

Archival Report

Neurocognitive Measures of Self-blame and Risk Prediction Models of Recurrence in Major Depressive Disorder

Andrew J. Lawrence, Daniel Stahl, Suqian Duan, Diede Fennema, Tanja Jaeckle, Allan H. Young, Paola Dazzan, Jorge Moll, and Roland Zahn

ABSTRACT

BACKGROUND: Overgeneralized self-blaming emotions, such as self-disgust, are core symptoms of major depressive disorder and prompt specific actions (i.e., action tendencies), which are more functionally relevant than the emotions themselves. We have recently shown, using a novel cognitive task, that when feeling self-blaming emotions, maladaptive action tendencies (feeling like hiding and feeling like creating a distance from oneself) and an overgeneralized perception of control are characteristic of major depressive disorder, even after remission of symptoms. Here, we probed the potential of this cognitive signature, and its combination with previously employed functional magnetic resonance imaging (fMRI) measures, to predict individual recurrence risk. For this purpose, we developed a user-friendly hybrid machine/statistical learning tool, which we make freely available.

METHODS: A total of 52 medication-free patients with remitted major depressive disorder, who had completed the action tendencies task and our self-blame fMRI task at baseline, were followed up clinically over 14 months to determine recurrence. Prospective prediction models included baseline maladaptive self-blame-related action tendencies and anterior temporal fMRI connectivity patterns across a set of frontolimbic a priori regions of interest, as well as including established clinical and standard psychological predictors. Prediction models used elastic net regularized logistic regression with nested 10-fold cross-validation.

RESULTS: Cross-validated discrimination was highly promising (area under the receiver-operating characteristic curve ≥ 0.86), and positive predictive values over 80% were achieved when including fMRI in multimodal models, but only up to 71% (area under the receiver-operating characteristic curve ≤ 0.74) when solely relying on cognitive and clinical measures.

CONCLUSIONS: This study shows the high potential of multimodal signatures of self-blaming biases to predict recurrence risk at an individual level and calls for external validation in an independent sample.

<https://doi.org/10.1016/j.bpsc.2021.06.010>

One central feature of cognitive models of vulnerability to major depressive disorder (MDD) is a tendency to excessively blame oneself (1). These models have pointed to overgeneralized self-blame as causing worthlessness and hopelessness, core symptoms of MDD that are distinctive compared with other recurrent emotional disorders such as panic disorder (2). Self-blame in MDD is strongly associated with complex emotions such as guilt, self-disgust, and shame (3). Social psychologists have highlighted that specific maladaptive action tendencies such as feeling like hiding were associated with self-blaming emotions (4) and distinguish adaptive from depressogenic self-blame more accurately than the label of the emotion (5). Indeed, using a novel action tendencies task, we have observed that self-blame-related action tendencies, in particular feeling like creating a distance from oneself and hiding, as well as an overgeneralized perception of control for other people's wrongdoing, were more frequent in patients with remitted MDD compared with control participants

cross-sectionally (6). In this comparison, feeling like hiding was also more common in patients with MDD when their friend was described to act badly toward them, which could be interpreted in different ways but may reflect an overgeneralized sense of responsibility (7). If measures of maladaptive self-blame-related action tendencies prospectively predicted future recurrence risk in remitted MDD, this would provide critical evidence for their role in MDD vulnerability. Apart from its pathophysiological importance, the identification of novel predictors of recurrence is needed for developing accurate risk prediction tools, because clinical variables are poor predictors at an individual level (8) and because there are no established biomarkers (9).

Over the last decade, the application of machine learning to magnetic resonance imaging (MRI)-based prediction of individual responses to antidepressant treatments (10–13) has raised great promise for developing predictive markers in MDD. Machine learning models are powerful at making

personalized predictions because they condense multiple variables, such as MRI signal in different brain regions, or different modalities (e.g., cognitive, structural MRI and functional MRI [fMRI]) (11) into one model used to classify individuals based on complex interactions between the multiple input variables (14). This approach has been used to predict whose symptoms will respond to a treatment (10,11,15,16). Successful development of novel treatments and stratification algorithms to improve long-term outcomes of recurrent MDD, however, requires predictors that capture the neurocognitive underpinnings of asymptomatic precursors of recurrent symptoms. Using a standard general linear model, we previously identified an fMRI signature of patients with MDD who will develop a recurring episode after recovering from previous depression over 1 year of follow-up (17), but the machine learning approach we used to show its predictive value at the individual level with a positive predictive value of 74% had two main weaknesses. First, it used a high variance leave-one-out cross-validation method (18,19). Second, it used the extracted clusters showing the most significant association with recurrence risk in our SPM general linear model. This technique overestimates model performance owing to selection of voxels and variables on the basis of another model comparing recurring episode and stable remission patients in the same sample (20). Here, we sought to overcome these two weaknesses.

As recently reviewed (21), guilt proneness has been previously selectively associated with subgenual cingulate cortex (SCC) activation on fMRI (22), whereas frontopolar activation was most reproducible for guilt versus other-blaming emotions irrespective of individual differences in guilt proneness (23–25). The latter is part of the default mode network, and its connectivity with the SCC has been implied in depressive rumination (26), which is typically of a self-critical nature. Interestingly, the SCC was abnormally active in current MDD, and normalization of its metabolism was associated with remission (27,28), and its connectivity predicted response to psychological versus antidepressant treatment (15), suggesting its central importance. The SCC, however, is only part of a network of regions involved in self-blaming feelings. Using fMRI, we demonstrated that proneness to self-blaming feelings in healthy people was associated with increased functional connectivity between the SCC and the right superior anterior temporal lobe (ATL) (29), which we had previously demonstrated to enable differentiated interpretations of the meaning of social behavior (e.g., differentiating actions as impolite or absent-minded, rather than just overgeneralizing them as “bad”) (30,31).

