

Structural neuroimaging biomarkers for obsessive-compulsive disorder in the ENIGMA-OCD consortium: medication matters

Willem B. Bruin*, M.Sc., Luke Taylor, B.Sc., Rajat M. Thomas, Ph.D., Jonathan P Shock, Ph.D., Paul Zhutovsky, M.Sc., Yoshinari Abe, M.D., Ph.D., Pino Alonso, M.D., Ph.D., Stephanie H. Ameis, M.D., M.Sc., Alan Anticevic, Ph.D., Paul D. Arnold, M.D., Ph.D., Francesca Assogna, Ph.D., Francesco Benedetti, M.D., Jan C. Beucke, Ph.D., Premika S.W. Boedhoe, Ph.D., Irene Bollettini, Ph.D., Anushree Bose, Ph.D., Silvia Brem, Ph.D., Brian P. Brennan, M.D., M.M.Sc., Jan K Buitelaar, M.D., Ph.D., Rosa Calvo, M.D., Ph.D., Yuqi Cheng, Ph.D., Kang Ik K. Cho, Ph.D., Sara Dallspezia, M.D., Damiaan Denys, M.D., Ph.D., Benjamin A. Ely, Ph.D., Jamie D. Feusner, M.D., Kate D. Fitzgerald, M.D., Jean-Paul Fouche, Ph.D., Egill A. Fridgeirsson, M.Sc., Patricia Gruner, Ph.D., Deniz A. Gürsel, M.Sc., Tobias U. Hauser, Ph.D., Yoshiyuki Hirano, Ph.D., Marcelo Q. Hoexter, M.D., Ph.D., Hao Hu, Ph.D., Chaim Huyser, M.D., Ph.D., Iliyan Ivanov, M.D., Anthony James, MRCP, MRPsych, Fern Jaspers-Fayer, Ph.D., Norbert Kathmann, Ph.D., Christian Kaufmann, Ph.D., Kathrin Koch, Ph.D., Masaru Kuno, M.D., Ph.D., Gerd Kvale, Ph.D., Jun Soo Kwon, M.D., Ph.D., Yanni Liu, Ph.D., Christine Lochner, Ph.D., Luisa Lázaro, M.D., Ph.D., Paulo Marques, Ph.D., Rachel Marsh, Ph.D., Ignacio Martínez-Zalacáin, M.Sc., David Mataix-Cols, Ph.D., José M. Menchón, M.D., Ph.D., Luciano Minuzzi, M.D., Ph.D., Pedro S. Moreira, M.Sc., Astrid Morer, M.D., Ph.D., Pedro Morgado, M.D., Ph.D., Akiko Nakagawa, M.D., Ph.D., Takashi Nakamae, M.D., Ph.D., Tomohiro Nakao, M.D., Ph.D., Janardhanan C. Narayanaswamy, M.D., Erika L. Nurmi, M.D., Ph.D., Joseph O'Neill, Ph.D., Jose C. Pariente, M.Sc., Chris Perriello, B.Sc., John Piacentini, Ph.D., Fabrizio Piras, Ph.D., Federica Piras, Ph.D., Y.C. Janardhan Reddy, M.D., Oana G. Rus-Oswald, Ph.D., Yuki Sakai, M.D., Ph.D., João R. Sato, Ph.D., Lianne Schmaal, Ph.D., Eiji Shimizu, M.D., Ph.D., H. Blair Simpson, M.D., Ph.D., Noam Soreni, M.D., Carles Soriano-Mas, Ph.D., Gianfranco Spalletta, M.D., Ph.D., Emily R. Stern, Ph.D., Michael C. Stevens, Ph.D., S. Evelyn Stewart, M.D., Philip R. Szeszko, Ph.D., David F. Tolin, Ph.D., Ganesan Venkatasubramanian, M.D., Ph.D., Zhen Wang, M.D., Ph.D., Je-Yeon Yun, M.D., Ph.D., Daan van Rooij, Ph.D., **ENIGMA-OCD consortium****, Paul M. Thompson, Ph.D., Odile A. van den Heuvel, M.D., Ph.D., Dan J. Stein, M.D., Ph.D., Guido A. van Wingen, Ph.D.

*corresponding author. Postal address: Meibergdreef 5, 1105 AZ, Amsterdam, The Netherlands; email: willembenjamin@gmail.com; telephone: +31208913752

** [see excel sheet for affiliations and full consortium member authors](#)

Abstract:

Objective: No diagnostic biomarkers are available for obsessive-compulsive disorder (OCD). Magnetic resonance imaging (MRI) studies have provided evidence for structural abnormalities in distinct brain regions, but effect sizes are small and have limited clinical relevance. To investigate whether individual patients can be distinguished from healthy controls, we performed multivariate analysis of structural neuroimaging data from the ENIGMA-OCD consortium. **Method:** We included 46 data sets with neuroimaging and clinical data from adult (≥ 18 years) and pediatric (< 18 years) samples. T_1 images from 2,304 OCD patients and 2,068 healthy controls were analyzed using standardized processing to extract regional measures of cortical thickness, surface area and subcortical volume. Machine learning classification performance was tested using cross-validation, and possible effects of clinical variables were investigated by stratification. **Results:** Classification performance for OCD versus controls using the complete sample with different classifiers and cross-validation strategies was poor (AUC—0.57 (standard deviation (SD)=0.02; $P_{\text{corr}}=0.19$) to 0.62 (SD=0.03; $P_{\text{corr}}<.001$)). When models were validated on completely new data from other sites, model performance did not exceed chance-level (AUC—0.51 (SD=0.11; $P_{\text{corr}}>.99$) to 0.54 (SD=0.08; $P_{\text{corr}}>.99$)). In contrast, good classification performance (>0.8 AUC) was achieved within subgroups of patients split according to their medication status. **Conclusions:** Parcellated structural MRI data do not enable good distinction between patients with OCD and controls. However, classifying subgroups of patients based on medication status enables good identification at the individual subject level. This underlines the need for longitudinal studies on the short- and long-term effects of medication on brain structure.

Manuscript:

Introduction

Obsessive-compulsive disorder (OCD) is a severe and disabling condition that occurs in 2-3% of the population [1]. It is characterized by recurrent, intrusive, irrational and distressing thoughts (obsessions) and repetitive behaviors or mental acts (compulsions) [2]. So far, no biomarkers that may aid differential diagnosis are available, and diagnosis relies entirely on recognition of behavioral features assessed by clinical interview [3]. Many neuroimaging studies have provided evidence for abnormalities in cortico-striato-thalamo-cortical (CSTC) circuits, as well as distributed changes in limbic, parietal and cerebellar regions [4, 5]. This has recently been confirmed by meta- and mega- analysis of neuroimaging studies within the Enhancing Neuro-Imaging and Genetics through Meta-Analysis (ENIGMA) consortium [6–8]. However, inference was done at the group-level, and the small effect sizes that were reported precludes clinical application.

