DOI:
10.1111/acps.12887

Document Version
Peer reviewed version

A Morphometric Signature of Depressive Symptoms in Unmedicated Mood Disorders

Toby Wise[1,2,3], Lindsey Marwood[1,4], Adam M Perkins[1,4], Andres Herane-Vives[1,5,7], Steve CR Williams[4,6], Allan H Young[1,4,7], Anthony J Cleare[1,4,7]*, Danilo Arnone[1,7]*

* These authors contributed equally to this work

[1]Centre for Affective Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[2]Wellcome Trust Centre for Neuroimaging, University College London, London, UK

[3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK

[4]National Institute for Health Research Biomedical Research Centre, South London and Maudsley NSH Foundation Trust, London, UK

[5]Departamento de Clínicas, Facultad de Medicina, Universidad Católica del Norte, Coquimbo, Chile

[6]Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[7]South London and Maudsley NHS Foundation Trust, London, UK

Word count:

Abstract: 183

Main article: 4707

Tables: 1

Figures: 3

Running title: Grey matter correlates of depressive and hypomanic symptoms

Contact: Toby Wise: t.wise@ucl.ac.uk; Danilo Arnone: danilo.arnone@kcl.ac.uk

Grey matter correlates of depressive and hypomanic symptoms

Abstract

Objective

A growing literature indicates that unipolar and bipolar depression are associated with alterations in grey matter volume. However, it is unclear to what degree these patterns of morphometric change reflect symptom dimensions. Here, we aimed to predict depressive symptoms and hypomanic symptoms based on patterns of grey matter volume using machine learning.

Method

We used machine learning methods combined with voxel-based morphometry to predict depressive and self-reported hypomanic symptoms from grey matter volume in a sample of 47 individuals with un-medicated unipolar and bipolar depression.

Results

We were able to predict depressive severity from grey matter volume in the antero-ventral bilateral insula in both unipolar and bipolar depression. Self-reported hypomanic symptoms did not predict grey matter loss with a significant degree of accuracy.

Discussion

The results of this study suggest that patterns of grey matter volume alteration in the insula are associated with depressive symptom severity across unipolar and bipolar depression. Studies using other modalities, and exploring other brain regions with a larger sample are warranted to identify other systems that may be associated with depressive and hypomanic symptoms across affective disorders.

Keywords: Depression, bipolar disorder, magnetic resonance imaging, MRI, Machine learning, VBM, DARTEL

Significant outcomes

- Grey matter volume, in regions previously shown to be commonly reduced in unipolar and bipolar depression, is associated with depressive symptom severity across these disorders. This relationship is most pronounced in the insula.
- Depressive severity can be predicted accurately at the individual level based on grey matter volume
- Self-reported hypomanic symptom severity was not able to be predicted accurately using grey matter volume

Limitations

- This study relies on retrospective, self-reported hypomanic symptoms which may be inaccurate
- The number of subjects with bipolar disorder was low relative to the number with unipolar depression.

Introduction

Affective disorders such as major depression and bipolar disorder are common disorders with profound effects on sufferers and society (1,2). Depressive episodes, often with a recurrent pattern, occur in both conditions and are generally considered the most burdensome disease contributor leading to substantial functional impairment. Identifying biological signatures of clinical symptoms of depression and mood elevation within affective disorders at the earliest opportunity is a research priority necessary to aid diagnostic identification and differentiation within affective disorders and provide effective treatment with increased specificity (3).

Recent research has indicated similarities in grey matter volume reduction across the brain between unipolar and bipolar disorders (4–7). Notably, we have demonstrated a specific pattern of volumetric reduction in the medial prefrontal cortex and insula bilaterally in both unipolar and bipolar disorders, which may represent the morphometric substrate of depressive symptoms across mood disorders (7). We have also identified a preferential pattern of grey matter loss in unipolar major depression compared to bipolar disorder in the left hippocampus, right middle temporal gyrus and right dorsolateral prefrontal cortex (7) which might aid clinical identification of bipolarity in the presence of a depressed mood state. Early identification of bipolarity at the brain level is a research target as diagnostic discrimination can be clinically very difficult as "unipolar" major depressive episodes are commonly the first presentation (8–10). Furthermore, individuals with a bipolar diathesis, even below the threshold for diagnosis with bipolar disorder, are more likely to develop more severe illness trajectories (11), and attempt suicide (10).

However, the networks we identified originated from a large meta-analysis of cross sectional studies and remain fundamentally untested in a new sample of people with mood disorders in relation to mood state. This is an important next step to precisely define the veracity of this pattern of brain abnormalities and their relationship to mood state across psychiatric disorders with overlapping clinical characteristics. This is especially relevant for regions like the insula known to be involved in a number of psychiatric presentations (12).

In this study, we sought to confirm whether our pattern of pre-identified grey matter loss in the prefrontal cortex and insula, common to both unipolar and bipolar depression, relate to symptoms of major depression in a sample of medication free individuals with mood disorders. We attempted to identify an unknown bipolar diathesis, by tentatively evaluating the predictive value of variation in grey matter volume in the brain regions shown to differentiate unipolar and bipolar disorders in our model (left hippocampus, right middle temporal gyrus and right dorsolateral prefrontal cortex). This was a secondary analysis because the study investigated depressed individuals in the absence of overt symptoms of mania or hypomania. Hence, we expected that alterations in the brain will be less noticeable in the absence of current symptoms. For this purpose we employed a self-report measure of hypomanic symptoms, the 'Hypomania Checklist' in view of its recognised sensitivity in detecting a bipolar diathesis in patients presenting with symptoms of unipolar depression (13).