In an independent previous sample of patients with remitted MDD, we demonstrated abnormal functional connectivity between the right superior ATL and frontolimbic networks when experiencing self-blaming feelings relative to feelings related to blaming others (compared with never-depressed control participants independent of medication status) (25). These frontolimbic networks included the SCC, frontopolar cortex, hypothalamus, and hippocampus. This concurred with Price and Drevets’ (27) neuroanatomical model of mood disorders, which includes the superior ATL because of its close anatomical connection with the medial prefrontal network including the subgenual cortex and the hypothalamus, thereby

providing a crucial and often forgotten link between the limbic forebrain (including the hippocampus) and the frontal cortex. The ATL is probably underreported because fMRI in this region requires optimized imaging parameters owing to the abundance of artifacts (32). It is notable that the original formulations of the default mode network based on positron emission tomography included the superior ATL (33). Our focus on the right superior ATL is based on its predominance for socio-emotional functions (23,31,34–37) relative to the left ATL. This concurs with the efficacy of right temporal electroconvulsive therapy in MDD (38), response to which was predicted by right ATL–frontal connectivity including the SCC (39).

The aim of the current study was to develop a user-friendly prediction tool and use it to provide robust estimates for the potential to predict MDD recurrence risk at the individual level, when using cognitive and fMRI-based signatures of self-blaming biases. Further, as one could argue that fMRI is relatively expensive for widespread use, we also probed the potential to replace it by our novel cognitive task for future clinical decision support systems.

METHODS AND MATERIALS

Participants and Standard Measures

Participants were recruited as a part of a prospective cohort study (from 2011 to 2014) in a clinical research facility (ethics reference: 07/H1003/194) (17). All participants provided informed consent (verbal for initial telephone screening and written otherwise). Recruitment has been described previously (40). Inclusion criteria were MDD according to the DSM-IV-TR (41), remitted for ≥ 6 months, psychotropic medication-free, right-handed, native English speaking, and normal or corrected-to-normal vision. Main exclusion criteria were current Axis I disorders including a history of substance or alcohol abuse, and past comorbid Axis I disorders being the likely cause of depressive symptoms. Patients were subsequently followed up clinically over 14 months using the well-validated Longitudinal Interval Follow-up Evaluation for DSM-IV (42). Raters were blinded to the baseline results; interrater reliability was excellent (17).

For our multimodal prediction model, we used a complete-cases analysis (see [Supplemental Methods](#)) including 52 patients with MDD with complete follow-up data that we were able to categorize into recurring episode over 14 months ($n = 18$) versus stable remission with no recurring episode ($n = 34$).

The following standard clinical and psychological measures were selected for use in the prediction model as based on previous reports establishing them as reproducible predictors of MDD recurrence at the group level: the number of previous episodes (categorized into nonrecurrent [i.e., 1 previous episode], recurrent [2–4 episodes], and highly recurrent [>5 episodes]) (43); the Beck Depression Inventory (44) as a measure of residual symptoms that are associated with recurrence risk (43) and that loads heavily on self-criticism, also associated with recurrence (45); the Global Assessment of Functioning Scale to capture comorbidity and functional impact (46); the Rosenberg Self-Esteem Scale (47,48); and the Positive and Negative Affect Schedule asking how people feel in general (49), in which the negative affect subscale was used as a measure of trait negative emotionality shown to be

Self-blame and Depression Risk Prediction

associated with poor prognosis (50) and the positive affect subscale was used because of its negative association with subsequent recurrence (51) in previous studies.

Experimental Stimuli

Stimuli for both fMRI and behavioral tasks were written sentences each presenting an abstract hypothetical social behavior contrary to social and moral values. Participants were asked to imagine the situation in each stimulus for two conditions differing by the agency of the participant. In the self-agency condition, the participant was described to act toward their friend, and in the other-agency condition, their friend acted toward them. Stimuli were based on the value-related moral sentiment task (VMST) used previously (17,40). The same social behaviors were used in both conditions (90 trials per condition, 50% per se negative social behaviors [e.g., “does act stingily”] and 50% negated positive behaviors [e.g., “does not act generously”]). Participants were asked to provide the name of their best friend of the same gender, with whom they were not related and not romantically involved.

After the fMRI and behavioral tasks, participants were asked to rate all items in the self- and other-agency conditions for unpleasantness (“How strongly would you feel unpleasant feelings?": scale of 1 [not unpleasant] to 7 [extremely unpleasant]). As in our previous article (17), we restricted our analyses to the items deemed most emotionally relevant to each participant (defined as items rated equal or higher than the individual’s median unpleasantness rating for each condition: self- and other-agency).

fMRI Acquisition and Paradigm

As previously described (Figure S1) (17), we used an fMRI protocol optimized for detection of ventral brain regions (see Supplemental Methods), and standard fMRI results were previously reported (17). Participants were presented with hypothetical social actions in self-agency and other-agency conditions (stimuli described above). Stimuli were presented for 5 seconds in three runs in pseudorandom order; runs were counterbalanced across participants and were interspersed with a baseline visual fixation of pattern condition ($n = 90$). In the scanner, participants were asked to decide whether each situation would feel mildly or very unpleasant to ensure that they paid attention to the task and that they made an emotional decision about the stimuli. We used a jittered (range = 500 ms) intertrial interval with mean duration of 4 seconds.

fMRI Analysis

As in our previous analysis (Supplemental Methods) (17) to measure functional connectivity, we employed the well-established psychophysiological interaction analysis (52), which requires the extraction of the right superior ATL signal time course [physiological variable, as done previously (29); 6-mm-radius sphere; Montreal Neurological Institute coordinates: $x = 58$, $y = 0$, $z = -12$] and the creation of an interaction term with the psychological variable (the contrast between the most highly unpleasant items in the self-agency condition vs. the visual fixation condition and highly unpleasant items in the other-agency condition vs. the visual fixation condition).

