

Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE *et al* (2005b). Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med* 352: 2515–2523.

Klumpp H, Ho SS, Taylor SF, Phan KL, Abelson JL, Liberzon I *et al* (2011). Trait anxiety modulates anterior cingulate activation to threat interference. *Depression and Anxiety* 28: 194–201.

Lamm C, Singer T (2010). The role of anterior insular cortex in social emotions. *Brain Struct Funct* 214: 579–591.

Lang PJ, Bradley MM, Cuthbert BN (2008). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual, Technical Report A-8*. University of Florida: Gainesville, FL, USA.

Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P (2004). Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet* 5: 32.

Mayberg HS, Brannan SK, Mahurin RK, Jerabek PA, Brickman JS, Tekell JL *et al* (1997). Cingulate function in depression: a potential predictor of treatment response. *Neuroreport* 8: 1057–1061.

Nicodemus KK, Callicott JH, Higier RG, Luna A, Nixon DC, Lipska BK *et al* (2010). Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Hum Genet* 127: 441–452.

Nitschke JB, Sarinopoulos I, Oathes DJ, Johnstone T, Whalen PJ, Davidson RJ *et al* (2009). Anticipatory activation in the amygdala and anterior cingulate in generalized anxiety disorder and prediction of treatment response. *Am J Psychiatry* 166: 302–310.

Norman SB, Cissell SH, Means-Christensen AJ, Stein MB (2006). Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). *Depress Anxiety* 23: 245–249.

O'Bryant SE, Xiao G, Barber R, Reisch J, Doody R, Fairchild T *et al* (2010). A serum protein-based algorithm for the detection of Alzheimer disease. *Arch Neurol* 67: 1077–1081.

Ochsner KN, Gross JJ (2005). The cognitive control of emotion. *Trends Cogn Sci* 9: 242–249.

Paulus MP, Stein MB (2010). Interoception in anxiety and depression. *Brain Struct Funct* 214: 451–463.

Qi Y, Bar-Joseph Z, Klein-Seetharaman J (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins Struct Funct, Bioinformatics* 63: 490–500.

Reiss S, Peterson RA, Gursky DM, McNally RJ (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behav Res Ther* 24: 1–8.

Roy-Byrne P, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ *et al* (2010). Delivery of evidence-based treatment for multiple anxiety disorders in primary care. *J Am Med Assoc* 303: 1921–1928.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN *et al* (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 54: 573–583.

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW (2009). A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage* 44: 839–848.

Sheehan D, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E *et al* (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 59: 22–33.

Sherbourne CD, Sullivan G, Craske MG, Roy-Byrne P, Golinelli D, Rose RD *et al* (2010). Functioning and disability levels in primary care outpatients with one or more anxiety disorders. *Psychol Med* 40: 2059–2068.

Shi T, Seligson D, Belldegrun AS, Palotie A, Horvath S (2004). Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Mod Pathol* 18: 547–557.

Shin LM, Liberzon I (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology* 35: 169–191.

Siegle GJ, Thompson WK, Collier A, Berman SR, Feldmiller J, Thase ME *et al* (2012). Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. *Arch Gen Psychiatry* 69: 913–924.

Strobl C, Malley J, Tutz G (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14: 323–348.

Talairach J, Tournoux P (1988). *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Thieme.

Ward MM, Pajevic S, Dreyfuss J, Malley JD (2006). Short-term prediction of mortality in patients with systemic lupus erythematosus: classification of outcomes using random forests. *Arthritis Care Res* 55: 74–80.

Whalen PJ, Johnstone T, Somerville LH, Nitschke JB, Polis S, Alexander AL *et al* (2008). A functional magnetic resonance imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. *Biol Psychiatry* 63: 858–863.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)