Analytic tools such as multivariate pattern analysis (MVPA) enable inference at the individual-level, which may result in better discrimination [3, 9]. MVPA techniques can be used to develop predictive models that extract common patterns from neuroimaging data to classify individuals based on their diagnosis. A major advantage of MVPA is its ability to use inter-regional correlations to detect subtle and spatially distributed effects compared to traditional methods of analysis [4]. Therefore, MVPA seems particularly suited for neuroimaging analyses in OCD, as abnormalities are typically distributed across the brain [10, 11]. Previous MVPA studies have been able to distinguish OCD patients from controls with accuracies ranging from 66-100% (reviewed in [12]). Although these results are promising, sample sizes have typically been small, limiting model performance optimization and leading to high variance in estimated accuracy and overly optimistic classification rates [13]. Additionally, most

studies have been performed using data from one research center to minimize technical (e.g., scanner hardware, protocols and diagnostic assessment) and clinical (e.g., age, medication status, disease chronicity and severity) heterogeneity. It is therefore not clear whether these results generalize to other centers, which would be required for clinical application [14–16].

Here, we used data from the ENIGMA-OCD consortium, including 4,372 participants recruited at 36 research institutes around the world, with a full range of technical and clinical heterogeneity. We assessed the ability of MVPA to distinguish OCD patients from healthy controls based using structural neuroimaging data at the individual subject level. We investigated machine learning classification performance in both single-site and multi-site samples using different validation strategies to assess generalizability, as well as effects of clinical variables, such as medication use.

Materials and methods

Study population

The ENIGMA-OCD working group includes 46 data sets from 36 international research institutes, with neuroimaging and clinical data from adult (≥ 18 years) and pediatric (< 18 years) samples. In total, we analyzed data from 4,372 participants, including 2,304 OCD patients ($n=1,801$ adult, $n=503$ child) and 2,068 healthy controls (HC; i.e., free of psychopathology; $n=1,629$ adult, $n=439$ child), with 38 of 46 datasets identical to those described in previous mega-analyses by this working group [6, 7, 17]. All participating sites obtained permission from their local institutional review boards or ethics committees to provide anonymized data for analysis, and all study participants provided written informed consent. Demographic and clinical characteristics of each cohort are detailed in supplementary **Table S1**. A complete overview of instruments used to obtain diagnosis and clinical information can be found elsewhere (Data Supplement 1, Supplementary Section S1) [7]. Diagnosis was determined in accordance with DSM [2]; MINI and SCID were used for adult samples and K-SADS, MINI-KIDS and ADIS were used for pediatric samples [18–22].

MRI processing

Structural T1-weighted brain MRI scans were acquired and processed locally at each site. Image acquisition parameters are listed elsewhere [7]. Parcellations were performed using FreeSurfer (FS) software version 5.3 (<http://surfer.nmr.mgh.harvard.edu>), following standardized ENIGMA protocols to harmonize analyses and quality control procedures across multiple sites (see

<http://enigma.usc.edu/protocols/imaging-protocols/>). Parcellations of 34 cortical (Desikan-Killiany atlas-based [23]) and 7 subcortical gray matter structures per hemisphere, lateral ventricle volumes, two whole-hemisphere measures and total intracranial volume were extracted, visually inspected and statistically evaluated for outliers (quality assurance is reported elsewhere [7]). Brain regions (features) used for classification included cortical thickness (CT), surface area (SA) and subcortical volumes of ROIs, two lateral ventricular and intra-cranial volumes (ICV), and two whole-hemisphere measures for SA and CT.

Multivariate classification and validation

Participants with >10% missing entries were excluded (N=276), and median imputation was used for missing MRI data on the training set. Continuous features were centered around median zero and scaled according to their interquartile range. FS variables were combined with covariates age, sex, and site by concatenating individual feature vectors. All analyses were performed separately for pediatric and adult patients, and both groups combined. Common MVPA classifiers were applied: support vector machine (SVM) with linear and non-linear (radial-basis-function (RBF)) kernels, logistic regression (LR) with L1 and L2 regularization, Gaussian processes classification (GPC) with a linear kernel, and two decision-tree based ensemble methods, namely the random forest classifier (RFC) and the XGBoost algorithm [24–27]. A deep neural network was also implemented (fully connected; 3 layers with 60, 40 and 20 nodes respectively). SVM and LR classifiers were combined with and without automatic dimensionality reduction via principal component analysis (PCA), using the minimal number of components explaining 90% of the variance. Hyper-parameters for SVM (linear and non-linear), LR and XGBoost were optimized using nested cross-validation; RFC and GPC were tuned following recommendations. Details on handling missing data, model implementation and hyper-parametrization can be found in Supplementary

Methods. The primary performance metric was area-under-the-receiver-operator-curve (AUC). Balanced accuracy, sensitivity and specificity are reported in supplement.

Multi-site classification of OCD patients versus HCs was assessed using different cross-validation (CV) approaches. First, we assessed multi-site classification using 10-fold CV to obtain maximally homogeneous train-test splits, with approximately the same number of subjects and the same proportion of samples coming from each site (internal validation). Next, we addressed leave-one-site-out (LOSO) CV, in which all but one site were used to train the models while the left out site was used to assess model performance (external validation). This may result in large between-sample heterogeneity of training and test sets, resulting in lower classification performance[28]. Because LOSO-CV has different fold sizes, we additionally performed 10-fold CV with LOSO-matched fold sizes, to evaluate whether differences in performance were due to differences in heterogeneity or fold size. Finally, we also performed single-site predictions using 10-fold CV to assess classification performance with reduced heterogeneity. Statistical significance of model performance was assessed directly through obtained AUC scores using the Mann-Whitney-U statistic for non-parametric testing (see supplement for details) [29]. Bonferroni-corrected level of significance was set at $\alpha=0.05$ for the number of classifiers and comparisons.