Regression coefficient averages over a priori regions of interest (ROIs) (depicted in Figure 1B and further described below) as defined previously using an independent sample (25) were extracted as raw values using MarsBaR (53) and used to capture self-blame-related fMRI connectivity with our ATL seed region, as well as to capture ATL connectivity with these ROIs irrespective of psychological condition, while covarying for root mean square movement parameters during the scan (obtained from the realignment process). In addition, we modeled standard blood oxygen level-dependent effects following the modeling approach as previously described (17) for self-blame in the SCC (28) and ATL seed region. This was in keeping with the psychophysiological interaction approach (52). To limit the number of variables, we modeled blood oxygen level-dependent effects only in our primary ROIs (SCC and ATL).

fMRI ROIs

To ensure independence of our ROIs from the current dataset, we used all a priori ROIs from our previous independent publication (25), but in order to reduce the number of ROIs, we chose only right-sided ROIs reflecting our right-sided ATL seed region, except for the midline regions, in which we used bilateral ROIs (see Supplemental Methods and Table S2).

Action Tendencies Task

As previously described (6), for each described social behavior, the participant was asked, “What would you feel like doing in response to this behavior?” “Please select the option that you feel that you would most strongly feel like doing” (see Figure S1): “verbally or physically attacking/punishing your best friend,” “verbally or physically attacking/punishing yourself,” “apologising/fixing what you have done,” “hiding,” “creating a distance from your best friend,” “creating a distance from yourself,” “no action,” or “other action.” They then rated the hypothetical situation regarding how much control they feel they would have (responses ranging from 1 [none at all] to 7 [complete]).

We have previously reported that self-agency-related hiding and self-distancing, as well as an overgeneralized perception of control for others’ wrongdoing, were characteristic of MDD at baseline compared with control participants (6). In the present article, we report their association with recurrence risk for the first time. As done previously (6), we separately calculated the proportion of trials in which hiding or self-distancing were selected in the self-agency condition for the highly unpleasant items as defined using individual median splits as described above. In order to reduce the number of variables, we z-transformed these proportions across the whole sample to combine hiding and self-distancing into a single average of z scores of maladaptive self-blaming action tendencies. We measured overgeneralized perception of control as done previously by subtracting each subject’s average perceived control rating in the highly unpleasant other-agency condition trials from the same quantity in their highly unpleasant self-agency trials. Perceived control is expected to be higher during self-agency, so a smaller value in the calculated difference measure indicates abnormally higher perception of control in the other-agency condition, which is conceptually related to the

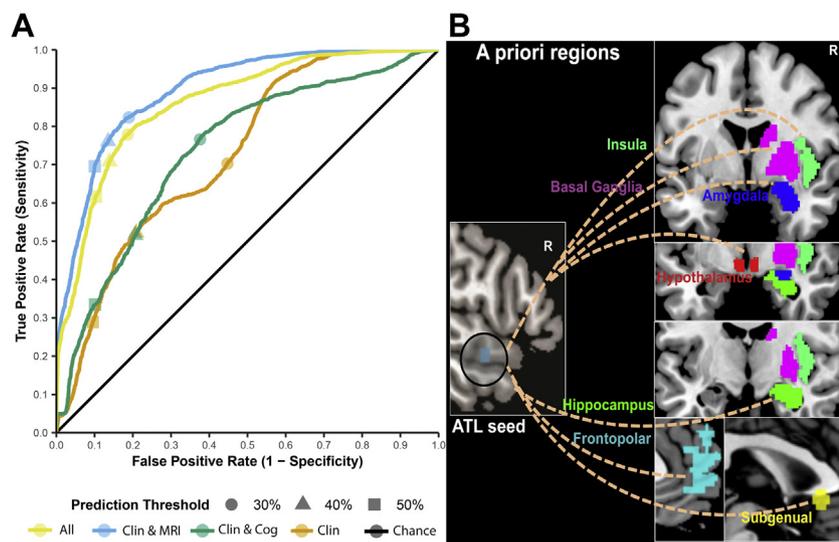


Figure 1. (A) Average cross-validated receiver operating characteristic curves for repeated nested regularized logistic regression models (the same 52 participants with major depressive disorder in all models, 30 nested inner and 100 outer loop repetitions using 10-fold cross-validation, R software). The clinical variables (Clin) model included standard scales and known clinical predictors. For the clinical and cognitive (Clin & Cog) model, maladaptive self-blaming action tendencies measures were added. For the clinical and functional magnetic resonance imaging (fMRI) (Clin & MRI) model, fMRI measures were added. The all modalities (All) model included cognitive, clinical, and fMRI measures. (B) All a priori regions of interest reported in a previous independent article (25) with close anatomical connections to our anterior temporal lobe (ATL) seed region and of sufficient anatomical specificity for data extraction were included (i.e., regions spanning multiple Brodmann areas such as the dorsolateral frontal cortex were discarded). The model included the midline subgenual cingulate cortex, frontopolar cortex (Brodmann area 10), and hypothalamus, as well as the right hemispheric hippocampus, amygdala, insula, and basal ganglia (striatum/pallidum) regions of interest to match our a priori right hemisphere ATL seed region (25). R, right.

notion of omnipotent responsibility guilt (i.e., caused by an exaggerated sense of responsibility for others' well-being) as shown to be associated with MDD (7).

