Archival Report

Biological Psychiatry: CNNI

Individualized Prediction and Clinical Staging of Bipolar Disorders Using Neuroanatomical Biomarkers

Benson Mwangi, Mon-Ju Wu, Bo Cao, Ives C. Passos, Luca Lavagnino, Zafer Keser, Giovana B. Zunta-Soares, Khader M. Hasan, Flavio Kapczinski, and Jair C. Soares

ABSTRACT

BACKGROUND: Neuroanatomical abnormalities in bipolar disorder (BD) have previously been reported. However, the utility of these abnormalities in distinguishing individual patients with BD from healthy control subjects and stratifying patients based on overall illness burden has not been investigated in a large cohort.

METHODS: We examined whether structural neuroimaging scans coupled with a machine learning algorithm are able to distinguish individual patients with BD from healthy control subjects in a large cohort of 256 subjects. Additionally, we investigated the relationship between machine learning–predicted probability scores and clinical characteristics of subjects, such as illness duration and clinical stages. Neuroimaging scans were acquired from 128 patients with BD and 128 healthy control subjects. Gray and white matter density maps were obtained and used to train a relevance vector machine learning algorithm, which was used to distinguish individual patients from healthy control subjects.

RESULTS: The relevance vector machine algorithm distinguished patients from healthy control subjects with 70.3% accuracy (74.2% specificity, 66.4% sensitivity, χ^2 $p < .005$) using white matter density data and 64.9% accuracy (71.1% specificity, 58.6% sensitivity, χ^2 $p < .005$) using gray matter density. Multiple brain regions, largely covering the frontolimbic system, were identified as "most relevant" in distinguishing both groups. Patients identified by the algorithm with high certainty (a high probability score) belonged to a subgroup with >10 total lifetime manic episodes including hospitalizations (late-stage BD).

CONCLUSIONS: These results indicate the presence of widespread structural brain abnormalities in BD that are associated with higher illness burden, which points to neuroprogression.

Keywords: Big data, Bipolar disorders, Clinical staging, Machine learning, Manic episodes, Neuroimaging

http://dx.doi.org/10.1016/j.bpsc.2016.01.001

Bipolar disorder (BD) is a debilitating illness with an approximate lifetime prevalence of 4%–5% in the general population (1). In the last 2 decades, neuroimaging studies have extensively reported volumetric abnormalities in patients with BD as compared with healthy control subjects. Specifically, reductions in gray matter density have been reported in the orbitofrontal cortex, superior temporal gyrus, and insula (2–4). In addition, reductions in white matter density have been reported in the corpus callosum, cingulate gyrus, and prefrontal white matter (5,6). A high number of total lifetime manic episodes has been associated with greater reductions in gray matter density in the prefrontal brain regions, cerebellum, and ventricular systems (7–10). However, although these studies have undoubtedly offered significant insights into neuroanatomical abnormalities of BD, subsequent findings have not been translated into objective and clinically useful biomarkers. A significant first step in realizing this goal is the ability to use neuroimaging scans and associated clinical measurements to objectively distinguish individual patients with BD from healthy control subjects as well as discern clinically relevant biological pathways and brain circuitries as hypothesized elsewhere (11–14).

Machine learning algorithms are ideal computational solutions with the ability to contribute to the search of much needed objective biomarkers for the following three reasons. First, machine learning algorithms allow predictions at an individual subject level and therefore are able to facilitate individualized clinical decisions (12,14). Second, these algorithms are largely multivariate and therefore are able to analyze multiple biological measurements simultaneously, as opposed to traditional univariate statistical methods, which are able to analyze only single measurements at a time (15). Third, machine learning algorithms use robust cross-validation methods to establish generalizability of results by testing the algorithm using previously unseen observations (16–19). A detailed overview of machine learning in psychiatric neuroimaging is provided elsewhere (15–17). Other fields of psychiatric research that have benefited from machine learning

include prediction of suicide attempts (20), prediction of treatment response (21), and others as summarized elsewhere (22). However, although promising, these techniques are still unavailable for clinical use.

