Psychological Medicine

cambridge.org/psm

Development and validation of a multimodal neuroimaging biomarker for electroconvulsive therapy outcome in depression: a multicenter machine learning analysis

Willem Benjamin Bruin[1], Leif Oltedal[2,3], Hauke Bartsch[2], Christopher Abbott[4], Miklos Argyelan[5,6], Tracy Barbour[7], Joan Camprodon[7], Samadrita Chowdhury[7], Randall Espinoza[8], Peter Mulders[9], Katherine Narr[10], Mardien Oudega[11], Didi Rhebergen[12], Freek ten Doesschate[1,13], Indira Tendolkar[9], Philip van Eijndhoven[9], Eric van Exel[11], Mike van Verseveld[13], Benjamin Wade[14], Jeroen van Waarde[13], Paul Zhutovsky[1], Annemiek Dols[11,*] and Guido van Wingen[1,15,*]

[1]Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam Neuroscience, Amsterdam, The Netherlands; [2]Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University Hospital, Bergen, Norway; [3]Department of Clinical Medicine, University of Bergen, Bergen, Norway; [4]Department of Psychiatry, University of New Mexico, Albuquerque, NM, USA; [5]The Feinstein Institutes for Medical Research, Manhasset, NY, USA; [6]The Zucker Hillside Hospital, Glen Oaks, NY, USA; [7]Division of Neuropsychiatry and Neuromodulation, Massachusetts General Hospital, Harvard Medical School. Boston, MA, USA; [8]Department of Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, USA; [9]Donders Institute for Brain, Cognition and Behavior, Department of Psychiatry, Nijmegen, The Netherlands; [10]Ahmanson-Lovelace Brain Mapping Center, Departments of Neurology, and Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, USA; [11]Department of Old Age Psychiatry, GGZinGeest, Department of Psychiatry, Amsterdam UMC, location VUmc, Amsterdam Neuroscience, Amsterdam, The Netherlands; [12]Mental Health Institute GGZ Centraal, Amersfoort; Department of Psychiatry, Amsterdam UMC, location VUmc, Amsterdam Neuroscience, Amsterdam, The Netherlands; [13]Rijnstate, Department of Psychiatry, Arnhem, The Netherlands; [14]Ahmanson-Lovelace Brain Mapping Center, Department of Neurology, UCLA, Los Angeles, USA and [15]Amsterdam Brain and Cognition, University of Amsterdam, The Netherlands

Abstract

Background. Electroconvulsive therapy (ECT) is the most effective intervention for patients with treatment resistant depression. A clinical decision support tool could guide patient selection to improve the overall response rate and avoid ineffective treatments with adverse effects. Initial small-scale, monocenter studies indicate that both structural magnetic resonance imaging (sMRI) and functional MRI (fMRI) biomarkers may predict ECT outcome, but it is not known whether those results can generalize to data from other centers. The objective of this study was to develop and validate neuroimaging biomarkers for ECT outcome in a multicenter setting.

Methods. Multimodal data (i.e. clinical, sMRI and resting-state fMRI) were collected from seven centers of the Global ECT-MRI Research Collaboration (GEMRIC). We used data from 189 depressed patients to evaluate which data modalities or combinations thereof could provide the best predictions for treatment remission (HAM-D score $\leqslant 7$) using a support vector machine classifier.

Results. Remission classification using a combination of gray matter volume and functional connectivity led to good performing models with average 0.82–0.83 area under the curve (AUC) when trained and tested on samples coming from the three largest centers ($N = 109$), and remained acceptable when validated using leave-one-site-out cross-validation (0.70–0.73 AUC).

Conclusions. These results show that multimodal neuroimaging data can be used to predict remission with ECT for individual patients across different treatment centers, despite significant variability in clinical characteristics across centers. Future development of a clinical decision support tool applying these biomarkers may be feasible.

Introduction

Electroconvulsive therapy (ECT) is currently the most effective intervention for patients with treatment resistant depression. Despite its high efficacy, ECT remains underutilized, as only 1–2% of patients with severe or persistent depression receive ECT (Slade, Jahn, Regenold, & Case, 2017). Although approximately 48% of treatment resistant patients recover with ECT, it is also associated with adverse cognitive effects and may be regarded as more invasive than other

Original Article

*Shared senior author.