Prediction Modeling

We developed a software tool (AL) to implement the proposed methods and make this tool available for use by the scientific and clinical community as an R package (<https://www.github.com/AndrewLawrence/dCVnet>). We adopted a hybrid approach combining statistical and machine learning to balance their benefits and limitations in this type of prediction problem. In contrast to traditional statistical methods, we reduced overfitting and allowed correlated predictors through regularization. We also carefully cross-validated performance by including all modeling decisions within the cross-validation. This hybrid approach can likewise be contrasted to more complex (and data-hungry) (54) machine learning methods: each model uses relatively few of the available predictors, which are chosen a priori based on previously published results and clinical utility. To further reduce complexity, we used a model that assumes that the effects of the predictors are linear and additive, and any interaction terms or predictor transformations must be prespecified. The software tool is called dCVnet, for “double cross-validation for the elastic net.” It employs elastic net-regularized logistic regression with double (also termed “nested”) cross-validation, consisting of an outer cross-validation of model performance measures with an inner cross-validation to independently tune elastic net hyperparameters.

Elastic Net Model. Elastic net-regularized binary logistic regression extends logistic regression with regularization. This acts as a penalty to model complexity and reduces this by shrinking coefficients toward zero (55). The elastic net regularization penalty comprises two types of penalty with different roles: 1) the L2 (Ridge) penalty allows correlated predictors to

jointly enter the model and stabilizes solutions and 2) the L1 (LASSO) penalty encourages variable selection (56). The amount of regularization (hyperparameter: lambda) and balance of the two types of regularization penalty (hyperparameter: alpha; indicating the fraction of the penalty that is of LASSO type) are hyperparameters of the elastic net procedure in that they determine the learning performance of the algorithm, and tuning alpha and lambda adapts the algorithm to different problem settings.

Cross-Validation and Tuning. We implemented double cross-validation in dCVnet to obtain performance estimates without optimistic bias at the same time as tuning hyperparameters based on cross-validated performance (57). For both inner and outer cross-validation loops, our tool employed repeated k-fold cross-validation (Figure S4), an unbiased estimate of expected (out-of-sample) prediction error with lower variance than bootstrap, hold-out, and leave-one-out cross-validation methods (56,58).

For this particular application of dCVnet, stable hyperparameter selection could be obtained (for all models) with 30 repetitions of 10-fold cross-validation, and stable cross-validation results in 100 repetitions of 10-fold cross-validation. For tuning, six logarithmically spaced values of alpha were considered between 0.01 (mostly Ridge) and 1.0 (a LASSO model). For each alpha, 100 lambda values were determined from the data by glmnet (59). The tuned alpha and lambda were selected based on the minimum mean square error.

Software. dCVnet was programmed in R software version 3.6.2 (R Foundation for Statistical Computing) and makes repeated calls to the glmnet 3.0.2 package's binomial elastic net fitting function to produce the multiple models required for nested k-fold cross-validation (59). The novelty of dCVnet lies in providing a documented and standardized implementation

Self-blame and Depression Risk Prediction

of this particular machine learning pipeline, making it accessible to researchers lacking programming experience required for more general machine learning software environments (e.g., sklearn, caret, tidymodels, or mlr). We make this software including documentation available (<https://www.github.com/andrewlawrence/dCVnet>) to improve reproducibility and future open source development of clinical decision support systems. We ran permutation tests to rule out biases and information leakage within dCVnet, which was confirmed by a distribution of the balanced accuracy measures from the permuted datasets around the expected mean of 50% (i.e., chance) (Figure S2, Table S1, and Supplemental Methods).

Model Evaluation. Predicted probabilities of depression recurrence were compared with the known recurrence status to calculate a variety of cross-validated prediction performance measures (60). These included common classification metrics (such as positive predictive value and negative predictive value) based on a 50% classification cutoff. For positive predictive value and negative predictive value, we adjusted our estimates for the 42% prevalence of recurrence in the wider study sample. We further considered threshold-independent performance measures: the Brier score (mean squared difference between predicted probability and binary outcome) and the area under the receiver-operating characteristic curve (AUC) (also termed the concordance statistic; see Supplemental Methods). Finally, we considered model calibration (i.e., the fidelity of the predicted probability to the observed proportion with the outcome) by calculating the intercept and slope of the calibration graph (60). Consistent with our aims, at this point in model development, the measure of prime interest is the AUC, as this reflects the discriminative potential of a model independent of the classification threshold, or model calibration.

RESULTS

As shown in Table 1, there were no demographic differences between groups. As expected, the recurring episode group showed a higher Beck Depression Inventory score at baseline. We obtained an AUC of at least 0.86, $\geq 80\%$ cross-validated

accuracy, and moderate calibration for classifying patients into recurring episode versus stable remission when combining multimodal information from fMRI and standard clinical and psychological measures, with or without our novel action tendencies task (Table 2, Figure 1). We further probed the contributions of the different modalities to predictive performance. Relying solely on standard clinical and psychological measures achieved an AUC of 0.73 and a positive predictive value of 68% (Table 2). Our results showed, however, that the novel cognitive measure, our action tendencies task, did not improve prediction performance to a relevant degree (AUC = 0.74; positive predictive value of 71%) (Table 2). We therefore also ran a prediction model that only contained the action tendencies task variables to demonstrate that they were associated with recurrence risk (AUC = 0.67) (Table 2), but this model performed slightly worse compared with known clinical and standard measures. Likewise, when using fMRI without standard measures, there was an association with recurrence risk (AUC = 0.69), but this was not as strong as in the multimodal model.

When comparing unregularized odds ratios from single-predictor logistic regressions with the regularized odds ratios, there was relatively little shrinkage of coefficients (i.e., a small lambda was chosen), reflecting the relatively good cross-validated performance of the model (Table 3). Interestingly, there were predictors that on their own had no relevant association with recurrence risk but contributed to the multivariate prediction. Maladaptive action tendencies were associated with recurrence risk by themselves and in the full regularized model (Table 3), whereas overgeneralized perception of control was not (Table 3).