We measured grey matter volume using voxel-based morphometry (14), which provides an effective method of mapping grey matter volume across the brain, providing an effective way of identifying localised volumetric changes in clinical groups. In effect, the method provides a measure of individual subjects' grey matter volume in each voxel of a structural image. While these images have most commonly been compared between groups using standard statistical methods, more recently machine learning techniques have been applied to these measures (3). An important feature of these methods is that, rather than simply comparing the value of each voxel between groups, they use the pattern of grey matter volume measurements in a particular region to either classify diagnostic groups or predict a continuous clinical measure. These methods have two major advantages over traditional univariate methods: firstly, their multivariate nature allows detection of complex patterns represented across multiple voxels, unlike traditional univariate analyses where each voxel is analysed independently, and secondly, they allow single-subject prediction, enabling us to identify an individual's level of symptoms based on brain structure. This is of particular relevance to psychiatric disorders because it allows more subtle detection of distributed patterns and individual-subject level

predictions of clinical status. We performed this study in a sample of medication-free individuals experiencing a depressive episode, focusing on current depressive symptoms and self-reported historic hypomanic/manic symptoms.

While previous studies have addressed similar questions (15,16), our work is the first to use multivariate methods to assess the relationship between volumetric patterns and symptom severity, and is also the first to investigate this relationship across unipolar and bipolar depression.

We firstly hypothesised that grey matter volume in regions demonstrating volumetric reduction across unipolar and bipolar disorders, including the bilateral insula and medial prefrontal cortex, would predict depressive symptom severity irrespective of diagnosis, providing a common substrate for depressive symptoms in affective disorders. Secondly, we explored the possibility that grey matter volume in regions previously shown to distinguish unipolar and bipolar disorders, including the hippocampus and dorsolateral prefrontal cortex, would predict self-reported hypomanic symptom severity.

Methods

Participants

Participants were recruited from public advertisements (17) and local psychological therapy services and underwent a diagnostic interview based on the Mini International Neuropsychiatric Interview (18). Subjects were recruited from our research programme investigating biological mechanisms underpinning major depression and bipolarity by using neuroimaging techniques.

Patients were assessed by a psychiatrist (DA) and a trained researcher (LM) and met Axis I DSM-IV criteria for unipolar major depression or bipolar disorder. Inter-rater reliability for depressive symptoms was high (intraclass correlation coefficient = 0.96, p = .004). The

diagnosis was further substantiated by review of medical notes and collateral information to exclude axis II comorbidity as far as possible. The Young Mania Rating Scale (19) was used to exclude current hypomania/mania mixed episodes in bipolar patients. Duration of illness was calculated as the number of years since the onset of any mood episode.

Inclusion criteria were: 1) experiencing a moderate to severe depressive episode at the time of inclusion assessed using the Montgomery-Åsberg Depression Rating Scale (MÅDRS, Montgomery & Åsberg, 1979) with a score of ≥18; 2) medication-free for ≥2 weeks (≥4 weeks for fluoxetine) and not receiving any psychological intervention at the time of scanning, 3) right handedness (assessed using the Edinburgh Handedness Scale) (21), and 4) not meeting criteria for substance/alcohol misuse disorder (participants were excluded if they reported any illicit substance use in the previous two months). Any unstable medical condition, medication that could affect safety, study results, analyses or interpretation were excluded. All participants were screened for MRI safety.

Depressive symptom severity was assessed using the MÅDRS, and the total score was used in correlational analyses. Historical hypomanic symptoms were assessed using the 33-item hypomania checklist (HCL), a self-report measure that has been shown to reliably detect symptoms of hypomania (13,22). The questionnaire features a range of symptoms present during “high” states (e.g. “I talk more”, “I need less sleep”) and the overall score is calculated as total number of items that the subject endorses. This questionnaire was chosen as it is sensitive to variation in hypomanic symptoms and provides a continuous measure of bipolarity adjunct to categorical diagnosis. The total scale score was used in correlational analysis in this study.

The research was approved by the relevant local ethics committee and informed consent was obtained from each participant. All participants were compensated for taking part in the research.

Structural imaging

High resolution T1-weighted structural images were acquired on one of two identical GE MR750 3 Tesla scanners at the same site, using identical sequences (TR=7.31ms, TE=3.02ms, 256 x 256 matrix, 196 slices, voxel size = 1.2 x 1.05 x 1.05mm).

Voxel-based morphometry

Structural images were pre-processed using voxel-based morphometry in SPM12 *(www.fil.ion.ucl.ac.uk/spm).* Images were segmented into grey matter, white matter, and cerebrospinal fluid and processed with DARTEL (23) before being normalised to MNI space. The modulated grey matter images were then smoothed with an 8mm FWHM Gaussian kernel and the resulting images were used as input for further analyses. Data were Z-scored prior to further machine learning analysis.

Machine learning

To identify distributed patterns of grey matter volume alteration that predict depressive and hypomanic symptoms, we used support vector regression (24) with a linear kernel implemented in the Scikit-learn module for Python (25). This is an extension of the support vector machine classification method that aims to predict a continuous target variable, in our case symptom severity scores, by finding a linear hyperplane with minimal distance from the observed data points, subject to penalisation of residuals greater than ε by constant *C* to limit overfitting. This can be used to predict target values associated with previously unseen data. Predictive accuracy was assessed using leave-one-out cross validation. This involves training the model on all but one of the observations and testing its predictive accuracy on the withheld sample, repeating this process for every possible combination of training and testing data. We assessed the statistical significance of the results using permutation testing with 1000 permutations, whereby the model was repeatedly evaluated on data with target MÅDRS or HCL scores randomly reshuffled to provide an empirically derived null distribution which is used to determine the probability of the observed predictive accuracy occurring under the null

hypothesis. The value of parameters *C* and ε were set using random search (26). Prediction accuracy was quantified using mean-squared error (MSE), however we also report mean absolute error for ease of interpretation.

