Original Research

Prediction of Illness Severity in Patients With Major Depression Using Structural MR Brain Scans

Benson Mwangi, MSc,* Keith Matthews, MD, PhD, and J. Douglas Steele, MD, PhD

Purpose: To develop a model for the prediction of Major Depressive Disorder (MDD) illness severity ratings from individual structural MRI brain scans.

Materials and Methods: Structural T1-weighted MRI scans were obtained from 30 patients with MDD recruited from two different scanning centers. Self-rated (Beck Depression Inventory; BDI), and clinician-rated (Hamilton Rating Scale for Depression, HRSD), syndrome-specific illness severity ratings were obtained just before scanning. Relevance vector regression (RVR) was used to predict the scores (BDI, HRSD) from T1-weighted MRI scans.

Results: It was possible to predict the BDI score (correlation between actual score and RVR predicted scores $r = 0.694$; $P < 0.0001$), but not the HRSD scores ($r = 0.34$; $P = 0.068$) from individual subjects. BDI scores from the most ill patients were predicted more accurately than those from patients who were least ill (standard deviation of difference between predicted and actual scores 2.5 versus 7.4, respectively).

Conclusion: These data suggest that T1-weighted MRI scans contain sufficient information about neurobiological change in patients with MDD to permit accurate predictions about illness severity, on an individual subject basis, particularly for the most ill patients.

Key Words: major depressive disorder; relevance vector regression; pattern classification; multicenter neuroimaging; BDI; Beck Depression Inventory; HRSD; Hamilton Depression Rating Scale
J. Magn. Reson. Imaging 2012;35:64–71.
© 2011 Wiley Periodicals, Inc.

MAJOR DEPRESSIVE DISORDER (MDD) is characterized by mood, cognitive, and in some cases, motor abnormalities (1). The World Health Organization ranks depression as an increasingly dominant contributor to the global burden of disease in terms of Disability Adjusted Life Years (2). Numerous studies have reported a relationship between MDD illness severity measures and functional and structural abnormalities using neuroimaging (3–6). However, conventional approaches can only identify such relationships at group level and not, thus far, at the level of the individual patient. Consequently, previous work has had little impact on informing or modifying clinical practice and the diagnosis, assessment and management of patients with MDD remains entirely clinically based.

Pattern classification techniques, such as support vector machines (SVMs) (7–11), have recently been applied to the classification of individual subject neuroimaging data with success (12–16). Typically, example data with known group membership (e.g., patient or control), are input to the SVM to allow a classification function to be derived. The trained classifier is then tested on data not used for training to determine accuracy of prediction. These techniques have most commonly been applied to two group classification problems (e.g., MDD versus controls), although have also been extended to multi-group classification (17). A further extension of this multi-group approach is the prediction of continuous, noncategorical measures, such as illness severity ratings, using individual scan data (18,19).

The SVM technique has been extended to regression to allow prediction of continuous variables (11). However, the SVM approach has several limitations (7). First, SVM requires estimation of margin "regularization" or "insensitivity" parameters for classification or regression respectively, using cross-validation, which is computationally expensive. Second, predictions made by the SVM approach are not probabilistic and so estimation of the confidence level of predictions is not straightforward. Third, the kernel functions must satisfy Mercer's conditions: i.e., the input matrix is symmetric and positive definite, therefore, limiting kernel function choices (9). Finally, for classification or regression purposes, SVM uses many basis functions (support vectors) which potentially limit generalizability (7). Overcoming these limitations, a sparse Bayesian learning method, Relevance Vector Regression (RVR), has been proposed (20). RVR has been applied successfully to make predictions about Alzheimer's disease (18,19,21) and has been reported to perform more accurately than SVM regression (21,22).

Centre for Neuroscience, Division of Medical Sciences, Ninewells Hospital and Medical School, University of Dundee, United Kingdom.
Contract grant sponsors: Gordon Small Charitable Trust; the Miller McKenzie Trust; and the Scottish Imaging Network (SINAPSE).
*Address reprint requests to: B.M., Center for Neurosciences, Division of Medical Sciences, Level 5, Mailbox 5, Ninewells Hospital and Medical School, University of Dundee DD1 9SY United Kingdom. E-mail: b.m.irungu@dundee.ac.uk
Received April 14, 2011; Accepted August 12, 2011.
DOI 10.1002/jmri.22806
View this article online at wileyonlinelibrary.com.

© 2011 Wiley Periodicals, Inc.