Clinical variables and sensitivity analysis

To explore the effects of clinical heterogeneity on classification performance, we selected subgroups with particular demographic and clinical characteristics: medication use, OCD severity, age of onset (AO) and duration of illness. Classifications performed were HC vs low (YBOCS \leq 24; mild-moderate[30]) and high severity (YBOCS $>$ 24; moderate-severe) OCD; HC vs early ($<$ 18yrs) and late AO (\geq 18yrs) OCD; HC vs short (\leq 7yrs) and long duration ($>$ 7yrs) OCD; and HC vs unmedicated and medicated OCD. For disease

duration and severity, median splits were used to define groups; the 18 year threshold for AO was chosen in line with prior ENIGMA-OCD mega-analyses[6, 7]. Finally, as particular clinical variables can co-occur, we performed a post-hoc sensitivity analysis to investigate the effects of potential clinical covariance for results with $AUC \geq 0.8$ (see Supplementary **Methods**).

Feature importance

To assess which brain regions and clinical variables contributed most to classification we used feature importance extracted from RFC combined with a permutation testing framework (see Supplementary **Methods**) [31].

Results

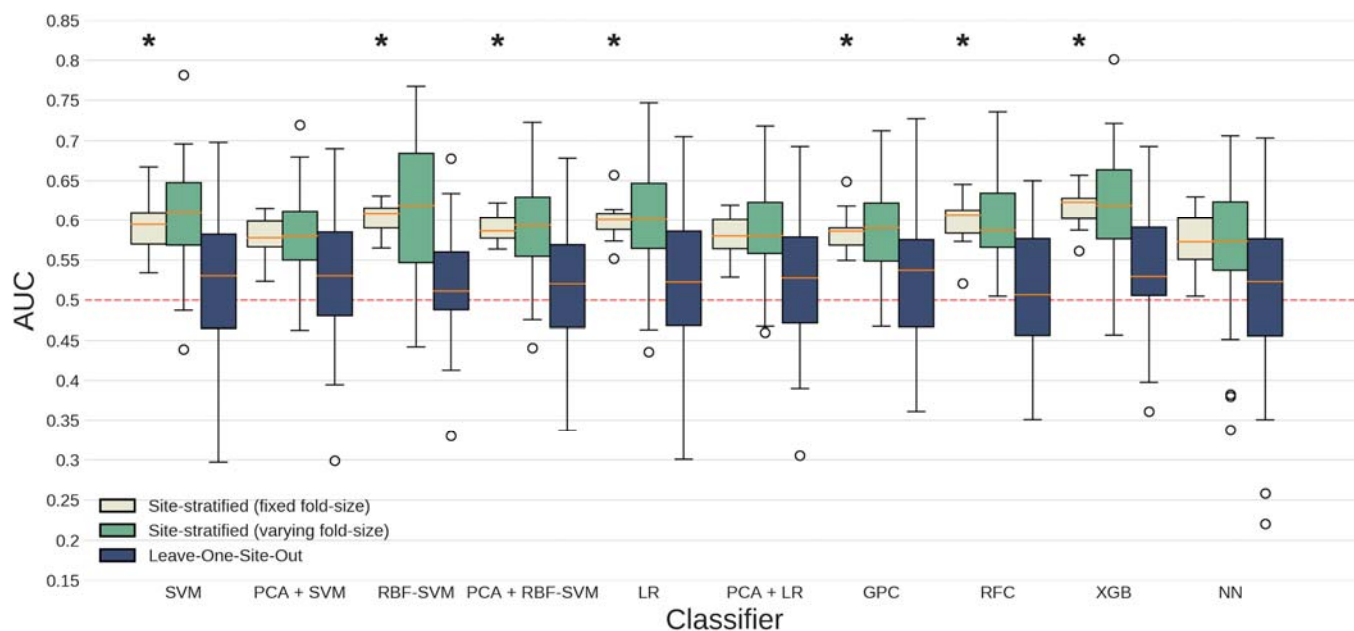


Figure 1. Performance for multi-site classification using different algorithms and cross-validation schemes. Boxplots summarize AUC scores obtained across CV-iterations; the dashed line represents chance-level performance and asterisks indicate scores significantly different from chance (Mann-Whitney-U statistic; $p < 0.05$ Bonferonni corrected (10 classifiers x 3 CV types), see eSupplement for details). SVM=Support Vector Machine, PCA=Principal Component Analysis, RBF=Radial Basis Function, LR=Logistic Regression, GPC=Gaussian Processes Classification, RFC=Random Forest Classifier, XGB=XGBoost, NN=Neural Network.

Multi-site classification

Three different CV approaches were used to assess the influence of sample heterogeneity. Results using various classification algorithms are summarized in **Figure 1**. Classification performance (AUC) using site-stratified CV (with training on combined samples and equal fold sizes) ranged between 0.57 (standard

deviation (SD)=0.02; P_{corr} =0.19) and 0.62 (SD=0.03; P_{corr} <.001) across different classifiers. All models had statistically significant performance after multiple comparison corrections except for PCA+LR, PCA+SVM and NN classifiers. LOSO-CV led to lower classification performance; 0.51 (SD=0.11; P_{corr} >.99) to 0.54 (SD=0.08; P_{corr} >.99) AUC with relatively high variance across folds (SD=0.07-0.11) and no classifiers surviving multiple comparison corrections. AUC values obtained through site-stratified CV with different fold sizes were similar to site-stratified CV results with equal fold sizes, ranging between 0.57 (SD=0.08; P_{corr} >.99) and 0.62 (SD=0.07; P_{corr} =.55). However, variance across CV-iterations was higher and comparable to that from LOSO-CV (SD; site-stratified fixed: 0.02–0.04; site-stratified variable: 0.05-0.08; LOSO: 0.07-0.11). A complete overview of classification results is provided in supplementary **Table S2**. Multi-site classification, performed separately on pediatric and adult samples yielded similar results, ranging from 0.56 (SD=0.03; P_{corr} >.99) to 0.62 (SD=0.06; P_{corr} =.71) and 0.56 (SD=0.03; P_{corr} =.69) to 0.61 (SD=0.02; P_{corr} =.008) AUC, respectively (see supplementary **Tables S3-4**). As site-stratified CV with equal fold-sizes resulted in the best performances, we used this strategy for further evaluation of intra-site performance and the influence of clinical variables. Only RFC classification performance is reported here, as differences between classifiers were minimal and this model was also used to extract feature importance.

Single-site classification

Single-site classification performance with 10-fold CV varied greatly, with AUCs ranging between 0.17-0.91 across different sites and classifiers (see supplementary **Table S5**). **Figure 2** summarizes RFC performances for each individual site. We assessed the correlation between the number of participants and classification performance across sites, which showed a non-significant trend (r_s =0.29, p =0.054).