Reducing the dimensionality of input data, a process known as feature selection, is crucial to ensure accurate predictions when using machine learning methods (27). We focused on brain regions shown to either differentiate unipolar from bipolar disorder, or be commonly affected in both disorders, based on our previous meta-analysis (7). Regions differentiating the two disorders included left hippocampus, right middle temporal gyrus, right middle frontal gyrus, cerebellar vermis, and left inferior parietal lobule (7), and these were used to create a region of interest mask that was used to predict hypomanic symptoms. The regions common to the two disorders included the ventromedial prefrontal cortex, anterior cingulate cortex, and insula, and these areas were used to predict depressive symptom severity in the present study. Thus feature selection was based on independent data and does not result in "double dipping" (27). Feature weights from the support vector regression were extracted and plotted to identify the voxels contributing most highly to the prediction.

Univariate analyses

To compare the performance of the support vector regression model in detecting regions of grey matter variation predicting depressive and hypomanic symptoms against traditional univariate methods, linear regression analyses in SPM were used to identify regions of grey matter alteration that correlated with MÅDRS and HCL scores, covarying for intracranial volume and scanner. As with the support vector regression, statistical analyses used a region of interest approach, focusing on regions identified in our previous meta-analysis (7). Statistical maps were thresholded with a voxelwise threshold of $< .001$ uncorrected and a cluster-wise significance threshold of $< .05$ FDR corrected.

	Unipolar Depression	**Bipolar Depression**	***p***
N	39	BP-I = 2; BP-II = 6	NA
Age, years	30.67 (8.71)	29.50 (6.21)	.58
Male/Female, n	9/30	3/5	.69
MÅDRS Score	28.82 (6.08)	25.62 (5.34)	< .001
YMRS Score†	NA	2.33 (3.0)	NA
HCL Score	18.10 (5.39)	26.75 (3.01)	.01
Illness Duration, years	9.39 (7.96)	9.75 (5.44)	.88
Age of onset, years	21.28 (8.63)	19.75 (3.45)	.41
Number of mood episodes	2 (3)	5 (9.5)	.009
Years of education	14.78 (2.70)	15.43 (3.60)	.058
Full time employment, %	50	35	0.02
Comorbid Diagnoses, n	GAD=9; SAD=5, OCD=4; PD=2 PTSD=2	None	NA

Table 1: Sample characteristics. Values are reported as mean (SD) except where indicated otherwise. Number of episodes are reported as median (interquartile range). MÅDRS: Montgomery-Åsberg Depression Rating Scale, YMRS: Young Mania Rating Scale, GAD: Generalized Anxiety Disorder, SAD: Social Anxiety Disorder, OCD: Obsessive Compulsive Disorder, PD: Panic Disorder, PTSD: Post-Traumatic Stress Disorder, BP-I: Bipolar Disorder Type I, BP-II: Bipolar Disorder Type II. †YMRS scores were recorded only for bipolar disorder subjects.

Results

Forty-seven currently depressed patients were recruited. Subjects were diagnosed with either unipolar depression or bipolar disorder (type I or II), and no subjects met criteria for bipolar disorder-not otherwise specified. The study was conducted between December 2013 and June 2016. Fifty-one patients that expressed an interest in participating in the study (see Wise et al., 2016 for details) gave consent and were enrolled into the study. Of these 2 did not complete the study (one participant did not complete the MRI scan while one withdrew consent). Two subjects were excluded from these analyses due to missing questionnaire data.

Demographic and clinical details for both groups are shown in Table 1. No subjects reported a history of psychotic illness outside the context of manic episodes. There were no significant clinical differences in participant characteristics between the two scanners used (HCL scores: $t = 0.04$, $p = 0.97$, MÅDRS scores: $t = 0.11$, $p = 0.91$). Age was not correlated with grey matter volume within our regions of interest, as assessed by univariate regression (no clusters with $p > .05$ FDR corrected).

Grey matter volume and symptom severity

Predictive accuracies are shown in Figure 1. We were able to predict depression severity with a mean squared error of 30.59 (mean absolute error = 4.38, Figure 2), which was shown to be significant using permutation tests ($p = 0.006$, permutation test). When examining feature weights from the support vector regression, the highest weights were observed in the bilateral insula, suggesting that grey matter volume in this region is most strongly predictive of symptom severity. The direction of these weights was particularly consistent in the anterior part of the insula (Figure 2). Support vector regression analysis was able to predict individual hypomania scores with a mean-squared error of 35.50 (mean absolute error = -4.36), which did not survive permutation testing ($p = 0.18$). Univariate analyses did not detect any clusters exhibiting significant correlations with HCL scores or MÅDRS scores.