DISCUSSION

We confirmed that combinations of clinical, psychological, and self-blame fMRI measures were highly promising candidates for predicting individual recurrence risk in MDD. Using a principled hybrid statistical/machine learning approach, we observed excellent cross-validated discriminative performance, with an AUC of 0.9 and positive predictive values above the suggested benchmark of 80% for clinically useful

Table 1. Demographic Characteristics and Symptoms

Variable	Remitted MDD		P Overall
	Recurring Group (n = 18)	Stable Group (n = 34)	
Age, Years, Mean (SD)	35.9 (13.2)	34.9 (13.9)	.79
Sex, n (%)			.98
Female	13 (72.2%)	23 (67.6%)	
Male	5 (27.8%)	11 (32.4%)	
Years of Education, Mean (SD)	16.1 (2.08)	17.1 (1.88)	.11
BDI Score, Mean (SD)	5.44 (3.81)	2.97 (3.11)	.03 ^a
MADRS Score, Mean (SD)	1.22 (1.56)	1.03 (1.49)	.67
GAF Score, Mean (SD)	83.6 (6.46)	86.5 (4.84)	.10

A total of 52 patients with fully remitted MDD had available baseline data for all predictors included in the multimodal prediction model and were followed up clinically over 14 months to determine recurrence. As expected, there were slightly higher residual depressive symptoms on the more sensitive BDI in the group with a subsequent recurrence.

BDI, Beck Depression Inventory; GAF, Global Assessment of Functioning; MADRS, Montgomery-Åsberg Depression Rating Scale; MDD, major depressive disorder.

^aSignificant group difference at an uncorrected two-sided threshold of $p \leq .05$.

Table 2. Cross-Validated Prediction Model Results

Metric	fMRI ^a	Cognitive ^b	Clinical ^c	Clinical ^c and Cognitive ^b	Clinical ^c and fMRI ^a	All Modalities
Accuracy	0.66	0.68	0.69	0.70	0.83	0.80
Sensitivity	0.32	0.33	0.28	0.33	0.69	0.61
Specificity	0.84	0.86	0.91	0.90	0.90	0.90
PPV	0.59	0.63	0.68	0.71	0.83	0.81
NPV	0.63	0.64	0.63	0.65	0.80	0.76
AUC	0.69	0.67	0.73	0.74	0.90	0.86
Brier Score	0.22	0.20	0.20	0.20	0.13	0.15
Calibration Intercept	-0.22	-0.12	-0.02	-0.10	0.02	-0.08
Calibration Slope	0.50	0.78	0.99	0.83	0.61	0.55

Cross-validated (100× repeated 10-fold cross-validation) performance measures from the different prediction models. Both were adjusted to the 42% prevalence found in the full sample (17).

AUC, area under the receiver-operating characteristic curve; fMRI, functional magnetic resonance imaging; NPV, negative predictive value; PPV, positive predictive value.

^aSelf-blame-related fMRI variables.

^bSelf-blame-related action tendencies task variables.

^cStandard psychological and clinical measures.

Table 3. Model Coefficients Predicting Recurrence

Modality	Predictor Variable	All Modalities β	All Modalities OR	Univariate OR
Clinical/Standard Psychological	Intercept	-2.26	0.10	-
	Global Assessment of Functioning score	-0.19	0.83	0.91
	Number of episodes categorized	1.37	3.94	3.19
	Rosenberg Self-Esteem Scale	-1.00	0.37	0.90
	PANAS negative affect	-2.56	0.08	0.96
	PANAS positive affect	-0.73	0.48	0.86
	Beck Depression Inventory	1.30	3.67	1.23
fMRI	Motion parameters	-1.29	0.28	0.86
	SCC-ATL connectivity (self-blame)	2.43	11.36	2.14
	Basal ganglia-ATL connectivity (self-blame)	2.76	15.80	3.49
	Frontopolar-ATL connectivity (self-blame)	-1.68	0.19	0.73
	Hypothalamus-ATL connectivity (self-blame)	-1.82	0.16	0.76
	Insula-ATL connectivity (self-blame)	-0.14	0.87	1.67
	Amygdala-ATL connectivity (self-blame)	1.46	4.31	1.05
	Hippocampus-ATL connectivity (self-blame)	1.15	3.16	1.38
	SCC-signal time course	1.95	7.03	6.62
	Basal ganglia-signal time course	-0.13	0.88	8.08
	Frontopolar-signal time course	0.10	1.11	6.69
	Hypothalamus-signal time course	1.02	2.77	2.23
	Insula-signal time course	1.25	3.49	4.14
	Amygdala-signal time course	0.09	1.09	1.62
	Hippocampus-signal time course	0.27	1.31	5.53
	ATL BOLD effect (self-blame)	2.14	8.50	1.07
	SCC BOLD effect (self-blame)	0.72	2.05	1.05
Cognitive	Perceived control for self-blame vs. other blame	-0.25	0.78	1.19
	Maladaptive action tendencies	0.91	2.48	2.53

Displayed are the multivariate semi-standardized regularized logistic regression coefficient (β) and OR for each predictor variable from the all modalities model ($n = 52$), which is generated by dCVnet. Recurrence was coded as the positive outcome, so that positive regression coefficients for any predictor can be interpreted as a positive association with recurrence risk. For reference, univariate unregularized logistic regression ORs are provided. For fMRI measures, self-blame indicates that the measurement is derived from a contrast between self-blame trials and fixation cross trials. Motion parameters were derived from the SPM realignment and used root mean squares of the three translation and three rotation parameters, averaged after z transformation. Figure S3 details the a priori rationale for each predictor.

ATL, anterior temporal lobe; BOLD, blood oxygen level-dependent; OR, odds ratio; PANAS, Positive and Negative Affect Schedule; SCC, subgenual cingulate cortex.