Long established, widely used and extensively validated MDD illness severity measures include the self-report Beck Depression Inventory (BDI, 23) and the observer-rated Hamilton Depression Rating Scale (HRSD, 24). Although studies have reported correlations between abnormalities of brain *function* and MR estimates of structural change in MDD at a group level, to our knowledge, no previous work has reported prediction of continuous MDD illness severity measures from individual scans, such as the BDI and HRSD. For example, Campbell et al (25) reviewed studies of the medial temporal lobe and concluded that small hippocampal volumes were typically reported with recurrent MDD. Hajek et al (26) concluded that there was evidence for reduced anterior cingulate volume in depression in patients with a family history of mood disorder. Koolschijn et al (27) concluded that the largest reductions in gray matter volume tend to be found in the anterior cingulate and orbitofrontal cortex with smaller reductions in the hippocampus and basal ganglia. Notably, no consistent increases in gray matter volume tend to have been reported in studies of MDD. A negative correlation between BDI scores and the right planum temporale volume in MDD has been reported (28). Chen et al (29) found correlations between gray matter volume (GMV) reductions and increased HRSD scores in bilateral dorsal prefrontal, bilateral medial frontal, left inferior frontal, left superior frontal, right orbitofrontal, and cingulate cortices. A further study (30) has reported significant correlations between HRSD scores and GMV in the left dorsolateral prefrontal cortex, post central gyrus, parahippocampal gyrus, inferior temporal gyrus and occipital lobe of MDD patients. A significant correlation between left hippocampal GMV of male patients and HRSD scores has also been reported (31). However, a further study (32), failed to find a correlation between symptom severity and GMV in MDD.

Correlations between neuropsychological measurements and gray matter abnormalities in MDD have also been reported. For example, reduced hippocampal volumes correlated with poorer performance in the Wisconsin Card Sorting Test (WCST) (33) but not performance on the Rey Auditory Verbal Learning Test (RAVLT). Similarly, Li et al (30) reported significant correlations between false alarms (errors of commission) on a *Go/No-go* task and GMV in the superior frontal gyrus, precentral gyrus, and superior frontal gyrus. Additionally, correlations between accuracy of performance on a facial memory task and GMV in occipital lobe and dorsolateral prefrontal cortex have also been reported in MDD patients (30). The presence of such correlations between discrete brain structure volumes and selective aspects of behavioral performance measures—some of which either contribute toward, or closely correlate with MDD symptom burden severity estimates—suggests that prediction of illness severity ratings using RVR may be possible.

The primary objective here was to train a RVR model that would predict BDI and HRSD rating scores given T1-weighted neuroimaging scans from a multi-center dataset. Multi-center data was of particular interest given the increasing popularity of multi-center recruitment to maximize data and to improve generalizability of results. A secondary objective was to investigate which brain regions most strongly influenced predictions.

Table 1
Demographic and Clinical Details*

Characteristic	Cohort A	Cohort B	Significance
Age/years	46.1(12.5)	44.7 (10.0)	$p=0.60^{a}$
Females/total	9/15	10/15	$p=0.41^{b}$
BDI	22.9 (8.2)	38.0 (9.2)	$p<0.001^{c,d}$
HRSD	23.2 (4.3)	28.9 (5.8)	$p=0.02^{d}$

*All tests are two-tailed.
[a]ANOVA.
[b]chi-square test.
[c]Patients versus patients.
[d]Independent groups t-test.

MATERIALS

Subjects

Thirty participants with MDD were recruited in total. Ethical approval was obtained from the relevant local Research Ethics Committees and written informed consent obtained from all participants. Participants were recruited through secondary care mental health service clinics within the Grampian and Lothian National Health Service (NHS) Board areas. In both centers, patients met criteria for a primary diagnosis of DSM IV MDD without significant comorbidity. Specifically, participants were excluded if they met criteria for other psychiatric diagnoses including personality disorder, a lifetime history of meeting DSM-IV criteria for substance misuse, gross structural brain abnormality observable radiologically, use of non-antidepressant and related medication and neurological disorder. Patients with co-morbid anxiety disorders were not excluded because anxiety symptoms are so commonly co-morbid with MDD. However, patients with claustrophobia were excluded as they were unlikely to tolerate scanning.

Diagnosis was made on the basis of detailed case note review, interview with clinical staff providing care for the patient, and a clinical interview performed by an experienced consultant psychiatrist (J.D.S.). All ratings for patients in both centers were performed by the same individual (J.D.S.) on the morning immediately *before* scanning. Typically, patients had many years of recurrent or continuous illness despite compliance with antidepressant treatment and general psychiatric management. For inclusion, patients had to have a minimum continuous duration of symptomatic illness greater than 3 months in the present episode, although typically the current illness duration significantly exceeded this. All medications were stable for a minimum of 1 month before scanning. For study inclusion the minimum illness severity rating on the Hamilton Depression Rating Scale (HRSD) (24) was greater than 21. For more details about the cohorts and clinical ratings see Table 1.