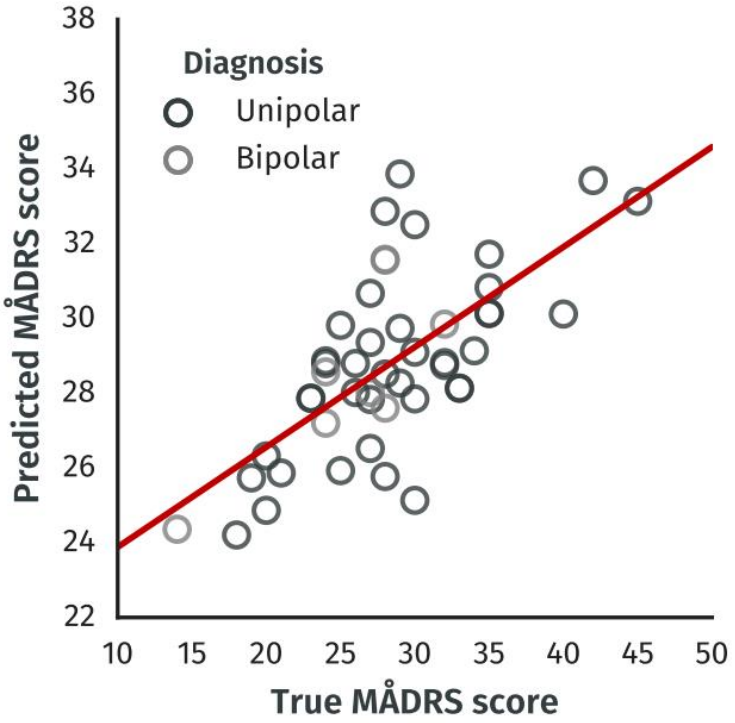

Figure 1: Relationship between true MÅDRS values and values predicted by the model. MÅDRS: Montgomery-Åsberg Depression Rating Scale.

Relationship between symptoms and predictive accuracy

In order to confirm the specificity of predictive accuracy to the symptoms in question, we investigated whether the volume in the regions associated with depression predicted hypomanic symptoms and vice versa. As expected, volume in regions thought to be associated with depression did not predict hypomanic symptoms ($p = .22$, mean squared error = 35.48, mean absolute error = 4.29), and volume in regions thought to be associated with hypomania did not predict depressive symptoms ($p = .67$, mean squared error = 35.70, mean absolute error = 4.41).

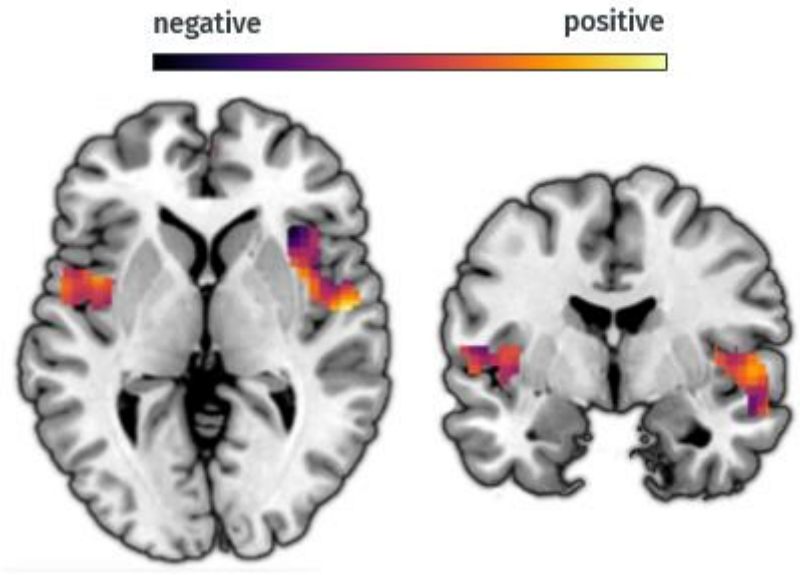

Figure 2: Support vector regression weights, focusing on the bilateral insula. Positive weights indicate that greater grey volume predicts higher MÅDRS scores, while negative weights predict the inverse. MÅDRS: Montgomery-Åsberg Depression Rating Scale.

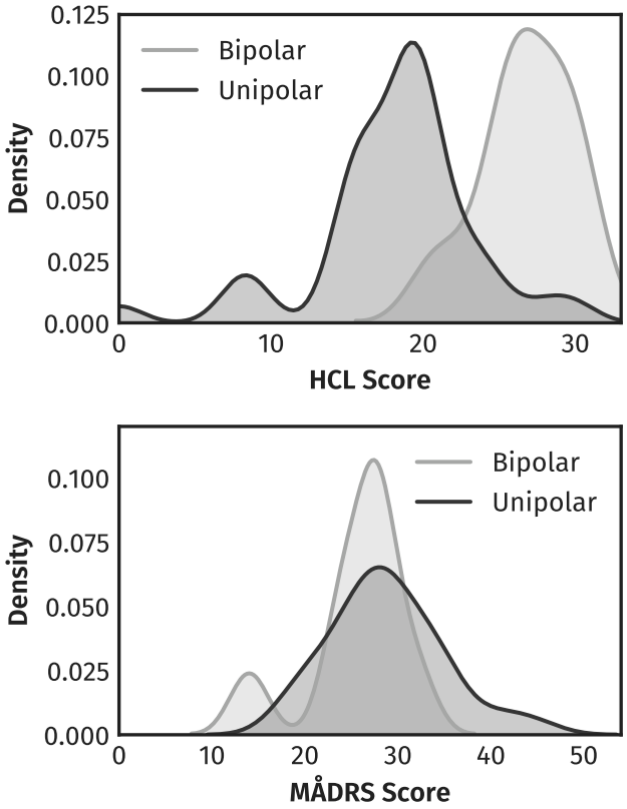

Figure 3: Distributions of outcome measures. HCL: Hypomania Checklist, MÅDRS = Montgomery Åsberg Depression Rating Scale.