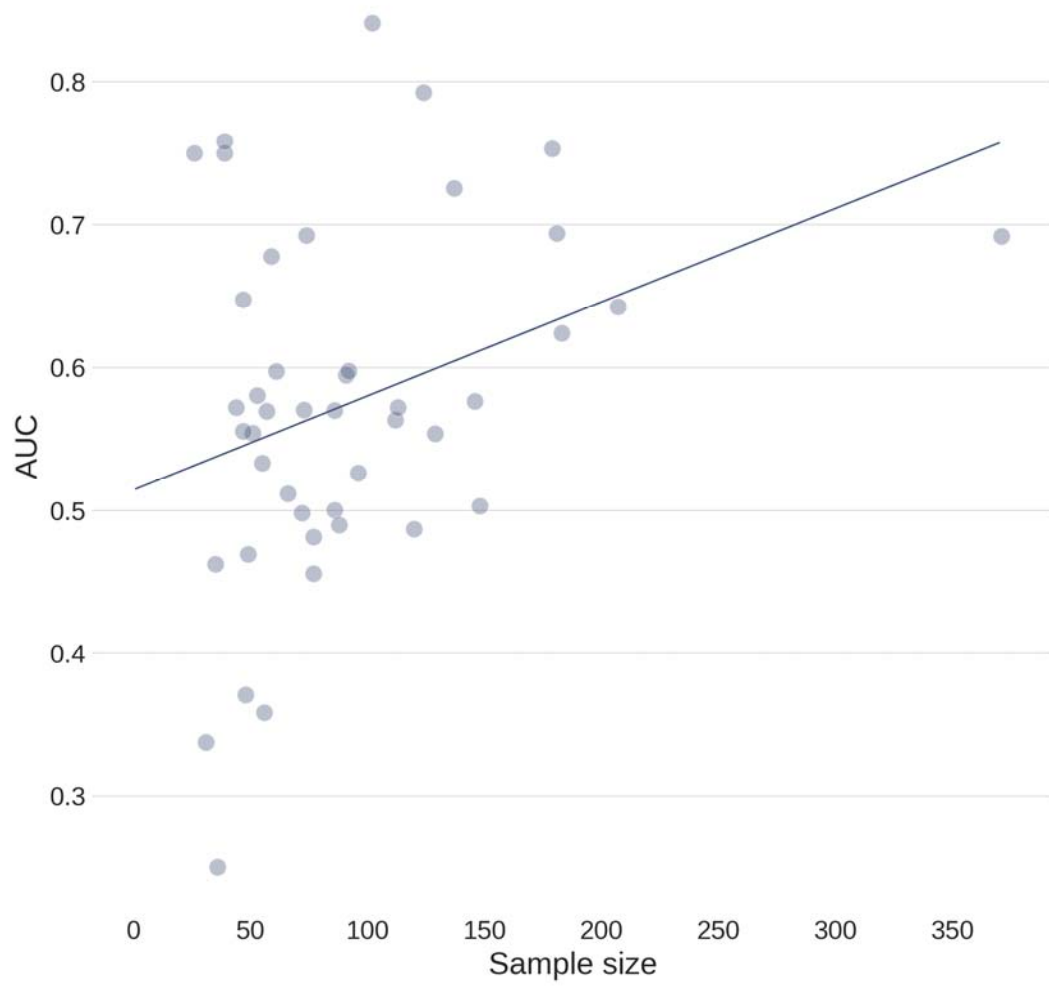


Figure 2. Scatterplot illustrating relationship between number of participants and classification performance across sites.

Only RFC classifier performance averaged across CV-iterations are plotted (Spearman correlation; $r_s=0.29$, $p=0.054$).

Clinical variables and sensitivity analysis

To assess the influence of different clinical variables on classification performance, we repeated the analysis for specific subgroups split according to medication use, AO, disease duration, and severity. A complete overview is provided in supplementary **Tables S6(a-d)**, and we report results using RFC on combined data with age, sex and site as covariates below. Medicated OCD vs HC classification resulted in 0.73 AUC ($SD=0.03$; $P_{corr}<.001$), unmedicated OCD vs HC in 0.61 ($SD=0.02$; $P_{corr}=.03$), and medicated vs

unmedicated OCD in 0.86 (SD=0.02; $P_{\text{corr}} < .001$) (see **Figure 3**). Early AO OCD vs HC classification resulted in 0.68 AUC (SD=0.03; $P_{\text{corr}} < .001$), late AO OCD vs HC in 0.73 (SD=0.02; $P_{\text{corr}} < .001$), and early vs late AO in 0.81 (SD=0.03; $P_{\text{corr}} < .001$). As no late AO patients were present in pediatric samples, classifications were re-run on adult samples only, resulting in 0.65 AUC (SD=0.04; $P_{\text{corr}} = .01$) for early AO vs HC, 0.70 (SD=0.03; $P_{\text{corr}} < .001$) for late AO vs HC, and 0.73 (SD=0.05; $P_{\text{corr}} < .001$) for early vs late AO. Classification of short disease duration OCD vs HC resulted in 0.68 AUC (SD=0.04; $P_{\text{corr}} < .001$), long disease duration vs HC in 0.71 (SD=0.02; $P_{\text{corr}} < .001$), and short vs long duration in 0.78 (SD=0.04; $P_{\text{corr}} < .001$). Finally, low severity OCD vs HC classification resulted in 0.60 AUC (SD=0.03; $P_{\text{corr}} = .15$), high severity OCD vs HC in 0.61 (SD=0.03; $P_{\text{corr}} = .04$), and low vs high severity OCD in 0.58 (SD=0.04; $P_{\text{corr}} > .99$). Medication status correlated significantly with disease duration ($r = -0.094$; $p < 10^{-05}$; Bonferroni corrected). We therefore performed additional classifications after further stratification (e.g., HC vs medicated + short duration OCD; HC vs unmedicated + short duration OCD, etc). Classifications with or without stratification for disease duration were comparable (see supplementary **Tables S7(a-c)** for full overview).

