Archival Report

Biological Psychiatry

Detecting Neuroimaging Biomarkers for Depression: A Meta-analysis of Multivariate Pattern Recognition Studies

Joseph Kambeitz, Carlos Cabral, Matthew D. Sacchet, Ian H. Gotlib, Roland Zahn, Mauricio H. Serpa, Martin Walter, Peter Falkai, and Nikolaos Koutsouleris

ABSTRACT

BACKGROUND: Multiple studies have examined functional and structural brain alteration in patients diagnosed with major depressive disorder (MDD). The introduction of multivariate statistical methods allows investigators to utilize data concerning these brain alterations to generate diagnostic models that accurately differentiate patients with MDD from healthy control subjects (HCs). However, there is substantial heterogeneity in the reported results, the methodological approaches, and the clinical characteristics of participants in these studies.

METHODS: We conducted a meta-analysis of all studies using neuroimaging (volumetric measures derived from T1-weighted images, task-based functional magnetic resonance imaging [MRI], resting-state MRI, or diffusion tensor imaging) in combination with multivariate statistical methods to differentiate patients diagnosed with MDD from HCs.

RESULTS: Thirty-three ($k = 33$) samples including 912 patients with MDD and 894 HCs were included in the meta-analysis. Across all studies, patients with MDD were separated from HCs with 77% sensitivity and 78% specificity. Classification based on resting-state MRI (85% sensitivity, 83% specificity) and on diffusion tensor imaging data (88% sensitivity, 92% specificity) outperformed classifications based on structural MRI (70% sensitivity, 71% specificity) and task-based functional MRI (74% sensitivity, 77% specificity).

CONCLUSIONS: Our results demonstrate the high representational capacity of multivariate statistical methods to identify neuroimaging-based biomarkers of depression. Future studies are needed to elucidate whether multivariate neuroimaging analysis has the potential to generate clinically useful tools for the differential diagnosis of affective disorders and the prediction of both treatment response and functional outcome.

Keywords: Affective disorder, Classification, Diagnosis, Prediction, Sensitivity, Specificity

http://dx.doi.org/10.1016/j.biopsych.2016.10.028

Major depressive disorder (MDD) has a lifetime prevalence of 14.6%, making it one of the most common psychiatric disorders worldwide (1). Reliable diagnosis of MDD is a primary prerequisite for effective pharmacological and psychological interventions (2). Currently, the diagnosis of depression is based on the phenomenological evaluation of symptoms and behavior by trained clinicians. Scientists have posited that neuroimaging holds "diagnostic potential" given findings in multiple studies of significant anomalies in brain structure (3–5), function (6–8), and neurochemistry (9,10) in patients with depression. Even though these meta-analyses indicate that brain changes are replicable across studies, the alterations are often small and do not allow a reliable differentiation between patients and control subjects (11). Thus, neuroimaging markers are not included in clinical practice to guide decisions concerning psychiatric diagnosis (12,13). This might result from the higher costs associated with neuroimaging examinations. Moreover, most of the previous neuroimaging studies in MDD have taken a univariate approach, which has important consequences in terms of the clinical applicability of the obtained results. For example, univariate approaches neglect the highly interconnected nature of the brain and, consequently, the statistical dependency of the given units of analysis (e.g., voxels or regions of interest) (14). Moreover, even if two groups (e.g., patients with depression and healthy control subjects [HCs]) differ at a statistically significant level with respect to a target variable (e.g., hippocampal volume), there is typically substantial overlap of the two distributions, hindering reliable differentiation of depressed from nondepressed individuals.

To address these limitations, investigators have begun to apply multivariate statistical methods to the analysis of neuroimaging data (15,16). By focusing on patterns of brain changes that are distributed across multiple regions, these methods allow for the generation of statistical models with high diagnostic or predictive power. In this context, a recent meta-analysis showed that patients with schizophrenia can be accurately differentiated from healthy volunteers in 80% of the cases using only neuroimaging-based diagnostic models (17). Moreover, these methods may facilitate the development

SEE COMMENTARY ON PAGE 306

 ISSN: 0006-3223

of neuroimaging tools to distinguish among different psychiatric disorders (18–21) or to predict clinical outcomes (22–24). Indeed, multiple proof-of-concept studies have successfully used multivariate statistical methods to guide the diagnosis of depression based on structural magnetic resonance imaging (sMRI) data (19,21,25–27), resting-state functional MRI (rsfMRI) data (26,28–34), and task-based functional MRI (fMRI) data (35–39). The sensitivity and the specificity reported in these studies both range from 70% to 90%. This variable diagnostic performance may be due to methodological differences among these studies with respect to the neuroimaging data modality, preprocessing protocol, classification algorithm, or the cross-validation (CV) procedure used. In addition, these studies differ with respect to demographic and clinical characteristics of depressed patients. Differences in performance and study heterogeneity make it difficult to evaluate the potential of neuroimaging to identify diagnostic biomarkers for depression. Here, we report the results of a meta-analysis conducted on studies that used multivariate statistical methods to differentiate patients with depression from HCs. This meta-analytic approach allows us to quantify the ability of multivariate methods to identify depression-related patterns in neuroimaging data. In this way, we investigate the neurobiological construct validity of the current clinical definition of MDD.