Relationship between demographic and clinical variables

We also explored relationships between demographic and clinical variables and HCL scores that could potentially affect the interpretation of our results. There was no significant correlation between age and HCL scores ($r(45) = -.02$, $p = .83$) or MÅDRS scores ($r(45) = -.05$, $p = .64$), and there was no significant difference between males and females in HCL scores ($t(45) = 1.59$, $p = .12$, $d = 0.40$) or MÅDRS scores ($t(45) = 0.88$, $p = .38$, $d = 0.22$). There was also no significant correlation between illness duration and HCL scores ($r(45) = -.19$, $p = .20$) or MÅDRS scores ($r(45) = .13$, $p = .39$), or between the number of mood episodes experienced and HCL ($\rho(45) = 0.12$, $p = 0.42$) or MÅDRS ($\rho(45) = 0.04$, 0.79). HCL scores and MÅDRS scores were positively correlated ($r(45) = .47$, $p < .001$). To ensure that our results were not influenced by confounding effects of comorbidities in some subjects, we compared symptom severity scores between subjects with and without comorbidity. This demonstrated no significant differences for either MÅDRS ($t(17.36) = 0.39$, $p = 0.70$) or HCL ($t(18.12) = 1.68$, $p = 0.11$) scores.

Differences between scanners

As our sample were scanned on two identical scanners, we also sought to determine any potential influence of using two physical locations on our results. As it is not possible to include such covariates in the support vector regression, we used a support vector machine cluster to examine differences in patterns of grey matter volume between scanners, using grey matter volume in our regions of interest to predict which scanner the subject was scanned on. This classification was significant in the regions we hypothesised to be related to bipolarity ($p = 0.023$, permutation test) but not in the regions we showed to predict depressive symptom severity ($p = 0.10$, permutation test) suggesting that patterns of grey matter volume are unlikely to be the source of our positive results, but may have influenced our analysis of hypomanic symptom severity and could potentially explain the negative result here.

Discussion

In this study, we investigated whether patterns of grey matter loss in the prefrontal cortex and bilateral insula found in our meta-analysis in both unipolar and bipolar disorders represent specific substrates of depressive symptoms. We also evaluated the significance of self-reported bipolar symptoms in differentiating grey matter loss in in the left hippocampus, right middle temporal gyrus and right dorsolateral prefrontal cortex. We showed that volume in regions previously shown to be commonly reduced in unipolar and bipolar disorders predicts depressive symptom severity, suggesting for the first time that volumetric alteration in these regions might represent a substrate of depressive symptoms across both affective disorders. In contrast, we found no reliable evidence that volume in regions shown to be differentially affected in unipolar and bipolar disorders were associated with self-reported measures of mood elevation.

The most highly weighted voxels in the prediction of depressive symptoms were located in the insula bilaterally. However, these weights ranged from positive to negative, indicating that a pattern of both elevated and reduced grey matter volume in the insula predicts depressive symptom severity. Strongest negative weights were present in the anterior and ventral portions of our insula regions of interests, whereas positive weights were observed in the posterior section. Other regions included in our model, such as the medial prefrontal cortex, had relatively weak weights, although this does not imply that they did not contribute to accurate prediction. The insula is known to play a crucial role in awareness of bodily states, and its role in depressive symptomatology may relate to aberrant interoceptive processing (28–30). The whole insula has recently been shown to be reduced in volume across other psychiatric disorders (12), suggesting the importance of evaluating grey matter volume sub-fields in future studies in the context of structural and functional networks to improve diagnostic specificity.

Our results show that a specific pattern of grey matter volume alteration, particularly in the insula, predicts depressive symptom severity, providing evidence for state effects in this specific sub-field. Notably, activity in this region has also been shown to predict response to both psychological and pharmacological treatments in depression (31), and our results provide further evidence of its association with depression symptomatology. Interestingly, predictive weights in the medial prefrontal cortex were low, suggesting that grey matter volume in this region is not related to current symptom severity. A potential explanation for this is that volumetric reductions in this region are in fact a result of life stress rather than depression *per se* (32).

Interestingly, we observed a gradient of negative to positive weights from the anterior to posterior sections of the insula. The anterior insula is known to play a key role in emotional awareness, in conjunction with prefrontal regions to which it is connected, while the posterior anterior functions as a low-level primary interoceptive sensory area, creating a gradient from simple representations of bodily states to a complex representation of emotional states. Speculatively, our results may reflect a shift from bottom-up to top-down processing of sensory information about bodily states in depression, however this proposal would require direct testing. It is also interesting to observe positive weights, suggesting increased volume associated with higher symptom severity, in this region as this contradicts previous work showing reduced volume in this region in patients versus controls (7). This may indicate that differences shown between patients do not necessarily reflect symptom severity; it is possible that while individuals with depression exhibit reduced grey matter volume on average relative to healthy individuals, volume within this group correlates positively with symptom severity.

These findings add to previous work investigating relationships between brain volume and depressive symptoms. Recent work examining subcortical volume and cortical thickness has failed to find any correlation with symptom severity (33,34), which raises concerns over the specificity of volumetric alterations to depression, especially in the context of research showing that volumetric alteration in several regions appears to be present across disorders (12).

However, these approaches are unable to account for spatially distributed patterns of volumetric change, unlike the multivariate methods employed here. Our results provide evidence that volumetric changes observed in case-control comparisons within unipolar and bipolar depression are indeed associated with symptoms of depression, suggesting that these do not represent a putative general factor underlying all psychopathology (12).

Our null results in relation to the prediction of bipolarity add to the only previous study which used functional MRI to show that variation in neural systems responsible for affective generation and regulation may be affected by the presence of bipolar symptoms within depressive disorders (35). This might indicate that self-rated symptoms might not be the most effective way to identify morphometric changes in our network, that neuropathology of bipolarity might be best characterised by functional networks or perhaps that the regions that distinguished unipolar and bipolar disorders in our previous meta-analysis are not sufficiently generalizable across samples. Having recruited currently depressed participants, results might also suggest a possible fluctuation in morphometric changes influenced by mood state at the time of scanning. However, it is important to note that our sample had relatively low hypomania scores due to the small number of severe bipolar cases (i.e. bipolar type I), which is likely to have limited our ability to detect such relationships, and as such this result should be interpreted with caution. Including larger numbers of bipolar type I subjects who have experienced more substantial mood elevation in the past could allow future studies to shed light on this issue. Future studies could also consider including a sample of individuals with bipolar disorder scanned in a hypomanic state to better discriminate the significance of mood state in the definition of this brain network in relation to objective rating scales such as the Young Mania Rating Scale.