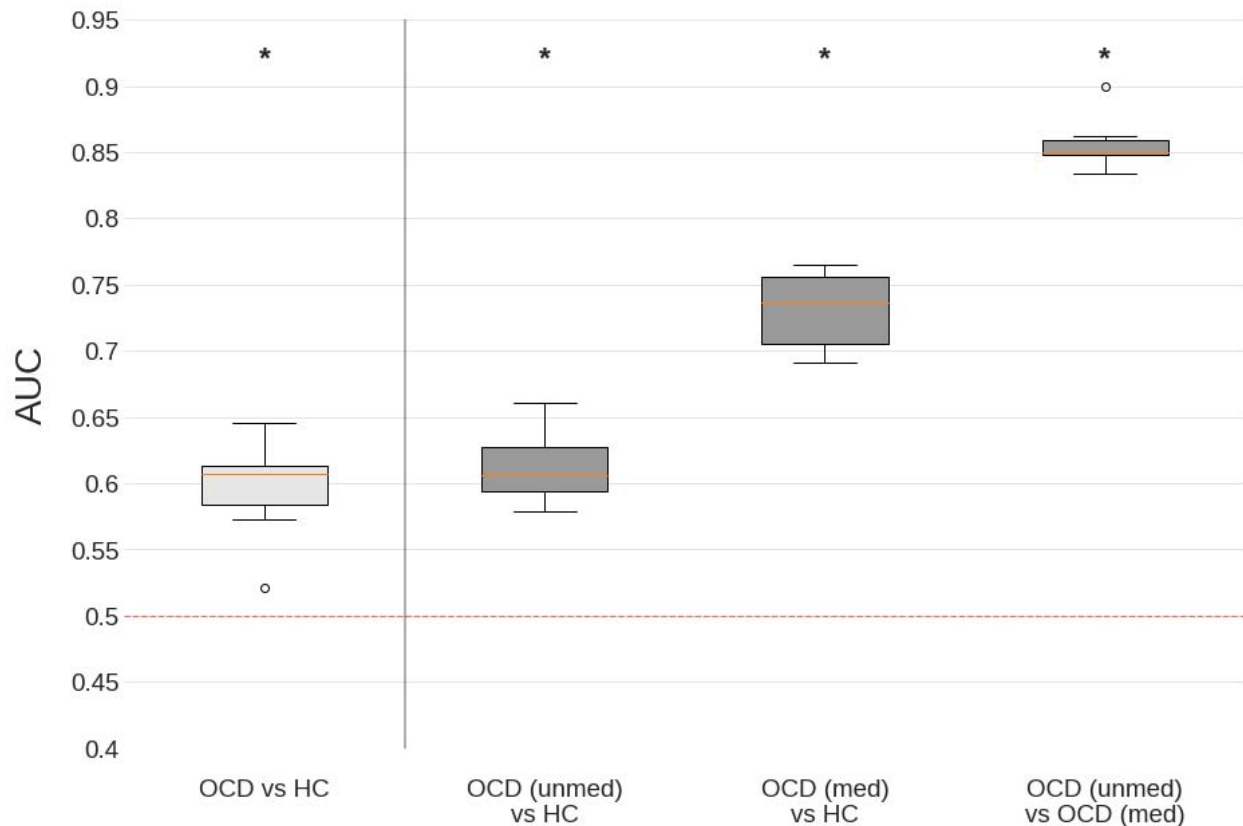


Figure 3. Performance for classification between subgroups of patients based on medication status. Only RFC classifier performance for combined data (both pediatric and adult samples) is shown here; Boxplots summarize AUC scores obtained across CV-iterations; the dashed line represents chance-level performance and asterisks indicate scores significantly different from chance (Mann-Whitney-U statistic; $p < 0.05$ Bonferonni corrected (10 classifiers x 3 CV types), see eSupplement for details). unmed=unmedicated, med=medicated.

Feature importance

We investigated which brain regions (features) contributed most to OCD vs HC classifications for site-stratified CV only, using the feature importance values from RFC and permutation testing. No features were selected consistently (survived false discovery rate (FDR) correction in >50% CV-iterations) for

main analyses (OCD patients vs HC classification) in either pediatric, adult or combined samples. However, for HC vs medicated OCD and medicated vs unmedicated OCD classification in combined samples, 7 and 36 significant- and consistently selected features were found, respectively. Additionally, 40 features were found for both HC vs late AO and early vs late AO patients classifications in combined samples. A complete overview of these findings (including features importance for classifications stratified for medication and AO in adult samples) can be found in supplementary **Tables S8-11**.

Discussion

We found that MVPA of parcellated structural neuroimaging data is unable to provide accurate distinction between OCD cases and controls. Classification of the complete sample using site-stratified CV ranged between 0.58 and 0.62 AUC, which is not sufficient for clinical application. Differences in performance between classifiers were minimal. Similar results were obtained for classifications performed separately on pediatric or adult samples. When validated on completely new data from other sites using LOSO-CV, model performance hardly exceeded chance-level. Our findings highlight the impact of validation schemes on classification performance and suggest poor discrimination between OCD patients and controls when combining data from multiple sites. In contrast, discrimination between subgroups of patients based on medication status enabled good individual subject classification.

Few diagnostic classifiers have been applied to OCD across multiple scanners and sites. Prior studies using structural MRI data to classify OCD using single-site samples yielded accuracies ranging from 0.72 up to 0.93 (reviewed in [12]). The wide range of performances observed in our individual site classification is in agreement with the published literature. Such a wide range may in part be explained by sample size, as larger samples tended to have higher AUC values [14, 28, 32]. However, this relationship does not necessarily hold true for large-scale multi-site studies, due to heterogeneity that arises from pooling samples with different scanning parameters, processing pipelines, inclusion criteria, demographic and clinical characteristics [12, 33]. All these factors can impact the data and obscure a pattern of abnormalities shared by all patients. Single-site studies that minimize heterogeneity may therefore yield higher classification performances, but limit the generalizability to new, unseen data and its use in clinical practice [14, 15]. LOSO-CV demonstrated that these structural MRI features do not provide a biomarker that enables generalization to new sites.

Classification within subgroups, split according to medication status, resulted in good performance even after accounting for correlated clinical variables (i.e., disease duration) through additional stratification. Evidence from rodent studies suggests that serotonin reuptake inhibitors (SRIs) mediate neuroplasticity in various cortical and subcortical structures through glio- and neuro-genesis [34–36]. However, little is known about how these findings might translate to humans and what the effects of long-term medication use are [37]. A few longitudinal studies suggest that SRI treatment normalizes brain volumes. One study reported significantly larger thalamic volumes in treatment-naïve pediatric patients compared to controls and these differences decreased following paroxetine treatment [38]. Another study reported smaller gray matter volume of CSTC-related regions in treatment-naïve patients that were no longer detectable following fluoxetine treatment [39]. Nonetheless, it remains unclear whether these structural changes are related to medication use or to symptom improvement.