Cite this article: Bruin WB *et al* (2024). Development and validation of a multimodal neuroimaging biomarker for electroconvulsive therapy outcome in depression: a multicenter machine learning analysis. *Psychological Medicine* **54**, 495–506. https://doi.org/10.1017/S0033291723002040

Received: 8 February 2023
Revised: 31 May 2023
Accepted: 3 July 2023
First published online: 24 July 2023

Keywords:
Biomarker; depression; ECT; machine learning; MRI; multimodal

Corresponding authors:
Willem Benjamin Bruin;
Email: w.b.bruin@amsterdamumc.nl;
Guido van Wingen;
Email: g.a.vanwingen@amsterdamumc.nl

treatment options because the use of anesthesia is necessary (Heijnen, Birkenhager, Wierdsma, & van den Broek, 2010). Furthermore, ECT is relatively expensive and non-responsiveness can only be determined after multiple sessions. Information that better predicts treatment outcome would enable patient selection thereby further improving the overall response rate and avoiding ineffective treatment with adverse effects. A personalized recommendation about the expected benefit of ECT could provide a valuable addition to the treating physician's clinical judgement and may increase its use in clinical practice.

Meta-analyses of studies that investigated predictors of ECT outcome have associated several clinical characteristics with beneficial ECT outcome, in particular no history of treatment resistance, older age, and presence of psychotic symptoms (Haq, Sitzmann, Goldman, Maixner, & Mickey, 2015; van Diermen et al., 2018). However, the effect sizes are small, limiting their use to guide individual patient selection. Recent studies have started using neuroimaging data to predict ECT outcome at the individual level using machine learning analysis, which can construct multivariate prediction models using all the available data. Initial small-scale studies have shown that both structural magnetic resonance imaging (MRI) and functional MRI findings can be used to predict ECT outcome with approximately 80% accuracy (Redlich et al., 2016; van Waarde et al., 2015). These initial results have been confirmed by subsequent studies, and a recent meta-analysis showed an average prediction accuracy of 82% and area under the receiver operator characteristic curve (AUC) of 83% (Cohen, Zantvoord, Wezenberg, Bockting, & van Wingen, 2021), which is considered to be excellent discrimination between groups (Hosmer, Lemeshow, & Sturdivant, 2013). Furthermore, a recent study using routine MRI data to predict ECT outcome in a relatively large sample of 71 patients reached an accuracy of 69% accuracy (Gartner et al., 2021).

Despite these promising results, the existing studies have been limited by small sample sizes and monocenter settings. This reduces the possibility for models to generalize to new patients across centers. Although machine learning models typically perform better when trained on large samples from the same center, classification accuracy of larger multicenter studies tends to decrease, presumably due to increased clinical (e.g. adults v. elderly) and technological (e.g. different MRI hardware and protocols) variability across centers (Schnack & Kahn, 2016). In order to develop robust and generalizable neuroimaging biomarkers for ECT outcome, we used data from the Global ECT-MRI Research Collaboration (GEMRIC) and validated classification performance in a multicenter setting (Oltedal et al., 2017). We used multimodal data (i.e. clinical, structural MRI [sMRI], and resting-state functional MRI [rs-fMRI]) and evaluated which data modalities or combinations thereof might provide the best predictions. Remission (17-item Hamilton Depression Rating Scale (HAM-D) score of $\leqslant$7 after treatment) was used as the primary outcome criterion. Remission may provide a better outcome criterion than response (at least 50% HAM-D reduction compared to baseline) and has become the gold standard for depression treatment, because patients who do not remit have a poorer prognosis and greater chance of relapse and recurrence than those who do (McIntyre & O'Donovan, 2004; van Diermen et al., 2018). Additionally, as most sites only contributed a small sample, we also evaluated model performance when only data from centers with $\geqslant$20 patients were used to provide classifiers with a minimum of approximately 10 examples per class per center, which potentially could increase classification performance (Abraham et al., 2017). Finally, we visualized the brain regions that were most informative to the classifications, in order to gain insight into the brain regions predictive of ECT outcome. To adhere to guidelines on reporting of diagnostic studies, we report our findings based on TRIPOD guidelines (Moons et al., 2015).