MRI Data Acquisition

T1-weighted MR images of brain structure were acquired at two imaging centers (one scanner per center): University of Aberdeen (Cohort A) and University of Edinburgh (Cohort B). Functional MRI data were also acquired and have been reported elsewhere (34,35). The same manufacturer and field strength of scanner was used in both centers: 1.5 Tesla (T) GE Medical Systems Signa. The T1 acquisition parameters were: Center A: repetition time (TR) 20 ms, echo time (TE) 6 ms, flip angle 35 degrees, 124 contiguous 1.6-mm axial slices of 256 × 256 voxels with an in-plane resolution of 0.938 mm^2. Center B: TR 20 ms, TE 6 ms, flip angle 15 degrees, 124 contiguous 1.7-mm axial slices of 256 × 256 voxels with an in-plane resolution of 0.938 mm^2.

METHODS

MR Image Processing

T1-weighted images were visually inspected for artifact then segmented into gray matter maps using the unified segmentation algorithm as implemented in statistical parametric mapping (SPM5) (36). This process comprises normalization to Montreal Neurological Institute (MNI) stereotactic space with modulation to control rescaling effects. Following previous studies using RVR (18,19,21,37) and as usual in voxel based morphometry, images were smoothed (with a 6-mm full-width half-maximum Gaussian kernel) to increase the signal-noise ratio.

Multiple linear regression was done using SPM with three covariates (Center, BDI and HRSD scores) to test the null hypotheses of no difference between the two scanner-patient groups with Center as the covariate of interest. The threshold of significance was defined at a voxel level as $P < 0.05$ and at a whole brain cluster level of $P < 0.05$ by requiring clusters to exceed 144 continuous voxels, this latter value being identified using a Monte-Carlo method (38). Brain regions which differed significantly between centers were excluded from the later predictive regression analyses by masking.

Whole brain smoothed voxel values (as above) were used as input feature vectors because RVR is a very sparse model, meaning it is dependent on only a small set of kernel functions for making predictions (20). This was achieved by pruning non-necessary kernel functions leading to fewer relevance vectors. Bishop and Tipping provide a detailed discussion of RVR sparsity issues (7,20). In contrast to a nonlinear kernel method, no kernel parameter estimation was required before training and testing the RVR predictor.

Sparse Bayesian Learning

RVR (7,20) is a kernel learning method formulated using a Bayesian framework. Given a supervised learning problem with N number of observations, $\{X_n, t_n\}_{n=1}^{N}$ with X_n representing the *input* data (e.g., structural T_1-weighted MRI scan) and t_n a corresponding *target* (e.g., BDI or HRSD continuous variable score). The objective is to predict a target score t given an unseen sample X. This *input-target* pair can be represented as a standard linear model

$$t_n = y(X_n; W) + \varepsilon_n \quad [1]$$

where $y(X_n; W) = W^T\phi(\tilde{x})$, $\phi : \Re^k \rightarrow \Re^z$ is a kernel mapping procedure from a k dimensional space to a z dimensional space (e.g., linear kernel or Gaussian kernel). This results in the formulation,

$$t_n = \sum_{n=1}^{N} \omega_n K(x, x_n) + b + \varepsilon_n \quad [2]$$

where $W = (\omega_1, \omega_1, \ldots, \omega_N)^T$, is a weighting vector b is a bias parameter and ε_n represents measurement noise assumed to be independent with zero mean Gaussian distribution and variance σ^2. In this study, a linear kernel $k(x, x_n) = x_i^T x_j$ was used. As a result, the maximum-likelihood estimation of the weight vector W is computed as

$$p(t/W, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\|t - \Phi W\|^2\right\} \quad [3]$$

where $t = (t_1 \ldots t_N)^T$, Φ is an $N \times (N + 1)$ design matrix where $\Phi = [\phi(x_1), \phi(x_2), \ldots \phi(x_N)]^T$ and $\phi(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \ldots, K(x_n, x_N)]^T$. This is similar to the common least square estimate and suffers from overfitting (20). To avoid this, the Bayesian approach is adopted by introducing an additional constraint in form of a prior distribution over the parameters (20). The zero-mean Gaussian prior distribution of the parameters is defined as;

$$p(W/\alpha) = \prod_{n=0}^{N} \mathrm{N}(\omega_n|0, \alpha_n^{-1}) \quad [4]$$

α is a vector of $N + 1$ hyperparameters, with each weighting being paired with an individual hyperparameter. Priors of the hyper-parameters α and the noise variance σ^2 are also defined and this process is explicitly explained in (7,20,39). Using the likelihood (Eq. [3]) and the prior (Eq. [4]) the Bayesian formulation is completed as;

$$\text{Posterior distribution} = \frac{\text{likelihood} \times \text{prior}}{\text{normalisation factor}} \quad [5]$$

The posterior distribution of the weights W is given as

$$p(W|t, \alpha, \sigma^2) = \frac{p(t|W, \sigma^2)p(W|\alpha)}{p(t|\alpha, \sigma^2)} \quad [6]$$