Features that enabled multivariate classifications for medicated OCD vs HC included thickness of right medial orbitofrontal, right superior frontal, bilateral rostral middle frontal and right pars triangularis cortices in both adult and combined samples, and left palladic and right lateral ventricle volumes in adult samples only. Features for medicated vs unmedicated OCD in adult and combined samples included widespread cortical thickness in frontal, temporal, parietal and occipital regions. Although these multivariate features are important for the classifications as a whole, this appears consistent with previous univariate ENIGMA-OCD meta- and mega-analyses that also reported medication effects [6, 7]. Compared to HC, pediatric OCD patients had larger thalamic volumes and this finding was specific to unmedicated patients. This finding is in line with the normalizing effects of paroxetine on thalamic volume described previously. However, most findings have pointed towards more pronounced brain abnormalities in medicated patients than in unmedicated patients compared to controls. For example, medicated pediatric patients were found to have smaller cortical surface area (mainly in frontal regions) that was not detected in unmedicated patients. Medicated adult OCD patients

showed thinner frontal, temporal and parietal cortices, and smaller hippocampal and larger pallidum volumes, whereas no differences were found for unmedicated adult patients. These results fit with the finding that the classification performance between medicated vs. unmedicated patients was better than that between cases versus controls, and may reflect minimization of heterogeneity in stratified patient groups compared to HC. Although these findings suggest that antidepressants affect brain structure that enables good single-subject classification, causal evidence for medication effects warrants prospective longitudinal studies. Furthermore, it remains unclear whether these findings are specific to OCD, or whether it is also present in other psychiatric disorders.

A number of limitations deserve emphasis. First, we used a sample pooled from existing data across the world, without harmonized protocols for scanning, inclusion criteria or demographic and clinical characteristics. These sources of heterogeneity may limit classification performance, but this also provides an opportunity for model development using independent data sets and the discovery of biomarkers that are reproducible across study sites. Second, limited information on medication use was available. We were therefore only able to distinguish patients on antidepressants with or without adjuvant antipsychotics versus those who had not received any medication. Medication history, medication dosage, and duration of use were unknown. Nonetheless, these coarsely defined medication groups enabled better case-control discrimination and good classification of medicated versus unmedicated cases. Third, there is a lack of information on comorbidity and OCD subtype in our dataset. Particular OCD subtypes may have different neural correlates, and this might limit the ability of MVPA models to find generalizable patterns in brain structure [12, 40]. Finally, it is possible that the brain features used for classification led to sub-optimal performance. OCD is thought to derive from abnormalities distributed at the network-level rather than focused on a single brain area, and FreeSurfer features might not be sufficiently sensitive to detect subtle alterations associated with OCD.

Taken together, this study provides a realistic estimate of the classification performance that can be achieved in a large, ecologically valid, multi-site sample of OCD participants using data on regional brain structure. Our findings show that parcellated structural MRI data does not enable a good overall distinction between patients with OCD and healthy controls. However, classifying subgroups of patients based on medication status enables good identification at the individual subject level. This underlines the need for longitudinal studies on the short- and long-term effects of medication on brain structure.

AUTHOR AND ARTICLE INFORMATION

Please address correspondence to W. B. Bruin (w.b.bruin@amsterdam.umc.nl). See the [online data supplement](#) for the complete list of ENIGMA OCD Working Group members.

DISCLOSURES

Pharmaceuticals, and Biohaven Pharmaceuticals. Dr. Walitza has received lecture honoraria Opopharma in the Dr. Baker has received research support from the National Institute of Mental Health (NIMH) and Valera Health. Dr. Brennan has received consulting fees from Rugen Therapeutics and Nobilis Therapeutics and research grant support from Eli Lilly, Transcept last 3 years. Her work was supported in the last 3 years by the Swiss National Science Foundation (SNF), diverse EU FP7s, HSM Hochspezialisierte Medizin of the Kanton Zurich, Switzerland, Bfarm Germany, Zinep, Hartmann Müller Stiftung, Olga Mayenfisch. Dr. Dan J. Stein has received research grants and/or consultancy honoraria from Lundbeck and Sun in the past 3 years. Dr. Paul M. Thompson has received research grant support from Biogen, Inc., for research unrelated to the topic of this manuscript. Dr. Ivanov has received honoraria from Lundbeck as a member of the Data Safety Monitoring Committee and research grants from the National Institute on Drug Abuse in the last 3 years. Dr. Pittenger has received research support and/or honoraria for consultation from Biohaven Pharmaceuticals, Blackthorn Therapeutics, Abide Therapeutics, and Brainsway, and royalties or honoraria from Oxford University Press and Elsevier in the past 3 years. Dr. Feusner has received an honorarium from Pfizer and consultation fees from NOCD, Inc. Dr. Piacentini has received research support from Pfizer Pharmaceuticals for research unrelated to the topic of this manuscript. Dr. Soreni has received support for Investigator Initiated Clinical Trial from Lundbeck LLC unrelated to the topic of this study. Dr. Buitelaar has been in the past 3 years a consultant to / member

of advisory board of / and/or speaker for Shire, Roche, Medice, and Servier. He is not an employee of any of these companies, and not a stock shareholder of any of these companies. He has no other financial or material support, including expert testimony, patents, royalties. Dr Mataix-Cols receives royalties for contributing articles to UpToDate (Wolters Kluwer Health), and for editorial work from Elsevier, all unrelated to the current work. In the last three years, Dr. Simpson has received research support for an industry-sponsored clinical trial from Biohaven Pharmaceuticals, royalties from UpToDate, Inc, and a stipend from JAMA Psychiatry for her role as Associate Editor. All other authors from the ENIGMA OCD working group have no conflicts of interest related to this study.