Methods

Participants

We performed a retrospective study using data from GEMRIC (v3.1, DOI:10.17605/OSF.IO/WD436), an international consortium that contains the largest multicenter database of neuroimaging scans of patients treated with ECT (Oltedal et al., 2017). All contributing sites received ethics approval from their local ethics committee or institutional review board. In addition, the centralized mega-analysis was approved by the Regional Ethics Committee South-East in Norway (No. 2018/769) (Oltedal et al., 2018). Analyses contained a selection of sMRI and rs-fMRI data from seven centers across Europe and North America, recorded from 189 clinically depressed patients according to ICD-10 (167 unipolar, 22 bipolar; see online Supplementary Table S1 for diagnoses per center) who had received right unilateral or bilateral ECT (or both). Depressed patients were eligible for ECT, typically after failure to respond to first-line treatments with conventional psychotherapy and antidepressant medications. The patients were included because of the availability of both high quality sMRI and rs-fMRI data. ECT parameters varied between different centers, including electrode placement. A description of center-specific ECT procedures and image acquisition is provided elsewhere (Oltedal et al., 2017). As GEMRIC consists of samples ranging from very small to relatively large (N = 14, 14, 15, 18, 19, 29, 38, 42), we performed all analyses on the entire cohort and on centers with $\geqslant$20 patients available (three centers, N = 109) in order to ensure classifiers were provided with sufficient data per center.

Choice of primary measure

Treatment outcome was measured using the HAM-D or Montgomery-Åsberg Depression Rating Scale (MADRS) that was converted to HAM-D (online Supplementary Methods), which are gold standard ratings for depression severity. Remission (minimal symptoms) was used as the primary outcome criterion and defined as post-ECT HAM-D score $\leqslant$7.

MRI data and preprocessing

MRI acquisition parameters are listed in online Supplementary Tables S2 and S3. Structural T1-weighted scans were acquired using 1.5 T and 3 T scanners with a minimum resolution of 1.33 mm^3 and preprocessed using the CAT12 toolbox (v12.6; http://www.neuro.uni-jena.de/cat/) for voxel-based morphometry (VBM). Images were segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), normalized to MNI space using DARTEL registration, resampled to 1.5 mm^3 isotropic and spatially smoothed with an 8 mm isotropic Gaussian kernel. GM volumes were masked at 0.2 to exclude WM.

150–266 rs-fMRI volumes were acquired with a TR of 1.7–3.0s, in-plane resolution of 2.4–3.75 mm, and slice thickness of 3–5 mm. Preprocessing was performed using ANTs (v2.2.0; https://github.com/ANTsX/ANTs) and FSL (v5.0.10;

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

http://fsl.fmrib.ox.ac.uk/), including brain extraction, boundary-based co-registration, motion correction, spatial smoothing with a 5 mm isotropic Gaussian kernel, and normalization to a 2 mm MNI template. Denoising was performed using ICA-AROMA, and depending on the type of analysis, high-pass ($f > 0.01$) or bandpass filtering ($0.009 < f < 0.08$) was applied together with WM and CSF nuisance regression (Pruim et al., 2015). Denoised rs-fMRI data were resampled to 4 mm isotropic.

Only subjects that passed quality control for both rs-fMRI and sMRI were included for analysis, leading to a final sample of 189 patients (online Supplementary Fig. S1 for a flowchart). Further details on MRI preprocessing and quality control are provided in online Supplementary Methods.