Because the prior and the likelihood are Gaussian, the posterior distribution is also Gaussian with mean μ and covariance Σ

$$\begin{aligned} p(W|t, \alpha, \sigma^2) &= \mathrm{N}(\mu, \textstyle\sum) \\ \mu &= \sigma^{-2} \textstyle\sum \Phi^T t \\ \textstyle\sum &= (\sigma^{-2}\Phi^T\Phi + A)^{-1} \\ A &= \mathrm{diag}(\alpha_0, \alpha_1, \ldots \alpha_N) \end{aligned} \quad [7]$$

Table 2
Correlation between Illness Severity ratings and Grey Matter Volume (GMV) Reduction

Clinical score	Anatomical region	MNI coordinates (x, y, z)	Z value
BDI	Medial frontal gyrus	−8, 30, −22	3.57
	Parahippocampal gyrus	−36, −36, −14	2.92
	Superior temporal gyrus	56, 8, −30	2.71
HRSD	Middle frontal gyrus	−30, 0, 62	3.09
	Cingulate gyrus	2, 2, 30	3.03
	Insula	46,−2, 2	2.63
	Cerebellum	30, −40, −44	3.43

where A is a diagonal matrix depicting the inverse of the noise for each weight. By integrating out the weights (Eq. [8]), which a marginal likelihood function is maximized (7,20).

$$p(t|\alpha,\sigma^2) = \int p(t|W,\sigma^2)p(W|\alpha)dw \qquad [8]$$

Maximization of this marginal likelihood function (Eq. [8]) aims to find the priors of the hyper-parameters A and σ^2. More details about this maximization procedure are given in (7,20,39). However, during this process, some of the estimated hyper-parameters α tend to infinity, suggesting a small prior variance meaning the parameter posterior probabilities are infinitely peaked at zero. In these cases the estimated weights ω_n are set to zero, this process being known as Automatic Relevance Determination (ARD, 40). This procedure makes models generated by the RVR method sparse with only a few non-zero weights which are subsequently used for predictions. Fundamentally, this process prunes nonessential weights making the model sparse, meaning only a small fraction of the data are used for predictions. These data are "relevance vectors" equivalent to "support vectors" in SVM. RVR predictions are defined as

$$t_{\text{new}} = \sum_{n=0}^{N} \mu_n \phi_{n^*} \qquad [9]$$

where μ is the posterior mean of the weights (Eq. [7]) and ϕ is an element of the $N \times N + 1$ matrix (Eq. [3]).

RVR Performance Evaluation

Leave-one-out cross-validation was used to test the trained RVR's ability to generalize from unseen data (10). Specifically, assuming the total number of subjects to be N, training was performed using $N - 1$ subjects, with each subject having an MR scan paired with an illness rating score, testing being done using the excluded sample. This calculation was repeated N times until all samples had been excluded once.

Rating score prediction performance was assessed using two measures. The Pearson's Correlation Coefficient (PRC) between predicted and actual rating scores was used to test the null hypothesis of no significant correlation. The differences between predicted and actual scores were assessed using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Voxel Based Morphometry

Additionally, using SPM5, a multiple linear regression calculation was used to test the null hypotheses of no relation between illness severity ratings (HRSD, BDI) and gray matter change, with Center as a covariate of no interest. The same cluster threshold as above was used to define the threshold of significance. Using a simple linear regression model we also investigated the relationship between the two rating scales and subject age.

RESULTS

Table 1 summarizes the socio-demographic and clinical information for the study participants. There were no significant differences between groups with respect to age and gender ratio, but cohort B patients were more symptomatic. All patients except one were receiving antidepressant medications. None were receiving other antidepressant treatments at the time of scanning. There were no correlations between scores on the BDI and age (r = −0.234; $P = 0.214$), nor HRSD score and age (r = 0.016; $P = 0.932$).

Table 2 shows significant correlations between gray matter reductions (scanning center being a covariate of no interest) and illness severity ratings. Increased HRSD scores were particularly associated with gray matter *reductions* in the middle frontal cortex, cingulate gyrus, insula, and cerebellum; increased BDI scores with reductions in the medial frontal gyrus, parahippocampal gyrus, and superior temporal gyrus. The threshold of significance was as defined above.

Figure 1 shows a scatter plot of BDI versus HRSD scores. The scores from the two rating scales were moderately correlated (r = 0.523, $R^2 = 0.27$; $P =$

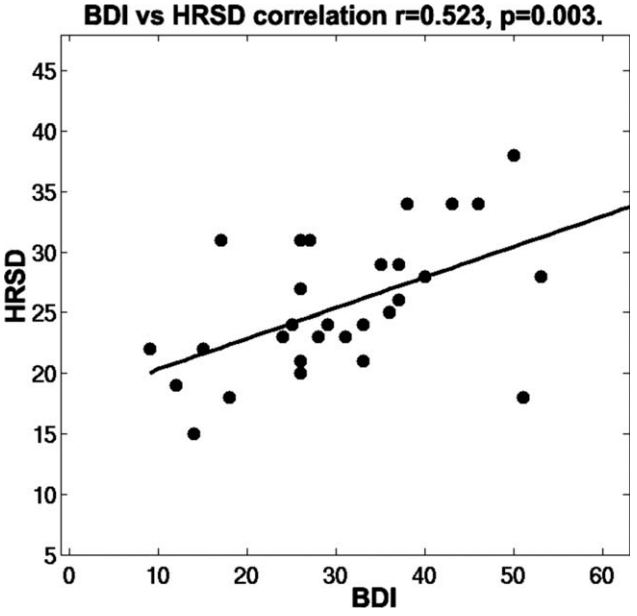

Figure 1. A scatter plot of BDI scores against HRSD scores.