GRANT SUPPORT

The ENIGMA-Obsessive Compulsive Disorder Working-Group gratefully acknowledges support from the NIH BD2K (Big Data to Knowledge) award U54 EB020403 (PI: Dr. Thompson) and Neuroscience Amsterdam, and an IPB grant to Dr. Schmaal & Dr. van den Heuvel. Supported by the Japan Society for the Promotion of Science (JSPS; KAKENHI Grants No. 18K15523 to Dr. Abe, No. 16K19778, No. 18K07608 to Dr. Nakamae, No. 16K04344, 19K03309 to Dr. Hirano, and No. 26461762 to Dr. Nakagawa); the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, São Paulo Research Foundation; Grant No. 2011/21357-9; 2018/04654-9 and 2018/21934-5); the EU FP7 project TACTICS (Grant No. 278948 to Dr. Buitelaar); the National Natural Science Foundation of China (No. 81560233 to Dr. Cheng); the International Obsessive-Compulsive Disorder Foundation (IOCDF) Research Award to Dr. Gruner; the Dutch Organization for Scientific Research (NWO) (grants 912-02-050, 907-00-012, 940-37-018, and 916.86.038); the Netherlands Society for Scientific Research (NWO-ZonMw VENI grant 916.86.036 to Dr. van den Heuvel; NWO-ZonMw AGIKO stipend 920-03-542 to Dr. de Vries), a NARSAD Young Investigator Award to Dr. van den Heuvel, and the Netherlands Brain Foundation (2010(1)-50 to Dr. van den Heuvel);

the Federal Ministry of Education and Research of Germany (No. BMBF-01GW0724 to Dr. Kathmann); the Alberta Innovates Translational Health Chair in Child and Youth Mental Health and funding from the Ontario Brain Institute to Dr. Arnold; the Deutsche Forschungsgemeinschaft (DFG; Grant No. KO 3744/7-1 to Dr. Koch); the Helse Vest Health Authority (No. 911754, 911880 to Dr. Kvale) and the Norwegian Research Council (No. HELSEFORSK 243675 to Dr. Kvale); the Wellcome Trust and a pump priming grant from the South London and Maudsley Trust, London, UK (Project Grant No. 064846 to Dr. Mataix-Cols); the Generalitat de Catalunya (AGAUR 2017 SGR 1247 to Dr. Menchón); the PhD-iHES program (FCT fellowship Grant No. PDE/BDE/113601/2015 to Dr. Moreira); the Japanese Ministry of Education, Culture, Sports, Science and Technology (Grant-in-Aid for Scientific Research (C) 22591262, 25461732, 16K10253 to Dr. Nakao); the Italian Ministry of Health (No. RC13-14-15-16-17-18A to Dr. Spalletta, Dr. Piras Fabrizio, Dr. Piras Federica); the National Center for Advancing Translational Sciences (Grants No. UL1TR000067/KL2TR00069 to Dr. Stern); the SA MRC funding and the National Research Foundation of South Africa to Dr. Stein and Dr. Lochner; the Canadian Institutes of Health Research, Michael Smith Foundation for Health Research, and British Columbia Provincial Health Services Authority funding Dr. Stewart; the Netherlands Organization for Scientific Research (NWO/ZonMW Vidi 917.15.318 to Dr. van Wingen); the Wellcome-DBT India Alliance (Grant No. 500236/Z/11/Z to Dr. Venkatasubramanian); the National Natural Science Foundation of China (No. 81371340) and the Shanghai Key Laboratory of Psychotic Disorders (No. 13dz2260500) to Dr. Wang; the Government of India grants from the Department of Science and Technology (Grants No. SR/S0/HS/0016/2011 to Prof. Y.C. Janardhan Reddy, and DST INSPIRE faculty grant -IFA12-LSBM-26 to Dr. Janardhanan C. Narayanaswamy) and from the Department of Biotechnology (Grants No. BT/PR13334/Med/30/259/2009 to Prof. Y.C. Janardhan Reddy, and No. BT/06/IYBA/2012 to Dr. Janardhanan C. Narayanaswamy); the Dana Foundation and NARSAD to Dr. Fitzgerald; the NIMH (Grants No. R01 MH107419 and No. K23MH082176 to Dr. Fitzgerald, Grant No. R21MH101441 to Dr. Marsh, Grant No. R21MH093889 to Drs. Marsh & Simpson,

Grant No. K23-MH104515 to Dr. Baker; Grant No. R33MH107589 to Dr. Stern, Grant No. R01MH111794 to Dr. Stern; Grant No. R01 MH081864 to Drs. O'Neill and Piacentini; Grant No. R01 MH085900 to Drs. O'Neill and Feusner, Grant No. K23 MH115206 to Dr. Gruner); the NIMH and the David Judah Fund at the Massachusetts General Hospital (Grant No. K23-MH092397 to Dr. Brennan); the Marató TV3 Foundation (Grants No. 01/2010 and 091710 to Dr. Lazaro); the Carlos III Health Institute (PI040829 to Dr Lazaro, Grant CPII16/00048 and Project Grants PI13/01958 & PI16/00889 to Dr. Soriano-Mas), co-funded by FEDER funds/European Regional Development Fund (ERDF), a way to build Europe; the AGAUR (2017 SGR 881 to Dr. Lazaro); the Michael Smith Foundation for Health Research funding Dr. Jaspers-Fayer and Dr. Stewart; the Carlos III Health Institute (PI14/00419 to Dr. Alonso, Grant No. FI17/00294 to Ignacio Martínez-Zalacáin, Project Grant No. PI16/00950 to Dr. Menchón); the Japan Agency for Medical Research and Development (AMED Brain/MINDS Beyond program Grant No. JP19dm0307002 to Dr. Shimizu); Oxford Health Services Research Committee grant to Dr A James and National Institute on Drug Abuse to Dr. Ivanov (NIH/NIDA Grant No. R21/DA046029 and No. R21/DA045218); the Wellcome & Royal Society (211155/Z/18/Z) to Tobias U Hauser; the Jacobs Foundation (2017-1261-04) to Tobias U Hauser, the Medical Research Foundation, and the Brain & Behavior Research Foundation (2018 NARSAD Young Investigator grant; 27023) to Tobias U. Hauser; the Swiss National Science Foundation (SNSF 320030_130237) to Susanne Walitza; the Harmann Müller foundation (1460) to Silvia Brem; NHMRC Career Development Fellowship (1140764 to Dr. Schmaal).