Feature extraction

We extracted commonly used MRI features from the preprocessed data. For sMRI, we used voxel-wise modulated GM maps (VBM) and 142 cortical and subcortical parcellations using the Neuromorphometrics atlas (NMM; provided by Neuromorphometrics, Inc). For rs-fMRI, we used a high-dimensional resting-state networks template from the UK BioBank dataset to extract 100 independent spatial components that were derived using group independent component analysis (ICA) (Alfaro-Almagro et al., 2018). 45 components reflecting non-neural signals and three components mainly located in cerebellar regions with insufficient EPI coverage were discarded, resulting in 52 components for analysis. Group information guided ICA was used to derive subject-specific time-series and spatial maps for each of the 52 signal components using the high-pass filtered preprocessed data (Du & Fan, 2013). Time-series were used to calculate individual functional connectivity (FC) matrices that described pairwise connectivity between signal components with Pearson correlations (ICA-FC). Additionally, we used an atlas-based approach from Power et al. and extracted time-series from 264 functional areas to compute FC matrices (Power FC) using the bandpass filtered preprocessed data (Power et al., 2011). Correlations were converted to z-scores with Fisher r-to-z transformation before performing the classification. The total number of features used were: 406 929 for voxel-wise VBM maps, 142 for NMM parcellations, 37 401 for Power-based FC, 1378 for ICA-based FC, and 26 629 for each of the 52 ICA spatial components identified as signal. Further details on feature extraction are provided in online Supplementary Methods.

Machine learning

Machine learning classifications were performed using linear support vector machine classifiers [SVM; LIBSVM for Python (https://www.csie.ntu.edu.tw/~cjlin/libsvm/) implemented in scikit-learn (v0.23.1; https://scikit-learn.org/)] and validated using stratified shuffle-split cross-validation (CV) with 100 iterations. At each iteration, data were randomly divided into independent training (80%) and test (20%) sets while preserving the proportion of remitters and non-remitters from each center to obtain maximally homogeneous splits. The model was always trained only on the training set and evaluated on the test set. The entire procedure was then repeated 100 times and the test performance is averaged as the final performance evaluation. This CV procedure is further referred to as 'internal validation'. In addition, we performed leave-one-site-out (LOSO) CV, in which all but one center was used to train the SVM while the remaining center was used to assess model performance (further referred to as 'external validation'). This procedure was repeated so that each center was used once for testing. LOSO reduces the risk of overfitting to data from a single center but may result in large between-sample heterogeneity of training and test sets, which could result in lower classification performance compared to internal validation. Hyper-parameters for the linear SVM were optimized using nested CV: a grid-search was performed across different values of C (0.001, 0.01, 0.01, 01, 1, 10, 100) using 10 inner stratified shuffle splits (this was done for both 'internal' and 'external' validation). We assessed classification performance using different sets of MRI features (VBM, NMM, ICA-DR FC, Power FC, and ICA spatial components), as well as using clinical data only (i.e. age, sex, and pre-ECT HAM-D scores) for baseline classification. Clinical data were always included for each classification. The primary performance metric was the AUC and reported metrics were averaged across CV iterations. Balanced accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are reported in online Supplementary Tables S6–S13.

Statistical significance of classification performance relative to chance was assessed using label permutation testing with 1000 iterations (Ojala & Garriga, 2010). Obtained p values were corrected for multiple comparisons using False Discovery Rate (FDR; two-stage [non-negative]; alpha = 0.05). 95% confidence intervals (CI) for AUC were computed using the modified Wald-method (Kottas, Kuss, & Zapf, 2014). To reduce the computational burden, only spatial ICA classifications that resulted in acceptable discrimination (>0.7 AUC) between groups were tested for significance (Hosmer et al., 2013). Finally, we assessed classification performance for multimodal classifications combining anatomical and functional data using feature concatenation: regional neuromorphometrics GM volumes with either ICA or Power-atlas-based FC, and voxel-wise GM with either ICA or Power-atlas-based FC. FDR correction was applied separately for classification results obtained using either internal or external validation; the full dataset or three largest centers only; and for unimodal, multimodal, and individual ICA spatial components, leading to 12 distinct families with qFDR set to $0.05/12 = 0.00417$. Details on classifier hyperparameter optimization and statistical significance testing are provided in online Supplementary Methods.

Anatomical localization

To investigate which regions contributed most to the classification, we employed a method to estimate p values for the weights of the SVM (Gaonkara, Shinohara, & Davatzikos, 2016). A statistic was computed incorporating the SVM weight component value and the size of the margin, and an analytical approximation to the null-distribution obtained through permutation testing was used to calculate p values. We only report p value feature importances for our best performing unimodal and multimodal models.