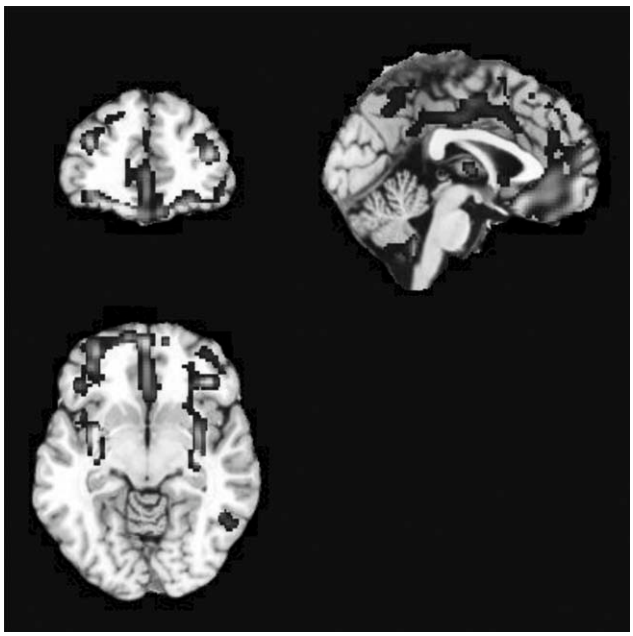

Figure 2. Brain regions which differed significantly between centers and were excluded from the RVR calculation.

0.003). Figure 2 shows the brain regions which differed due to the between-scanner effect and which were masked out of the RVR calculation. These include middle frontal gyrus, inferior parietal lobule and superior frontal gyrus. A previous study explored Aberdeen and Edinburgh scanner effects using a different group of only healthy subjects, each subject being scanned twice on the two scanners (41). The same scanner affected brain regions were reported.

Table 3 shows RVR prediction results for both BDI and HRSD scales. The BDI had the highest correlation between RVR predicted score and actual score ($r = 0.694$; $P < 0.0001$). The HRSD scale correlation was not significant ($r = 0.34$; $P = 0.068$), but there was a suggestion of a trend.

Figure 3 shows scatter plots of predicted versus actual score for the BDI and HRSD. BDI score predictions show a wider variance for lower scores. The range of BDI scores was therefore divided into three and the standard deviation of the difference between the predicted and actual scores calculated, for the highest and lowest 30th percentiles of BDI scores. The standard deviations were found to be 2.5 and 7.4 for the highest and lowest, respectively, indicating that BDI scores from more ill patients could be predicted more accurately.

Figure 4 shows brain regions that contributed to the significant ($P < 0.0001$) BDI score prediction. More highly weighted regions are indicated by colors ranging from blue through green to yellow (highest). The brain regions which contribute to classification are widely distributed throughout the brain, but there are noticeably more strongly predictive loci in the hippocampus (−32, −25, −17), superior temporal gyrus (47, −38, 7), and medial brain structures including the cingulate sulcus (12, 45, 18).

DISCUSSION

Using data from single MR structural brain scans from patients with MDD, we were able to reliably predict severity of illness according to scores on the self-rated BDI, but not the observer-rated HRDS, although a possible trend was observed in the latter case. Individual subject BDI rating scores could be predicted successfully from T1-weighted images with the smallest standard deviation in prediction observed for the most ill patients. This is consistent with the observation that the most pronounced gray matter reductions tend to be reported in the most ill patients (42,43). It may also suggest a limitation of the method, in that illness severity prediction may be inherently more difficult with less severely ill patients. The lack of correlation between illness severity measures and age, however, indicates that better prediction was not simply occurring in older patients some of whom may have had more marked brain abnormalities.

We chose the RVR method as it has been reported to be superior to SVR (21). We additionally chose to avoid dimensionality reduction techniques such as principal component analysis (PCA), as RVR is a fairly sparse algorithm which utilizes a fraction of its basis functions to make predictions. A study of Alzheimer's disease prediction (18) similarly used voxels from the whole brain as input features into RVR for predicting clinical scores, without dimensionality reduction, with some success.

A linear kernel, in contrast to a nonlinear kernel, has the advantage of providing accessible information on how strongly different brain regions contributed to prediction, as characterized by RVR weighting. Whereas widespread regions contributed to prediction, regions which particularly contributed to accurate prediction of the BDI score included the hippocampus and medial frontal cortex including the cingulate. These regions have been reported to have abnormal structure and function in many group level studies of depressive illness (3,35,43). The work reported here indicates that structural brain abnormalities in these regions contain information to allow illness severity prediction at an individual subject level.