This study has a number of strengths. Firstly, our sample consisted entirely of individuals free from psychotropic medication. Medications used to treat affective disorders can have effects on grey matter volume (36,37), and our medication-free sample limits the likelihood that such effects are responsible for the results. We excluded individuals who met criteria for

substance/alcohol misuse or dependence. We also corroborated clinical diagnoses by reviewing medical notes and by gathering available collateral information to exclude axis II comorbidity as far as possible. Additionally, despite our sample being scanned on two scanners, we showed no difference in patterns of grey matter volume between scanners in the regions driving our prediction of depressive symptom severity suggesting that this was unlikely to be driving these results. However, we did find differences between scanners in regions hypothesised to be associated with hypomanic symptom severity, and it is possible that this issue may have limited our ability to detect relationships with symptom severity in these areas.

Another important strength of this study is our use of a machine learning method that offers a number of advantages over traditional univariate voxel-based analysis, including the ability to account for distributed patterns of variation in grey matter volume and to predict bipolarity on an individual basis. As a result, the null results observed in our univariate analyses are likely due to the inability of traditional univariate statistics to identify more complex patterns of grey matter volume variation that predict depressive symptom severity. Notably, this pattern of results implies that the relationship between grey matter volume and symptom severity in these regions is a complex one.

However, it should be noted that this study does have a number of limitations. Firstly, we chose to base our feature section for the analysis examining depressive symptoms on regions shown in our previous work to identify unipolar and bipolar disorders. Hence, it is possible that although we were able to confirm morphometric changes in this network, there may be other regions that do accurately predict symptoms. It is difficult to perform data-driven feature selection with the sample size used here, as defining features on the same data that is used to test predictive accuracy inevitably biases the success of the analysis due to the circular nature of this procedure. In view of this concern, we chose to use a truly independently defined set of features to avoid an inflated false positive rate. Studies using larger samples would be beneficial to enable data-driven feature selection on independent subsets of available data.

Additionally, it is important to note that our sample size was relatively modest, and this may have limited our ability to detect significant predictive regions of volumetric variation. Future research in larger samples may be more successful in this respect. However it is unlikely that this would have led to higher accuracy than detected in cross validation, which typically tends to show exaggerated accuracy (38). Furthermore, although we included patients with a wide range of hypomanic symptoms to represent the entire bipolar severity spectrum, including those diagnosed with bipolar I/II disorder, few patients had experienced manic episodes and our sample was dominated by individuals with unipolar depression. Un-medicated individuals with bipolar depression experiencing a depressive episode are challenging to recruit, making this a difficult task (17). We were however also interested in predicting levels of depressive symptoms and bipolarity in individuals presenting with a major depressive episode in view of the well-described existence of sub-threshold hypomanic symptoms in subjects presenting with symptoms of depression. In this group, not uncommonly individuals report features consistent with unipolar depression at the time of clinical contact but are likely characterised by a bipolar diathesis, have an increased chance of presenting as bipolar disorder at some point during the natural history of their condition, and respond less favourably to antidepressant treatment. As a result our coverage of the full bipolar spectrum could be improved, and it would be helpful to include subjects with a history of severe manic episodes and subgroups in different mood states. We also recruited relatively few subjects with very low scores on the HCL, and it may be that the few subjects reporting low scores on the HCL limited the training success of the model at the low end of the spectrum.

It should also be noted that we carefully recorded medication free status in our inclusion criteria. Detailed historical information could not be obtained however beyond the criteria set for inclusion and at least theoretically, medication exposure including lifetime exposure could have influenced our results. Other limitations included the presence of comorbidities in some of the depressed patients although comparing symptom severity scores between subjects with and without comorbidity did not change the results.

Finally, measuring historic bipolar symptoms with self-report measures is open to reporting and recall biases. In the absence of direct observation of psychopathology, the HCL remains nevertheless one of the most reliable self-report or clinician-rated measures for historic symptoms of bipolarity (13). Importantly, a high score on the HCL does not necessarily indicate a diagnosis of bipolar disorder and it was not our intention to use the HCL as a diagnostic tool. Overall, the HCL has been shown to be sensitive to low-level hypomanic symptoms and to be a clinically meaningful dimension within unipolar depression (39–41). Although it was not our primary aim, our results indicate that further research is warranted to evaluate the clinical relevance of the predictive value of high HCL scores at brain level in longitudinal studies. In this context, repeated clinical assessments would be highly valuable not only to confirm affective diagnoses but also reduce potential contamination from axis II disorders which are difficult to fully exclude in studies with a cross sectional design.

Results from this report suggest that structural alterations in the brain are associated with depressive but not bipolar symptoms. Further research could validate this model by assessing its specificity to current depressive symptoms. For example, if the model is truly predicting current depressive symptom severity rather than a cumulative effect of lifetime depressive episodes, it should predict low levels of symptoms in remitted patient groups. Additionally, it is likely that there are other systems that may play a role in both depressive and hypomanic symptoms. There is therefore a clear need for further research into other aspects of other neurobiological parameters that may be associated with depression bipolarity, such as inflammation (42), stress (43), and fast acting neurotransmission (44). Finally, the degree to which such biological parameters can provide prognostic value in terms of stratifying treatment choice and clinical response is an area for future trials.

In conclusion, in this work we identified a pattern of volumetric change that predicted depressive symptom severity across unipolar and bipolar disorders, but were unable to do so for hypomanic symptom severity. This suggests that depressive symptoms are associated

with a specific pattern of grey matter volume alteration, but that further work is required to validate the neurobiological relationship with these symptoms.