Results

Demographic data

Demographic data are presented in Table 1. Of the 189 included patients, 76 were remitters and 113 non-remitters. In line with previous literature, remitting patients were older and showed

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

Table 1. Demographics of patients included in data analysis, with subject demographics and comparisons between ECT remitters and non-remitters

	Total sample ($n = 189$)		Remitters ($n = 76$)		Non-remitters ($n = 113$)		
	mean	std	mean	std	mean	std	p
Age	51.7	15.5	56.3	14.2	48.6	15.5	<0.001*
Sex (m/f)	83/106	n.a.	32/44	n.a.	51/62	n.a.	0.79
Laterality (RUL/BL; $n = 188$)	148/40	n.a.	60/15	n.a.	88/25	n.a.	0.86
HAM-D pre-treatment	25.0	7.7	25.8	8.2	24.5	7.2	0.26
HAM-D post-treatment	11.0	8.3	3.3	2.3	16.2	6.6	<0.001*
HAM-D change	14.0	10.7	22.5	8.3	8.2	7.8	<0.001*
Diagnosis (UP+/UP-/BP+/BP-)	32/135/2/20	n.a.	23/44/1/8	n.a.	9/91/1/12	n.a.	<0.001*
Total ECT sessions ($n = 186$)	13.4	6.2	12.9	6.7	13.8	5.8	0.35

Abbreviations: BL, bilateral ECT initially; BP, bipolar depression with/without psychotic symptoms (BP+/–); f, female; HAM-D, Hamilton Rating scale for depression; m, male; n.a., not available; RUL, right unilateral ECT initially; UP, unipolar depression with/without psychotic symptoms (UP+/–). Asterisks depict significance using independent t test or χ^2 test.

more presence of psychotic symptoms (van Diermen et al., 2018). No significant differences in sex, initial electrode placement, symptom severity at baseline, and total number of ECT sessions were observed.

We assessed differences in sample demographics and clinical characteristics between different centers regardless of ECT outcome using one-way analysis of variance (ANOVA) and χ^2. Age [$F(7181) = 14.08$, $p < 0.001$], pre-treatment HAM-D scores [$F(7181) = 7.40$, $p < 0.001$], post-treatment HAM-D scores [$F(7181) = 5.24$, $p < 0.001$], HAM-D change [$F(7181) = 8.65$, $p < 0.001$], number of ECT sessions [$F(7178) = 10.78$, $p < 0.001$], depression type [$X^2(21, N = 189) = 77.58$, $p < 0.001$], and initial electrode placement [$X^2(7, N = 189) = 109.8$, $p < 0.001$] differed significantly between centers. In contrast, sex did not differ between centers [$X^2(7, N = 189) = 3.84$, $p = 0.80$]. Demographic data for the three largest centers (with $N \geqslant 20$) used for additional analyses are described in online Supplementary Tables S5 and S6. Differences in sample demographics and clinical characteristics between the three largest centers were similar to those seen in the entire sample. These findings highlight that there is considerable clinical heterogeneity between centers.

Remission prediction

All centers

Unimodal neuroimaging: Performance for remission classification was evaluated with internal (site stratified shuffle splits) cross-validation across centers using all samples ($N = 189$). Remission classification performance was poor with AUCs (averaged across 100 CV iterations) ranging between 0.58 and 0.67 for clinical data only (i.e. age, sex, and pre-ECT HAM-D scores) and VBM, NMM, ICA-FC, and Power-FC feature sets that also included the clinical data in all analyses (Fig. 1*a*). Nonetheless, these AUCs—except for NMM—were statistically significant following permutation testing with multiple comparison correction. Classification using external (LOSO) cross-validation hardly exceeded chance-level performance, with AUCs (averaged across sites) ranging between 0.51 and 0.58 and none were statistically significant (Fig. 1*a*). Classification using ICA networks did not exceed AUC > 0.7 for either internal or external validation. Notably, one ICA component centered around the anterior temporal lobes that included the amygdala and hippocampus resulted in 0.70 AUC but did not obtain statistical significance following