To our knowledge, this is the first study reporting prediction of individual subject MDD illness severity scores using structural MRI scans. The "gold

Table 3
RVR Prediction Results for the Two Rating Scales

Statistical measure	Rating scale	
	BDI	HRSD
Pearson correlation coefficient and significance	0.694 $P<0.0001$	0.34 $P=0.068$
Mean absolute error	6.3	5.1
Root mean squared error	8.3	6.6
Standard deviation of residuals	8.5	6.7
Standard deviation of residuals (Scores split into ranges: 1, Low; 2, Medium; 3, High)	Low-7.38 Medium-7.07 High-2.46	Low-7.12 Medium-2.44 High-6.58

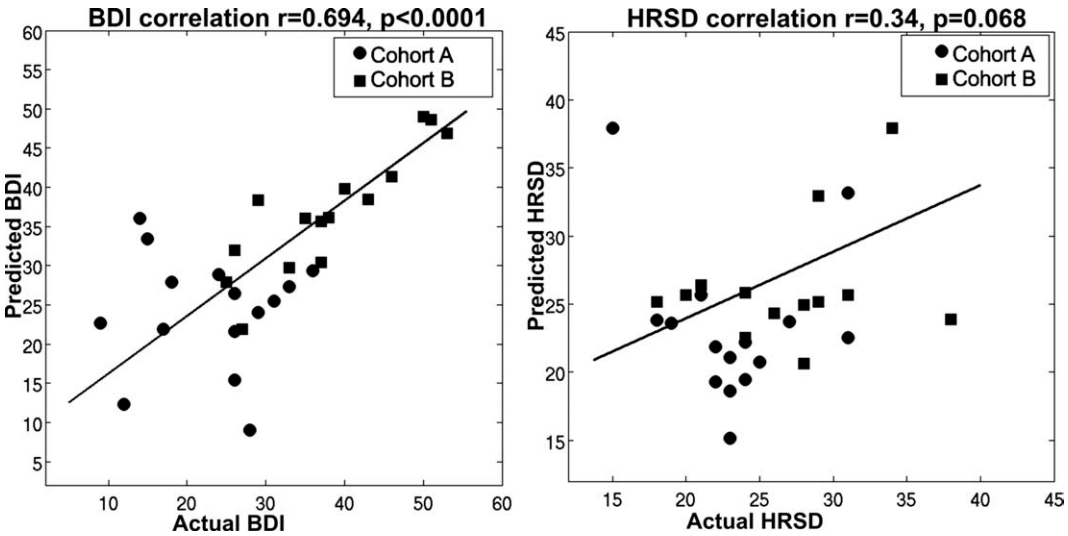

Figure 3. Scatter plots for BDI and Hamilton actual versus predicted scores. BDI predictions were significant ($P < 0.0001$) but Hamilton scores were not ($P = 0.068$).

standard" method of assessing illness severity will of course remain clinical assessment for the foreseeable future. The RVR approach is a departure from established neuroimaging analysis methods, which identify group level differences and correlations with illness severity. Extensive further work is required to develop a technique that can be used routinely in a clinical context. However the results of this study indicate that RVR has potential to provide useful information on individual patients with MDD using MRI.

The reason for the HRSD scores being "less predictable" than the BDI scores is unclear. Replication using different data would be important to determine if this is a general finding. The BDI and HRSD attempt to measure different things and there is no consensus in psychiatry which, if either, is best. The BDI focuses more on measuring cognitive symptoms compared with the HRSD. The BDI is a self-administered rating scale; the HRSD is administered by a semi-structured interview with the patient and scores are estimated by a trained observer. However, the rating scales are structured differently with both scales giving variable emphasis on somatic versus psychological symptoms. It has been suggested that the HRSD score is predominantly affected by somatic symptoms and to a greater extent than the BDI (44,45). Conversely, the BDI measures MDD symptoms which are stable over time and therefore thought to underestimate the extent of improvement during treatment as compared to HRSD (44). Self-rating scales are often regarded as more susceptible to inter-patient differences in self-rating of illness than clinician rating scales, but advantages of self-rated scales have also been identified (44). Consistent with previous studies (46,47), we found only a moderate degree of correlation between BDI and HRSD scores, which may explain some of the difference in ability to predict scores.