Declaration of interest

AJC has in the last three years received honoraria for speaking from Astra Zeneca (AZ), honoraria for consulting from Allergan and Livanova and research grant support from Lundbeck. AHY has given paid lectures and sits on advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders. AMP is supported by Bionomics Limited. DA has received travel grants from Janssen-Cilag and Servier. SCRW has received grant funding from the Medical Research Council (UK); Wellcome Trust (UK); National Institute for Health Research (UK) and support for investigator led studies from Takeda, Pfizer, Lundbeck, P1Vital, Roche and Eli Lilly. No other disclosures were reported.

Acknowledgements

This research was funded by Academy of Medical Sciences grant AMS-SGCL8 to DA, a National Institute of Health Research (NIHR) PhD studentship to TW supervised by AJC and DA, a Medical Research Council (MRC) / IoPPN Excellence PhD studentship to LM, supervised by AJC and AMP, and departmental funds generated by AJC and SCRW. DA, AJC, AMP, SCRW, TW and AHY have received support from the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. TW receives funding support from the Wellcome Trust. AHV was supported by a Chilean Bicentennial Fund Scholarship from the Bicentennial Fund for Human Capital Development (Becas Chile) and by the Psychiatric Research Trust. This study represents independent research part funded by the NIHR/Wellcome Trust, King's Clinical Research Facility and the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, MRC, Mental Health Research Network (MHRN), or the Department of Health. AHV was supported by a Chilean Bicentennial Fund Scholarship from the Bicentennial Fund for Human Capital Development (Becas Chile) and by the Psychiatric

Research Trust. The authors would like to thank the staff of the NIHR/Wellcome Trust Clinical Research Facility at King's College Hospital, the Centre for Neuroimaging Science at King's College London and Dean Broadhurst, and Michael Kelly of the MHRN for their support in the conduct of the study. We thank all participants of this study for their support. The funders had no role in design and conduct of the study, collection, management, analysis, data interpretation, preparation, review, approval of the manuscript and decision to submit for publication.

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). J Am Med Assoc. 2003 Jun 18;289(23):3095–105.

2. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RMA, Petukhova M, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry. 2007 May;64(5):543–52.

3. Wise T, Cleare AJ, Herane A, Young AH, Arnone D. Diagnostic and therapeutic utility of neuroimaging in depression: an overview. Neuropsychiatr Dis Treat. 2014;10:1509–22.

4. Arnone D, Job D, Selvaraj S, Abe O, Amico F, Cheng Y, et al. Computational meta-analysis of statistical parametric maps in major depression. Hum Brain Mapp [Internet]. 2016 Feb 1 [cited 2016 Feb 8]; Available from: http://onlinelibrary.wiley.com/doi/10.1002/hbm.23108/abstract

5. Kempton MJ, Salvador Z, Munafò MR, Geddes JR, Simmons A, Frangou S, et al. Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. Arch Gen Psychiatry. 2011 Jul;68(7):675–90.

6. Selvaraj S, Arnone D, Job D, Stanfield A, Farrow TF, Nugent AC, et al. Grey matter differences in bipolar disorder: a meta-analysis of voxel-based morphometry studies. Bipolar Disord. 2012;14(2):135–145.

7. Wise T, Radua J, Via E, Cardoner N, Abe O, Adams TM, et al. Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. Mol Psychiatry. 2017 Oct;22(10):1455.

8. Akiskal HS, Bourgeois ML, Angst J, Post R, Möller H-J, Hirschfeld R. Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders. J Affect Disord. 2000 Sep;59, Supplement 1:S5–30.

9. Angst J. The bipolar spectrum. Br J Psychiatry. 2007 Mar 1;190(3):189–91.

10. Cassano GB, Rucci P, Frank E, Fagiolini A, Dell'Osso L, Shear MK, et al. The Mood Spectrum in Unipolar and Bipolar Disorder: Arguments for a Unitary Approach. Am J Psychiatry. 2004 Jul 1;161(7):1264–9.

11. Woo YS, Shim IH, Wang H-R, Song HR, Jun T-Y, Bahk W-M. A diagnosis of bipolar spectrum disorder predicts diagnostic conversion from unipolar depression to bipolar disorder: a 5-year retrospective study. J Affect Disord. 2015 Mar 15;174:83–8.

12. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a Common Neurobiological Substrate for Mental Illness. JAMA Psychiatry. 2015 Apr 1;72(4):305–15.

13. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: Towards a self-assessment tool for hypomanic symptoms in outpatients. J Affect Disord. 2005 Oct;88(2):217–33.

14. Ashburner J, Friston KJ. Voxel-Based Morphometry—The Methods. NeuroImage. 2000 Jun;11(6):805–21.

15. Chang M, Womer FY, Edmiston EK, Bai C, Zhou Q, Jiang X, et al. Neurobiological Commonalities and Distinctions Among Three Major Psychiatric Diagnostic Categories: A Structural MRI Study. Schizophr Bull. 2018 Jan 13;44(1):65–74.

16. Cai Y, Liu J, Zhang L, Liao M, Zhang Y, Wang L, et al. Grey matter volume abnormalities in patients with bipolar I depressive disorder and unipolar depressive disorder: a voxel-based morphometry study. Neurosci Bull. 2014 Dec 11;1–9.

17. Wise T, Arnone D, Marwood L, Zahn R, Lythe KE, Young AH. Recruiting for research studies using online public advertisements: examples from research in affective disorders. Neuropsychiatr Dis Treat. 2016;12:279–85.

18. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. 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. 1998;59 Suppl 20:22-33;quiz 34-57.

19. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 1978 Nov 1;133(5):429–35.

20. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979 Apr 1;134(4):382–9.

21. Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia. 1971 Mar;9(1):97–113.

22. Feng Y, Xiang Y-T, Huang W, Wang G, Feng L, Tian T-F, et al. The 33-item Hypomania Checklist (HCL-33): A new self-completed screening instrument for bipolar disorder. J Affect Disord. 2016 Jan 15;190:214–20.

23. Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage. 2007 Oct 15;38(1):95–113.

24. Drucker H, Burges C, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Mozer M, Jordan M, Petsche T, editors. Advances in Neural Information Processing Systems. MIT Press, Cambridge; 1997.

25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011 Oct;12:2825−2830.

26. Bergstra J, Bengio Y. Random Search for Hyper-parameter Optimization. J Mach Learn Res. 2012 Feb;13(1):281–305.

27. Mwangi B, Tian TS, Soares JC. A Review of Feature Reduction Techniques in Neuroimaging. Neuroinformatics. 2013 Sep 8;12(2):229–44.

28. Avery JA, Drevets WC, Moseman SE, Bodurka J, Barcalow JC, Simmons WK. Major Depressive Disorder Is Associated With Abnormal Interoceptive Activity and Functional Connectivity in the Insula. Biol Psychiatry. 2014 Aug 1;76(3):258–66.

29. Simmons WK, Burrows K, Avery JA, Kerr KL, Bodurka J, Savage CR, et al. Depression-Related Increases and Decreases in Appetite: Dissociable Patterns of Aberrant Activity in Reward and Interoceptive Neurocircuitry. Am J Psychiatry. 2016 Jan 22;173(4):418–28.

30. Barrett LF, Quigley KS, Hamilton P. An active inference theory of allostasis and interoception in depression. Phil Trans R Soc B. 2016 Nov 19;371(1708):20160011.

31. McGrath CL, Kelley ME, Holtzheimer PE, Dunlop BW, Craighead WE, Franco AR, et al. Toward a neuroimaging treatment selection biomarker for major depressive disorder. JAMA Psychiatry. 2013 Aug;70(8):821–9.

32. Ansell EB, Rando K, Tuit K, Guarnaccia J, Sinha R. Cumulative Adversity and Smaller Gray Matter Volume in Medial Prefrontal, Anterior Cingulate, and Insula Regions. Biol Psychiatry. 2012 Jul 1;72(1):57–64.

33. Schmaal L, Veltman DJ, van Erp TGM, Sämann PG, Frodl T, Jahanshad N, et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. Mol Psychiatry [Internet]. 2015 Jun 30 [cited 2015 Sep 16]; Available from: http://www.nature.com/mp/journal/vaop/ncurrent/full/mp201569a.html

34. Schmaal L, Hibar DP, Sämann PG, Hall GB, Baune BT, Jahanshad N, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Mol Psychiatry [Internet]. 2016 May 3 [cited 2017 May 11]; Available from: http://www.nature.com/mp/journal/vaop/ncurrent/full/mp201660a.html

35. Fournier JC, Keener MT, Mullin BC, Hafeman DM, LaBarbara EJ, Stiffler RS, et al. Heterogeneity of amygdala response in major depressive disorder: the impact of lifetime subthreshold mania. Psychol Med. 2013;43(02):293–302.

36. Arnone D, McKie S, Elliott R, Juhasz G, Thomas EJ, Downey D, et al. State-dependent changes in hippocampal grey matter in depression. Mol Psychiatry. 2013 Dec;18(12):1265–72.

37. Hajek T, Cullis J, Novak T, Kopecek M, Höschl C, Blagdon R, et al. Hippocampal volumes in bipolar disorders: opposing effects of illness burden and lithium treatment. Bipolar Disord. 2012;14(3):261–270.

38. Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods [Internet]. [cited 2015 Jan 19]; Available from: http://www.sciencedirect.com/science/article/pii/S0165027015000114

39. Rybakowski JK, Dudek D, Pawlowski T, Lojko D, Siwek M, Kiejna A. Use of the Hypomania Checklist-32 and the Mood Disorder Questionnaire for detecting bipolarity in 1,051 patients with major depressive disorder. Eur Psychiatry. 2012 Nov;27(8):577–81.

40. Fornaro M, Ventriglio A, De Pasquale C, Pistorio ML, De Berardis D, Cattaneo CI, et al. Sensation seeking in major depressive patients: Relationship to sub-threshold bipolarity and cyclothymic temperament. J Affect Disord. 2013 Jun;148(2–3):375–83.

41. Hu C, Xiang Y-T, Ungvari GS, Dickerson FB, Kilbourne AM, Si T-M, et al. Undiagnosed bipolar disorder in patients treated for major depression in China. J Affect Disord. 2012 Oct;140(2):181–6.

42. Strawbridge R, Arnone D, Danese A, Papadopoulos A, Herane Vives A, Cleare AJ. Inflammation and clinical response to treatment in depression: A meta-analysis. Eur Neuropsychopharmacol. 2015 Oct;25(10):1532–43.

43. Herane Vives A, De Angel V, Papadopoulos A, Strawbridge R, Wise T, Young AH, et al. The relationship between cortisol, stress and psychiatric illness: New insights using hair analysis. J Psychiatr Res. 2015 Nov 1;70(Supplement C):38–49.

44. Arnone D, Mumuni AN, Jauhar S, Condon B, Cavanagh J. Indirect evidence of selective glial involvement in glutamate-based mechanisms of mood regulation in depression: Meta-analysis of absolute prefrontal neuro-metabolic concentrations. Eur Neuropsychopharmacol. 2015 Aug;25(8):1109–17.