References

1. Ruscio AM, Stein DJ, Chiu WT, Kessler RC, Mattar L, Thiébaud MR, et al. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Psychiatry Res.* 2010;15:513–517.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)*. *Diagnostic Stat Man Ment Disord 4th Ed TR.* 2013:280.
3. Abramowitz JS, Taylor S, McKay D. Obsessive-compulsive disorder. *Lancet.* 2009;374:491–499.
4. Menzies L, Chamberlain SR, Laird AR, Thelen SM, Sahakian BJ, Bullmore ET. Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: The orbitofronto-striatal model revisited. *Neurosci Biobehav Rev.* 2008;32:525–549.
5. van den Heuvel OA, van Wingen G, Soriano-Mas C, Alonso P, Chamberlain SR, Nakamae T, et al. Brain circuitry of compulsivity. *Eur Neuropsychopharmacol.* 2016;26:810–827.
6. Boedhoe PSW, Schmaal L, Abe Y, Ameis SH, Arnold PD, Batistuzzo MC, et al. Distinct Subcortical Volume Alterations in Pediatric and Adult OCD: A Worldwide Meta- and Mega-Analysis. *Am J Psychiatry.* 2017;174:60–69.
7. Boedhoe PSW, Schmaal L, Abe Y, Alonso P, Ameis SH, Anticevic A, et al. Cortical Abnormalities Associated With Pediatric and Adult Obsessive-Compulsive Disorder: Findings From the ENIGMA Obsessive-Compulsive Disorder Working Group. *Am J Psychiatry.* 2018;175:453–462.
8. Kong XZ, Boedhoe PSW, Abe Y, Alonso P, Ameis SH, Arnold PD, et al. Mapping Cortical and Subcortical Asymmetry in Obsessive-Compulsive Disorder: Findings From the ENIGMA Consortium. *Biol Psychiatry.* 2019:1–13.
9. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci*

- Biobehav Rev. 2012;36:1140–1152.
10. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*. 2016. 2016. <https://doi.org/10.1016/j.neuroimage.2016.02.079>.
 11. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer’s disease. *Brain*. 2008;131:681–689.
 12. Bruin W, Denys D, van Wingen G. Diagnostic neuroimaging markers of obsessive-compulsive disorder: Initial evidence from structural and functional MRI studies. *Prog Neuro-Psychopharmacology Biol Psychiatry*. 2018:0–1.
 13. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*. 2017:1–10.
 14. Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: Brain models in translational neuroimaging. *Nat Neurosci Rev*. 2017;20:365–377.
 15. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav Rev*. 2015;57:328–349.
 16. Nunes A. Using structural MRI to identify bipolar disorders – 13 site machine learning study in 3020 individuals from the ENIGMA Bipolar Disorders Working Group. *Mol Psychiatry*. 2018. 2018. <https://doi.org/10.1038/s41380-018-0228-9>.
 17. Boedhoe PSW, Heymans MW, Schmaal L, Abe Y, Alonso P, Ameis SH, et al. An Empirical Comparison of Meta- and Mega-Analysis With Data From the ENIGMA Obsessive-Compulsive Disorder Working Group. 2019;12:1–8.
 18. Silverman WK, Saavedra LM, Pina AA. Test-retest reliability of anxiety symptoms and diagnoses with the anxiety disorders interview schedule for DSM-IV: Child and parent versions. *J Am Acad*

- Child Adolesc Psychiatry. 2001. 2001. <https://doi.org/10.1097/00004583-200108000-00016>.
19. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36:980–988.
 20. Sheehan D V., Sheehan KH, Shytle RD, Janavs J, Bannon Y, Rogers JE, et al. Reliability and Validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *J Clin Psychiatry*. 2010;71:313–326.
 21. First MB. Structured Clinical Interview for the DSM (SCID) . *Encycl. Clin. Psychol.*, Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015. p. 1–6.
 22. Sheehan D V., Sheehan KH, Lecrubier Y, 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*, vol. 59, 1998. p. 22–33.
 23. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968–980.
 24. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning* (2nd edition). 2009.
 25. Rasmussen, C. E., Williams CKI. *Gaussian processes for machine learning*. vol. 14. 2004.
 26. Chang C, Lin C. LIBSVM: A Library for Support Vector Machines. *ACM Trans Intell Syst Technol*. 2013;2:1–39.
 27. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17- Augu, New York, New York, USA: ACM Press; 2016. p. 785–794.
 28. Abraham A, Milham MP., Di Martino A, Craddock RC, Samaras D, Thirion B, et al. Deriving

- reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage*. 2017;147:736–745.
29. Dinga R, Penninx BWJH, Veltman DJ, Schmaal L, Marquand AF. Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines. *BioRxiv*. 2019:743138.
 30. Storch EA, De Nadai AS, Conceição Do Rosário M, Shavitt RG, Torres AR, Ferrão YA, et al. Defining clinical severity in adults with obsessive-compulsive disorder. *Compr Psychiatry*. 2015;63:30–35.
 31. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8.
 32. Nieuwenhuis M, van Haren NEM, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage*. 2012;61:606–612.
 33. Schnack HG, Kahn RS. Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Front Psychiatry*. 2016;7.
 34. Czéh B, Müller-Keuker JH, Rygula R, Abumaria N, Hiemke C, Domenici E, et al. Chronic social stress inhibits cell proliferation in the adult medial prefrontal cortex: Hemispheric asymmetry and reversal by fluoxetine treatment. *Neuropsychopharmacology*. 2007. 2007. <https://doi.org/10.1038/sj.npp.1301275>.
 35. Kodama M, Fujioka T, Duman RS. Chronic olanzapine or fluoxetine administration increases cell proliferation in hippocampus and prefrontal cortex of adult rat. *Biol Psychiatry*. 2004. 2004. <https://doi.org/10.1016/j.biopsych.2004.07.008>.
 36. Soumier A, Banasr M, Lortet S, Masméjean F, Bernard N, Kerkerian-Le-Goff L, et al. Mechanisms contributing to the phase-dependent regulation of neurogenesis by the novel antidepressant, agomelatine, in the adult rat hippocampus. *Neuropsychopharmacology*. 2009. 2009. <https://doi.org/10.1038/npp.2009.72>.

37. Homberg JR, Schubert D, Gaspar P. New perspectives on the neurodevelopmental effects of SSRIs. *Trends Pharmacol Sci.* 2010;31:60–65.
38. Gilbert AR, Mac Master FP, Keshavan MS, Rosenberg DR, Narula V, Stewart CM, et al. Decrease in Thalamic Volumes of Pediatric Patients With Obsessive-compulsive Disorder Who Are Taking Paroxetine. *Arch Gen Psychiatry.* 2003;57:449.
39. Hoexter MQ, De Souza Duran FL, D’Alcante CC, Dougherty DD, Shavitt RG, Lopes AC, et al. Gray matter volumes in obsessive-compulsive disorder before and after fluoxetine or cognitive-behavior therapy: A randomized clinical trial. *Neuropsychopharmacology.* 2012;37:734–745.
40. Mataix-Cols D, van den Heuvel OA. Common and Distinct Neural Correlates of Obsessive-Compulsive and Related Disorders. *Psychiatr Clin North Am.* 2006;29:391–410.