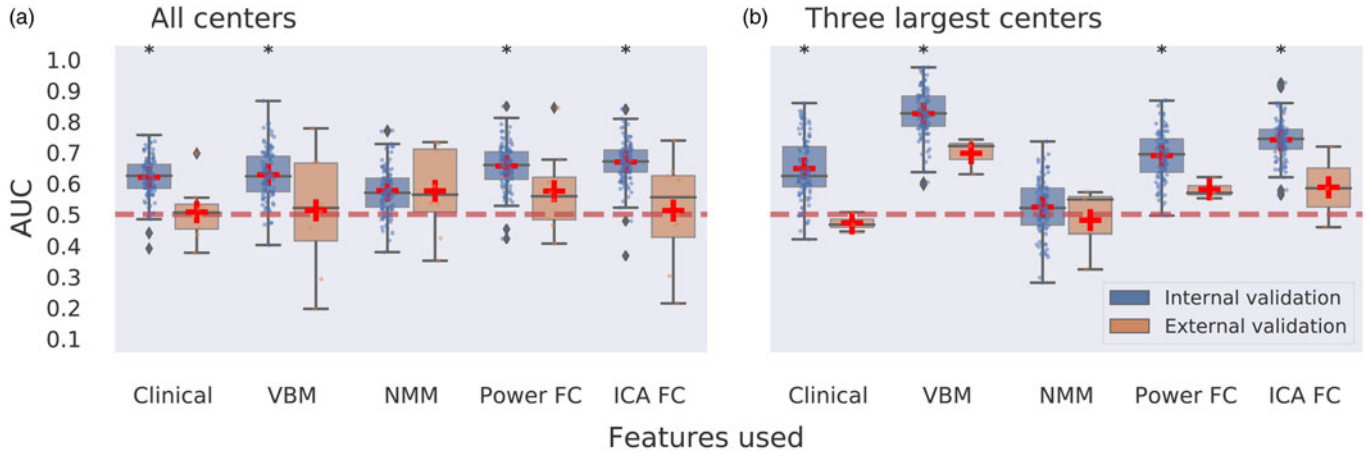

Figure 1. Multicenter predictions for ECT treatment remission using unimodal MR data modalities. Panel *a* depicts classification performance using data from all centers and different MR modalities with internal validation (AUC is averaged over 100 stratified cross-validation splits) and external validation (leave-one-site-out cross-validation, scores are averaged across different centers left out for model testing). Panel *b* depicts classification performance using data from the three largest centers with internal and external validation. VBM, voxel-based morphometry; NMM, Neuromorphometrics atlas; FC, functional connectivity; ICA, group information guided independent component analysis. Red dashed line depicts chance-level performance (0.5 AUC). Asterisks indicate significant difference from chance level after permutation testing with false discovery rate correction for multiple comparisons ($p < 0.05$, corrected).

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

with multiple comparison correction ($p_{FDRcorrected} = 0.2078$; $p_{uncorrected} = 0.0019$) (Fig. 2*a*). More comprehensive classification results, including balanced accuracy, sensitivity, specificity, PPV, and NPV, *p* values for AUC statistical significance and 95% CIs, are provided in online Supplementary Table S7.

Multimodal neuroimaging: Classification using a combination of sMRI, fMRI, and clinical data led to a maximum of 0.68 AUC using internal validation which was significantly different from chance level and a maximum of 0.64 AUC for external validation which did not obtain significance (Fig. 3*a*; online Supplementary Table S8).

Three largest centers

Unimodal neuroimaging

We next assessed prediction performance only using data from three centers with $N \geqslant 20$ ($N = 109$) to provide the machine learning classifier with sufficient samples per center. Classification performance with internal validation ranged between 0.52 and 0.83 AUC across different features used, and 0.65 AUC was obtained for classifications using clinical variables only (Fig. 1*b*). All these AUCs, except for NMM, showed statistical significance. Notably, the highest performance was achieved using voxel-wise GM data with 0.83 AUC. Two out of 52 ICA networks resulted in AUC > 0.7 (Fig. 2). One component centered around the temporal lobes resulted in 0.75 AUC, and a frontopolar network resulted in 0.80 AUC, but neither obtained statistical significance after multiple comparison correction. Classifications performed with external validation ranged between 0.47 and 0.70 AUC (Fig. 1*b*). The performance obtained with voxel-wise GM data using internal validation was reduced from 0.83 AUC to 0.70 AUC with external validation and failed to obtain statistical significance following permutation testing with multiple comparison correction ($p_{FDRcorrected} = 0.0899$; $p_{uncorrected} = 0.0089$). None of the ICA networks resulted in AUC > 0.7 with external validation (online Supplementary Table S9).