Self-blame and Depression Risk Prediction

markers (61). In contrast, our novel cognitive task to capture self-blame-related action tendencies was not sufficient to replace fMRI. We nevertheless obtained AUC values above 0.7, which is considered fair discriminatory value (62), and there was an association of maladaptive action tendencies with subsequent recurrence risk (odds ratio above 2.5). Yet, there was no relevant improvement on known clinical and standard measures regarding its discriminative value. These findings of modest predictive utility are in keeping with the previous literature showing that clinical and standard measures in themselves are unable to make accurate predictions of individual recurrence risk (8,63).

Despite the lack of evidence for the predictive utility of the action tendencies task in this dataset, there are several strategies that could be pursued to improve the value of experimental cognitive tasks in future prediction models. First, our task was purely text based and lacked immersive features. Immersion could be improved by adopting a virtual reality paradigm (64). Second, any single task is unlikely to capture all relevant aspects of vulnerability; thus, future work could integrate a wider range of cognitive measures: for example, a recently developed cognitive task measuring effort and reward decisions showed promising but modest effect sizes for predicting recurrence in MDD (65) resulting in similar AUC values of around 0.70 as our action tendencies task, and its combination with action tendencies may outperform the separate models.

Although a wide variety of clinical and standard measures with known predictive value were assessed, performance of the model using just these features was modest (AUC = 0.73), and one might ask whether important predictors were omitted. Recently, promising results were shown for recurrence risk prediction using a machine learning model that included childhood trauma (66,67); although this needs further replication, we did not measure childhood trauma, and adding this to our prediction model may improve its accuracy in the future. Despite their theoretical importance (2), our model did not include stressful life events during the follow-up period because our aim was prospective prediction from baseline data. We note that although 60% of our 93 patients with completed follow-up data in the overall study reported stressful life events during the 14-month follow-up period (as determined by clinical interview and standardized questionnaires), there was no difference in the rate of occurrence of these events between the recurring episode (61%) and stable remission (59%) groups. This is consistent with the hypothesis that life events trigger recurrence not directly, but rather through their interaction with other factors, such as self-blaming biases (2), a hypothesis that merits further investigation.

Previous studies using machine learning approaches to imaging data in depression have usually relied on voxel-based methods (68). Yet, these methods do not allow integrating cognitive and clinical variables in a straightforward way. Hence, we decided to employ a simpler approach to the imaging analysis, relying on ROIs, and this approach may be more feasible for future clinical applications. By employing the elastic net as a machine learning extension of logistic regression, we alleviated overfitting by shrinking the regression coefficients toward zero, which allows for automatic variable

selection by omitting some predictors. The nested cross-validation is vital to provide realistic estimates of out-of-sample prediction accuracy and thereby estimate internal validity (69), a shortcoming of many studies of clinical predictors that commonly employ shared cross-validation for both model tuning and performance estimation. Although this prognostic model is at an early stage, our approach follows the Medical Research Council Prognosis Research Strategy Partnership (69) guidelines for developing and reporting prognostic models.

On a more cautionary note, for prognostic model development, our sample size was relatively small (70). Although we deliberately adopted more stable hybrid methods than the voxelwise support vector machine considered by Varoquaux (70), we must nevertheless replicate these findings in a larger independent sample before drawing clinical conclusions. Further, we included patients who were fully remitted from symptoms and had no relevant comorbidity; thus, generalization to patients with partial remission and comorbidity will need to be investigated.

Conclusions

Maladaptive self-blame-related action tendencies and fMRI measures predicted subsequent recurrence risk in our MDD sample. When including fMRI, our statistical learning-based risk prediction tool showed promising potential to predict recurrence risk at an individual level. This calls for external validation in an independent sample.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by a Medical Research Council Clinician Scientist Fellowship (Grant No. G0902304 [to RZ]), the LABS-D'Or Hospital Network (Rio de Janeiro, Brazil) (to JM), and a Medical Research Council Doctoral Training Partnership (Grant No. 2064430 [to DF]). RZ and DS were partly funded and AJL was fully funded by the National Institute for Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and by NARSAD Independent Investigator Grant No. 24715 from the Brain & Behavior Research Foundation. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

We are grateful to Dr. Karen Lythe for collecting the primary data.

RZ has collaborations with e-health companies Depsee Ltd, EMIS PLC, and Alloc Modulo Ltd. All other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Centre for Affective Disorders (AJL, SD, DF, TJ, AHY, PD, RZ), Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London; Department of Biostatistics (DS), Institute of Psychiatry, Psychology and Neuroscience, King's College London; and the National Service for Affective Disorders (AHY, PD, RZ), South London and Maudsley NHS Foundation Trust, London, United Kingdom; and the Cognitive and Behavioral Neuroscience Unit (JM, RZ), D'Or Institute for Research and Education, Rio de Janeiro, Brazil.

Address correspondence to Roland Zahn, M.D., Ph.D., at roland.zahn@kcl.ac.uk.

Received Jan 22, 2021; revised Apr 30, 2021; accepted Jun 13, 2021.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.bpsc.2021.06.010>.