Limitations of this study should be noted. Patients were taking a variety of antidepressant medications which reflects standard psychiatric clinical practice. A link between brain structure changes and medication is possible, although it is not possible to test for a dose effect function directly as there is no accepted "common scale" for different antidepressants. However, it is highly unlikely these results reflect an

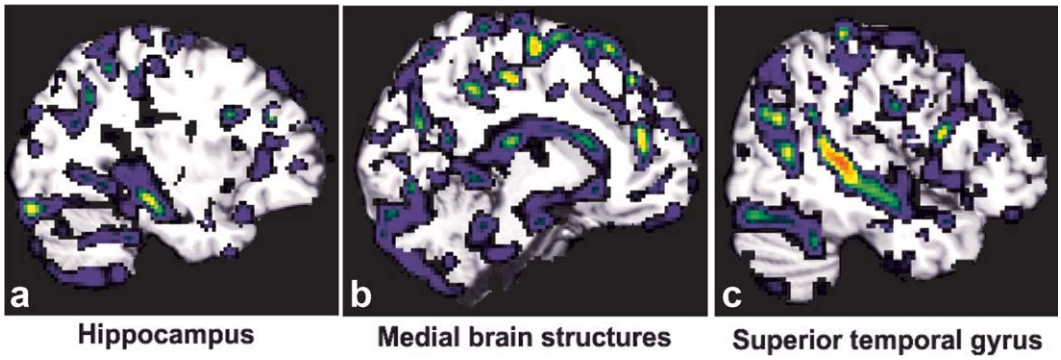

Figure 4. Weighted image derived from RVR weights used for predicting BDI scores. Progressively higher weighting indicated by blue (low), to green (medium), then yellow (high). The hippocampus, medial brain regions and superior temporal gyrus predicted BDI scores most strongly.

antidepressant effect as a negative correlation between increased illness severity and gray matter volume was found. Gray matter reductions are thought unlikely to be due to medication as there is little evidence for antidepressants *decreasing* measures of brain volume. Indeed there is evidence for antidepressant-associated neurogenesis in brain regions such as the hippocampus (48). It is possible that illness duration affects brain structure. Here, illness duration and severity may have been correlated, although the oldest patients were not the most ill patients. It was not possible to obtain accurate information on total illness duration due to incomplete recovery between previous episodes of illness and incomplete information in the case notes. A planned study with different participants could test this hypothesis. Scanner affected brain regions masked out of the RVR calculation included parts of the medial frontal and temporal lobes. This could have affected predictive accuracy and be avoided in future studies by acquiring a larger dataset from a single center. There are a relatively small number of patients in the current study and the results will have to be confirmed in a larger cohort.

In summary, using a multi-center dataset of 30 patients with MDD, it was possible to predict illness severity as quantified by the self-report BDI score. Although brain regions supporting prediction were widespread, they included regions such as the hippocampus and cingulate that have often been reported to be structurally and functionally abnormal at a group level in studies of MDD. With refinement and replication, studies of predictive classification may further advance our understanding of the regional neurobiology of MDD and offer the potential to develop diagnostic biomarkers.

ACKNOWLEDGMENTS

The study was funded by the Gordon Small Charitable Trust and the Miller McKenzie Trust. BM's PhD studentship was funded by the Scottish Imaging Network (SINAPSE).