METHODS AND MATERIALS

Search and Study Selection Strategy

We searched the electronic PubMed database from January 1, 1950, up to June 31, 2015 (see the Supplement for details). Subsequently, we screened studies according to the following criteria: To be included in the meta-analysis, a paper needed to report results of a neuroimaging-based, supervised, multivariate two-group classification model separating MDD patients from HCs. Studies were included if the following measures of classification performance were available or if data allowed for the calculation of the following parameters: true positives (TP), true negatives (TN), false positives (FP), false negatives (FN). In cases in which insufficient data were reported, the authors were asked to provide additional information regarding their published reports. The results of the literature search are presented in a flowchart following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (40) (see the Supplement and Supplemental Figure S1).

Data Extraction

The main outcome was the diagnostic accuracy of the multivariate diagnostic models when applied to patients with MDD and HCs as indicated by sensitivity [= TP / (TP + FN)] and specificity [= TN / (TN + FP)]. Additional information was extracted from the selected studies as follows: names of the authors, year of publication, demographic characteristics of HC and patient groups [group size, age, sex, medication status, symptoms as measured by the Hamilton Depression Rating Scale (HAMD) (41) or the Beck Depression Inventory (42)], neuroimaging modality (volumetric measures derived of T1-weighted MRI images sMRI, task-based fMRI, rsfMRI, positron emission tomography, single photon emission computed tomography, diffusion tensor imaging [DTI], scanner type, image resolution), characteristics of the neuroimaging preprocessing, configuration of the classification algorithm, and type of the cross-validation procedure (e.g., leave-one-out, *k*-fold cross-validation). To ensure accuracy of data extraction, two authors separately performed extraction and disagreements were resolved in a consensus conference.

Data Analysis

In the present analysis we implemented a random-effects, bivariate meta-analytical model as introduced by Reitsma *et al.* (43). Results of the meta-analysis are presented in forest plots separately for sensitivity and specificity. Summary estimates for sensitivity and specificity are provided separately for sMRI, task-based fMRI, rsfMRI, or DTI studies, and for all studies combined. The robustness of the results and the effects of potentially confounding variables (e.g., age, sex ratio, year of publication) were investigated by adding moderator variables to the bivariate regression model. Furthermore, we tested for differences between studies in the clinical variables using univariate analysis of variance. Publication bias was assessed by creating funnel plots by plotting log diagnostic odds ratios (logDORs) for all studies against $\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ with n_1 and n_2 representing the sample sizes of the patient and the HC group, respectively. This measure is proportional to the inverted square root of the effective sample size (ESS): $\frac{1}{\sqrt{ESS}}$. In case of a publication bias, the distribution of studies in the funnel plot is asymmetrical. A statistical test for funnel plot asymmetry is provided by a regression of logDOR with $\frac{1}{\sqrt{ESS}}$ weighted by ESS (44). As an exploratory analysis, we generated a multivariate regression model using the elastic net algorithm to predict logDOR of individual studies based on 23 clinical and methodological variables (see the Supplement for details). All computations were performed using the *R* statistical programming language version 3.3.1 (45) with the packages mada (46) and glmnet (47).

RESULTS

Meta-analysis

The initial literature search identified 641 studies of interest. After screening all studies and applying the inclusion criteria, 608 studies were excluded (see the Supplement and Supplemental Figure S1 for a flowchart of the literature search). The final sample consisted of 33 studies with a total of 912 patients (mean age: 34.27 years) and 894 HCs (mean age: 32.81 years). From those studies, 14 samples used sMRI (19–21,25–27,48–54), 9 samples used rsfMRI (26,29,31,33,54–58), 9 samples used fMRI (35–37,39,59–63), and 6 samples used DTI (26,64–67) to build predictive models (see the Supplement and Supplemental Table S1 for an overview of the characteristics of the studies; please note that some studies provide more than one sample). There were no studies available using single photon emission computed tomography methodology. One study reported 85% classification accuracy using $^{[18]}$fluorodeoxyglucose positron emission tomography but was excluded from further analysis due to the small number of available studies (68).