REFERENCES

1. Abramson LY, Seligman MEP, Teasdale JD (1978): Learned helplessness in humans - Critique and reformulation. *J Abnorm Psychol* 87:49–74.
2. Harris T (2001): Recent developments in understanding the psychosocial aspects of depression. *Br Med Bull* 57:17–32.
3. Zahn R, Lythe KE, Gethin JA, Green S, Deakin JF, Young AH, *et al.* (2015): The role of self-blame and worthlessness in the psychopathology of major depressive disorder. *J Affect Disord* 186:337–341.
4. Roseman IJ, Wiest C, Swartz TS (1994): Phenomenology, behaviors, and goals differentiate discrete emotions. *J Pers Soc Psychol* 67:206–221.
5. Tangney JP, Stuewig J, Mashek DJ (2007): Moral emotions and moral behavior. *Annu Rev Psychol* 58:345–372.
6. Duan S, Lawrence A, Valmaggia L, Moll J, Zahn R (2020): The role of blame-related action tendencies in the vulnerability to major depressive disorder. medRxiv. <https://doi.org/10.1101/2020.11.15.20232058>.
7. O'Connor LE, Berry JW, Weiss J, Gilbert P (2002): Guilt, fear, submission, and empathy in depression. *J Affect Disord* 71:19–27.
8. Nigatu YT, Liu Y, Wang J (2016): External validation of the international risk prediction algorithm for major depressive episode in the US general population: The PredictD-US study. *BMC Psychiatry* 16:256.
9. Berwian IM, Walter H, Seifritz E, Huys QJ (2017): Predicting relapse after antidepressant withdrawal - A systematic review. *Psychol Med* 47:426–437.
10. Williams LM, Korgaonkar MS, Song YC, Paton R, Eagles S, Goldstein-Piekarski A, *et al.* (2015): Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology* 40:2398–2408.
11. Schmaal L, Marquand AF, Rhebergen D, van Tol M-J, Ruhé HG, van der Wee NJA, *et al.* (2015): Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: A multivariate pattern recognition study. *Biol Psychol* 78:278–286.
12. Fu CH, Mourao-Miranda J, Costafreda SG, Khanna A, Marquand AF, Williams SC, *et al.* (2008): Pattern classification of sad facial processing: Toward the development of neurobiological markers in depression. *Biol Psychol* 63:656–662.
13. Marquand AF, Mourao-Miranda J, Brammer MJ, Cleare AJ, Fu CH (2008): Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. *Neuroreport* 19:1507–1511.
14. Sato JR, Thomaz CE, Cardoso EF, Fujita A, Martin Mda G, Amaro E Jr (2008): Hyperplane navigation: A method to set individual scores in fMRI group datasets. *Neuroimage* 42:1473–1480.
15. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, *et al.* (2017): Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. *Am J Psychiatry* 174:533–545.
16. Frässle S, Marquand AF, Schmaal L, Dinga R, Veltman DJ, van der Wee NJA, *et al.* (2020): Predicting individual clinical trajectories of depression with generative embedding. *Neuroimage Clin* 26:102213.
17. Lythe KE, Moll J, Gethin JA, Workman C, Green S, Lambon Ralph MA, *et al.* (2015): Self-blame-selective hyperconnectivity between anterior temporal and subgenual cortices and prediction of recurrent depressive episodes. *JAMA Psychiatry* 72:1119–1126.
18. Kohavi R (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. Montreal, Quebec: Canada: Morgan Kaufmann, 1137–1143.
19. Efron B (1983): Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc* 78:316–331.
20. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009): Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12:535–540.
21. Zahn R, de Oliveira-Souza R, Moll J (2020): Moral motivation and the basal forebrain. *Neurosci Biobehav Rev* 108:207–217.
22. Lythe KE, Gethin JA, Workman CI, Lambon Ralph MA, Deakin JFW, Moll J, *et al.* (2020): Subgenual activation and the finger of blame: individual differences and depression vulnerability [published online ahead of print Sep 25]. *Psychol Med*.
23. Zahn R, Moll J, Paiva M, Garrido G, Krueger F, Huey ED, *et al.* (2009): The neural basis of human social values: Evidence from functional MRI. *Cereb Cortex* 19:276–283.
24. Zahn R, de Oliveira-Souza R, Bramati I, Garrido G, Moll J (2009): Subgenual cingulate activity reflects individual differences in empathic concern. *Neurosci Lett* 457:107–110.
25. Green S, Lambon Ralph MA, Moll J, Deakin JFW, Zahn R (2012): Guilt-selective functional disconnection of anterior temporal and subgenual cortices in major depressive disorder. *Arch Gen Psychiatry* 69:1014–1021.
26. Hamilton JP, Farmer M, Fogelman P, Gotlib IH (2015): Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry* 78:224–230.
27. Price JL, Drevets WC (2010): Neurocircuitry of mood disorders. *Neuropsychopharmacology* 35:192–216.
28. Ressler KJ, Mayberg HS (2007): Targeting abnormal neural circuits in mood and anxiety disorders: From the laboratory to the clinic. *Nat Neurosci* 10:1116–1124.
29. Green S, Ralph MAL, Moll J, Stamatika EA, Grafman J, Zahn R (2010): Selective functional integration between anterior temporal and distinct fronto-mesolimbic regions during guilt and indignation. *Neuroimage* 52:1720–1726.
30. Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J (2007): Social concepts are represented in the superior anterior temporal cortex. *Proc Natl Acad Sci U S A* 104:6430–6435.
31. Zahn R, Moll J, Iyengar V, Huey ED, Tierney M, Krueger F, *et al.* (2009): Social conceptual impairments in frontotemporal lobar degeneration with right anterior temporal hypometabolism. *Brain* 132:604–616.
32. Wastling SJ, Barker GJ (2015): Designing hyperbolic secant excitation pulses to reduce signal dropout in gradient-echo echo-planar imaging. *Magn Reson Med* 74:661–672.
33. Buckner RL, Andrews-Hanna JR, Schacter DL (2008): The brain's default network: Anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 1124:1–38.
34. Pobric G, Lambon Ralph MA, Zahn R (2016): Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *J Cogn Neurosci* 28:351–360.
35. Simmons WK, Reddish M, Bellgowan PSF, Martin A (2009): The selectivity and functional connectivity of the anterior temporal lobes. *Cereb Cortex* 20:813–825.
36. Olson IR, McCoy D, Klobusicky E, Ross LA (2013): Social cognition and the anterior temporal lobes: A review and theoretical framework. *Soc Cogn Affect Neurosci* 8:123–133.
37. Zahn R, Green S, Beaumont H, Burns A, Moll J, Caine D, *et al.* (2017): Frontotemporal lobar degeneration and social behaviour: Dissociation between the knowledge of its consequences and its conceptual meaning. *Cortex* 93:107–118.
38. Kellner CH, Knapp R, Husain MM, Rasmussen K, Sampson S, Cullum M, *et al.* (2010): Bifrontal, bitemporal and right unilateral electrode placement in ECT: Randomised trial. *Br J Psychiatry* 196:226–234.
39. Leaver AM, Wade B, Vasavada M, Hellemann G, Joshi SH, Espinoza R, *et al.* (2018): Fronto-temporal connectivity predicts ECT outcome in major depression. *Front Psychiatry* 9:11.
40. Zahn R, Lythe KE, Gethin JA, Green S, Deakin JF, Workman C, *et al.* (2015): Negative emotions towards others are diminished in remitted major depression. *Eur Psychiatry* 30:448–453.
41. American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders, 4th ed text rev.* Washington, DC: American Psychiatric Press.
42. Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, *et al.* (1987): The Longitudinal Interval Follow-Up Evaluation: A comprehensive method for assessing outcome in prospective longitudinal studies. *Arch Gen Psychiatry* 44:540–548.
43. Buckman JEJ, Underwood A, Clarke K, Saunders R, Hollon SD, Fearon P, *et al.* (2018): Risk factors for relapse and recurrence of