Several studies have enlisted machine learning algorithms to distinguish patients with BD from healthy control subjects and reported above chance (>50%) prediction accuracy. Schnack *et al.* (23) recently reported 61% prediction accuracy in discriminating patients with BD from healthy control subjects using structural neuroimaging scans from 132 subjects. In a two-cohort replication study of 80 subjects, Rocha-Rego *et al.* (24) reported prediction accuracies of 72%–73%. Finally, a recent multicenter study reported prediction accuracies acquired from two centers each with a total of 58 subjects of 62% and 87.6%, respectively, using neuroimaging scans (25). However, although these studies have made significant contributions in establishing the predictive validity of neuroimaging scans in BD, multiple research questions remain unanswered. First, predictive results (specificity and sensitivity) have not been established using large samples. Second, the utility of these computational algorithms in supporting clinically relevant applications (e.g., patient stratification or clinical staging) has not been fully explored.

The three main objectives of the present study were to establish the utility of structural T1-weighted scan data together with a machine learning algorithm in distinguishing patients with BD from healthy control subjects; to elucidate gray and white matter neuroanatomical characteristics most relevant in distinguishing patients with BD from healthy control subjects; and, as a post hoc test, to investigate the relationship between prediction outputs or probability scores of the machine learning algorithm and individual subjects' clinical stages. Subjects were assigned multiple clinical stages (healthy control, early-stage bipolar disorder type I [BDI], intermediate-stage BDI, late-stage BDI, and bipolar disorder type II [BDII]) based on the International Society for Bipolar Disorders task report on staging systems (26). We hypothesized that patients with BD predicted by the machine learning algorithm with high certainty (high probability scores) would belong to the BDI late-stage subgroup.

METHODS AND MATERIALS

Subjects

This study was approved by the local institutional review board, and written informed consent was obtained from all subjects. Study participants included 128 patients with a DSM-IV diagnosis of BD and 128 demographically matched healthy control subjects (Table 1). A diagnosis of BD was established by a research psychiatrist (JCS) using the Structured Clinical Interview for DSM-IV (27). Recruited demographically matched healthy control subjects did not have first-degree relatives with Axis I DSM-IV psychiatric disorders. Participants with a history of head trauma, neurologic disorders, and current medical condition such as active liver disease or kidney problems were excluded from the study. Current mood status of patients was evaluated using the Young Mania Rating Scale (28) and the 21-item Hamilton Depression Rating Scale (29) (Table 1).

Image Acquisition and Preprocessing

All structural neuroimaging scans were acquired using a Philips 1.5-tesla magnetic resonance imaging scanner (Philips

Table 1. Demographic and Clinical Details

	Healthy Control Subjects	BD Patients[a]	*p* Value
Age, Years	36.33 (12.25)	37.56 (11.6)	.4086[b]
Women/Total	84 (128)	92 (128)	.2807[c]
BD Type	—		—
BDI		96/128	
BDII		32/128	
HDRS	.72 (1.07)	12.92 (7.84)	< .0001[b]
YMRS	.31 (.84)	6.9 (6.8)	< .0001[b]
GAF	91.46 (6.12)	62.52 (11.63)	< .0001[b]
MADRS	.38 (1.01)	17.18 (11.1)	< .0001[b]
Currently Taking or Previously Taken Any Psychotropic Medication	—	120	—
Current Mood	—		—
Euthymic		34	
Depressed		64	
Manic		6	
Hypomanic		7	
Mixed		14	
Undetermined		3	

Values are presented as mean (SD) or *n*.

BD, bipolar disorder; BDI, bipolar disorder type I; BDII, bipolar disorder type II; GAF, Global Assessment of Functioning; HDRS, Hamilton Depression Rating Scale; MADRS, Montgomery–Åsberg Depression Rating Scale; YMRS, Young Mania Rating Scale.

[a]Comorbidities included generalized anxiety disorder ($n = 22$), obsessive-compulsive disorder ($n = 11$), panic disorder ($n = 36$), posttraumatic stress disorder ($n = 25$), specific/simple phobia ($n = 12$), social phobia ($n = 24$), agoraphobia ($n = 26$), alcohol abuse ($n = 6$), anorexia ($n = 1$), anxiety disorder ($n = 4$), binge-eating disorder ($n = 3$), bulimia ($n = 2$).

[b]Student *t* test.

[c]χ^2 test.