REFERENCES

1. Sobin C, Sackeim H. Psychomotor symptoms of depression. Am J Psychiatry 1997;154:1.
2. World Health Organization. Mental Health: new understanding, new hope. Geneva, Switzerland: World Health Organization; 2001.
3. Ebmeier K, Donaghey C, Steele J. Recent developments and current controversies in depression. Lancet 2006;367:153–167.
4. Shah P, Ebmeier K, Glabus M, Goodwin G. Cortical grey matter reductions associated with treatment-resistant unipolar depression. Controlled magnetic resonance imaging study. Br J Psychiatry 1998;172:527–532.
5. Takahashi T, Yucel M, Lorenzetti V, et al. An MRI study of the superior temporal subregions in patients with current and past major depression. Prog Neuropsychopharamacol Biol Psychiatry 2010;34:98–103.
6. Heinzel A, Grimm S, Beck J, et al. Segregated neural representation of psychological and somatic-vegetative symptoms in severe major depression. Neurosci Lett 2004;456:49–53.
7. Bishop C. Pattern recognition and machine learning. London: Springer-Verlag; 2007.
8. Duda R, Hart P, Sork D. Pattern classification. London: John Wiley and Sons; 2001.
9. Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. Cambridge: University Press; 2004.
10. Theodoridis S, Koutroumbas K. Pattern recognition, 3rd ed. London: Academic Press; 2006.
11. Vapnik V. Statistical learning theory. New York: John Wiley and Sons; 1998.
12. Costafreda S, Chu C, Ashburner J, Fu C. Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS One 2009;4:e6353.
13. Fu C, Mourao-Miranda J, Costafreda S, et al. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol Psychiatry 2008;63:656–662.
14. Kloppel S, Stonnington C, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131:681–689.
15. Koutsouleris N, Meisenzahl E, Davatzikos C, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch Gen Psychiatry 2009;66:702–712.
16. Lao Z, Shen D, Xue Z, Karacali B, Resnick S, Davatzikos C. Morphological classification of brains via highdimensional shape transformations and machine learning methods. Neuroimage 2004;21:46–57.
17. Mourão-Miranda J, Reynaud E, McGlone F, Calvert G, Brammer M. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. Neuroimage 2006;33:1055–1065.
18. Stonnington C, Chu C, Kloppel S, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage 2010;51:1405–1413.
19. Wang Y, Fan Y, Bhatt P, Davatzikos C. High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. Neuroimage 2010;50:1519–1535.
20. Tipping M. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 2001;1:211–244.
21. Franke K, Ziegler G, Kloppel S, Gaser C; Alzheimer's Disease Neuroimaging Initiative. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. Neuroimage 2010;50:883–892.
22. Formisano E, De Martino F, Valente G. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. Magn Reson Imaging 2008;26:921–934.
23. Beck A, Ward H, Mendelson M. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571.
24. Hamilton M. Rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
25. Campbell S, Marriott M, Nahmias C, MacQueen G. Lower hippocampal volume in patients suffering from depression: a meta-analysis. Am J Psychiatry 2004;161:598–607.
26. Hajek T, Koxeny J, Kopecek M, Alda M, Hoschl C. Reduced subgenual cingulate volumes in mood disorders: a meta-analysis. J Psychiatry Neurosci 2008;33:91–99.
27. Koolschijn P, van Haren N, Lensvelt-Mulders G, Hulshoff Pol H, Khan R. Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. Hum Brain Mapp 2009;30:3719–3735.
28. Takahashi T, Yucel M, Lorenzetti V, et al. An MRI study of the superior temporal subregions in patients with current and past major depression. Prog NeuroPsychopharamacol Biol Psychiatry 2010;34:98–103.
29. Chen C, Ridler K, Suckling J, et al. Brain Imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. Biol Psychiatry 2007;62:407–414.
30. Li C, Lin C, Chou K, Che I, Hsieh J, Wu C, Lin W, Su T. Structural and cognitive deficits remitting and non-remitting recurrent depression: a voxel-based morphometry study. Neuroimage 2010;50:347–356.
31. Vakili K, Pillay S, Lafer B, et al. Hippocampal volume in primary unipolar major depression: a magnetic resonance imaging study. Biol Psychiatry 2000;47:1087–1090.
32. Kim JM, Hamilton P, Gotlib H. Reduced caudate gray matter volume in women with major depressive disorder. Psychiatry Res 2008;164:114–122.

33. Frodl T, Schaub A, Banac S, et al. Reduced hippocampal volume correlates with executive dysfunctioning in major depression. J Psychiatry Neurosci 2006;31:316–325.
34. Kumar P, Waiter G, Ahearn T, Milders M, Reid I, Steele J. Abnormal temporal difference reward-learning signals in major depression. Brain 2008;131:2084–2093.
35. Steele J, Kumar P, Ebmeier K. Blunted response to feedback information in depressive illness. Brain 2007;130:2367–2374.
36. Friston K, Ashburner J, Kiebel S, Nichols T, Penny W. Statistical parametric mapping: the analysis of functional brain images. London: Academic Press; 2007.
37. Chu C, Ni Y, Tan G, Saunders CJ, J A. Kernel regression for fMRI pattern prediction. Neuroimage 2011;56:662–673.
38. Slotnick S, Moo L, Segal J, Hart J. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn Brain Res 2003;17:75–82.
39. Rasmussen C, Williams C. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2006.
40. MacKay DJC. Bayesian methods for backpropagation networks. Models of Neural Networks III: New York: Springer; 1994. p 211–254.
41. Moorhead T, Gountouna V, Job D, et al. Prospective multi-centre voxel based morphometry study employing scanner specific segmentations: procedure development using Calibrain Structural MRI data. BMC Med Imaging 2009;9:8.
42. Ebmeier KP, Donaghey C, Steele JD. Recent developments and current controversies in depression. Lancet 2006;367:153–167.
43. Lorenzetti V, Allen NB, Fornito A, Yucel M. Structural brain abnormalities in major depressive disorder: a selective review of recent MRI studies. J Affect Disord 2009;117:1–17.
44. Cusin C, Yang H, Yeung A, Fava M. Rating scales for depression. Curr Clin Psychiatry 2010:7–35.
45. Bagby M, Ryder A, Schuller D, Marshall M. The Hamilton depression rating scale: has the Gold standard become a lead weight? Am J Psychiatry 2004;161:161–112.
46. Carter J, Frampton C, Mulder R, Luty S, Joyce P. The relationship of demographic, clinical, cognitive and personality variables to the discrepancy between self and clinician rated depression. J Affect Disord 2010;124:202–206.
47. Freedland K, Skala J, Carney R, et al. The depression interview and structured Hamilton (DISH): rationale, development, characteristics, and clinical validity. Psychosom Med 2002;64:897–905.
48. Sapolsky RM. Depression, antidepressants, and the shrinking hippocampus. Proc Natl Acad Sci U S A 2001;98:12320–12322.