Self-blame and Depression Risk Prediction

- depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clin Psychol Rev* 64:13–38.
44. Beck AT, Steer RA, Carbin MG (1988): Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev* 8:77–100.
 45. LeMoult J, Kircanski K, Prasad G, Gotlib IH (2017): Negative self-referential processing predicts the recurrence of major depressive episodes. *Clin Psychol Sci* 5:174–181.
 46. First MB, Spitzer RL, Gibbon M, Williams JBW (2002): *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York, NY: Biometrics Research, New York State Psychiatric Institute.
 47. Rosenberg M (1989): *Society and the Adolescent Self-Image*. Middletown, CT: Wesley University Press.
 48. Ormel J, Oldehinkel AJ, Vollebergh W (2004): Vulnerability before, during, and after a major depressive episode - A 3-wave population-based study. *Arch Gen Psychiatry* 61:990–996.
 49. Watson D, Clark LA, Tellegen A (1988): Development and validation of brief measures of positive and negative affect: The PANAS scales. *J Pers Soc Psychol* 54:1063–1070.
 50. Morris BH, Bylsma LM, Rottenberg J (2009): Does emotion predict the course of major depressive disorder? A review of prospective studies. *Br J Clin Psychol* 48:255–273.
 51. de Jonge M, Dekker JJM, Kikkert MJ, Peen J, van Rijsbergen GD, Bockting CLH (2017): The role of affect in predicting depressive symptomatology in remitted recurrently depressed patients. *J Affect Disord* 210:66–71.
 52. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997): Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
 53. Brett M, Anton J-L, Valabregue R, Poline J-B (2002): Region of interest analysis using an SPM toolbox. *Neuroimage* 13:210–217.
 54. van der Ploeg T, Austin PC, Steyerberg EW (2014): Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 14:137.
 55. Zou H, Hastie T (2005): Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* 67:301–320.
 56. Hastie T, Tibshirani R, Friedman JH (2001): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.
 57. Cawley GC, Talbot NLC (2010): On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 11:2079–2107.
 58. Kim J-H (2009): Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 53:3735–3745.
 59. Friedman JH, Hastie T, Tibshirani R (2010): Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1–22.
 60. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* (2010): Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 21:128–138.
 61. Savitz JB, Rauch SL, Drevets WC (2013): Clinical application of brain imaging for the diagnosis of mood disorders: The current state of play. *Mol Psychiatry* 18:528–539.
 62. Fischer JE, Bachmann LM, Jaeschke R (2003): A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Med* 1043–1051.
 63. Wojnarowski C, Firth N, Finegan M, Delgadillo J (2019): Predictors of depression relapse and recurrence after cognitive behavioural therapy: A systematic review and meta-analysis. *Behav Cogn Psychother* 47:514–529.
 64. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M (2016): Virtual reality in the psychological treatment for mental health problems: A systematic review of recent evidence. *Psychiatry Res* 236:189–195.
 65. Berwian IM, Wenzel JG, Collins AGE, Seifritz E, Stephan KE, Walter H, *et al.* (2020): Computational mechanisms of effort and reward decisions in patients with depression and their association with relapse after antidepressant discontinuation. *JAMA Psychiatry* 77:513–522.
 66. Ruhe HG, Mocking RJT, Figueroa CA, Seeverens PWJ, Ikani N, Tyborowska A, *et al.* (2019): Emotional biases and recurrence in major depressive disorder. Results of 2.5 years follow-up of drug-free cohort vulnerable for recurrence. *Front Psychiatry* 10.
 67. Opel N, Redlich R, Dohm K, Zaremba D, Goltermann J, Repple J, *et al.* (2019): Mediation of the influence of childhood maltreatment on depression relapse by cortical structure: A 2-year longitudinal observational study. *Lancet Psychiatry* 6:318–326.
 68. Kambeitz J, Cabral C, Sacchet MD, Gotlib IH, Zahn R, Serpa MH, *et al.* (2017): Detecting neuroimaging biomarkers for depression: A meta-analysis of multivariate pattern recognition studies. *Biol Psychiatry* 82:330–338.
 69. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, *et al.* (2013): *Prognosis Research Strategy (PROGRESS) 3: Prognostic model research*. *PLoS Med* 10:e1001381.
 70. Varoquaux G (2018): Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage* 180:68–77.