Multimodal neuroimaging

Classification combining voxel-wise GM with ICA-based FC led to the best performing model, with 0.83 AUC using internal validation and 0.70 AUC using external validation. Classifications for voxel-wise GM with the Power-atlas FC led to similar performances of 0.82 AUC for internal validation and 0.73 AUC for external validation. All of the aforementioned AUCs were statistically significant for both internal and external validation (Fig. 3*b*). Classification performance for regional NMM with ICA-based FC resulted in 0.75 AUC with internal validation and 0.51 AUC for external validation. Classifications for regional NMM with Power-atlas FC led to 0.67 AUC using internal validation and 0.55 AUC for external validation. AUCs obtained for classifications using regional neuromorphometrics and FC were statistically significant for internal validation but not for external validation (online Supplementary Table S10).

Response and post-ECT severity prediction

As previous predictive studies and clinical trials have used both remission (HAM-D score of ⩽7 after treatment) and response (at least 50% HAM-D score reduction compared to baseline) as outcome criterion, we additionally assessed response classification performance. Of the 189 included patients, 113 patients were ECT responders and 76 non-responders. Demographics of the included sample and results for response prediction are provided in online Supplementary Results (online Supplementary Figs. S2 and S3; online Supplementary Tables S2, S11–S14). To summarize, the majority of the classification models performed poorly with AUC < 0.7 with internal validation, and none of the models remained significant with external validation after permutation testing with FDR correction.

Additionally, we investigated whether regression could be used to directly predict post-ECT severity (i.e. HAM-D score) outcomes. To this end, we applied support vector regression (SVR) on data used for our best performing classification model (voxel-wise GM combined with ICA-based FC on the three largest samples with internal validation). The entire machine learning procedure, including nested grid-search for hyperparametrization, was performed identically as above. The results showed a positive significant correlation ($r = 0.33$, $p < 0.001$) between the predicted and post-ECT HAM-D scores (online Supplementary Fig. S4). However, other regression performance metrics like the R2 score (0.092) and mean absolute error (5.73 HAM-D scores) indicated poor overall performance.

Learning curves

To evaluate the relation between training sample size and classification performance, we examined learning curves for the best performing models (i.e. remission classification using data from the three largest centers) by using different proportions of training data. Classification accuracy reached 0.83–0.84 AUC for unimodal (voxel-wise GM) and multimodal (voxel-wise GM and ICA-based FC) classifiers, with averaged AUC > 0.75 for classifications using 50% of data for training ($N = 55$) and AUC > 0.8 for 70% of data used for training ($N = 76$). See online Supplementary Fig. S5 for full learning curves. Both learning curves did not appear saturated, suggesting that model performance could still increase when using larger training samples.

Anatomical localization

We investigated which brain regions contributed most to treatment classification for the best performing unimodal and multimodal models. Voxel-wise GM data from the three largest samples resulted in the best unimodal classification with an AUC of 0.83. The obtained feature importance *p* values were plotted for GM weights only as we were interested in brain regions rather than the influence of covariates. As shown in Fig. 4, regions located in dorsomedial prefrontal (dmPFC), precuneus and thalamus exhibited high contribution to the classification task. The sign of weights within thalamus was mostly negative, implying a high chance for non-remission classification, whereas signs of weights within dmPFC and precuneus were mostly positive, implying a high chance for remission classification. Note that these results reflected the contribution of these brain regions to the multivariate pattern used by the SVM classifier.

Next, we investigated the most contributing brain regions in our best performing multimodal model (using a combination of voxel-wise GM and ICA-based FC). The resulting *p* values obtained for multimodal voxel-wise GM were visually identical to those obtained from the unimodal model described above (online Supplementary Fig. S6) and highly correlated to each other ($r = 0.99$, $p < 0.001$) with a dice similarity coefficient of 0.92. Similarly, the significant feature importances for ICA-based FC in both the unimodal (online Supplementary

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

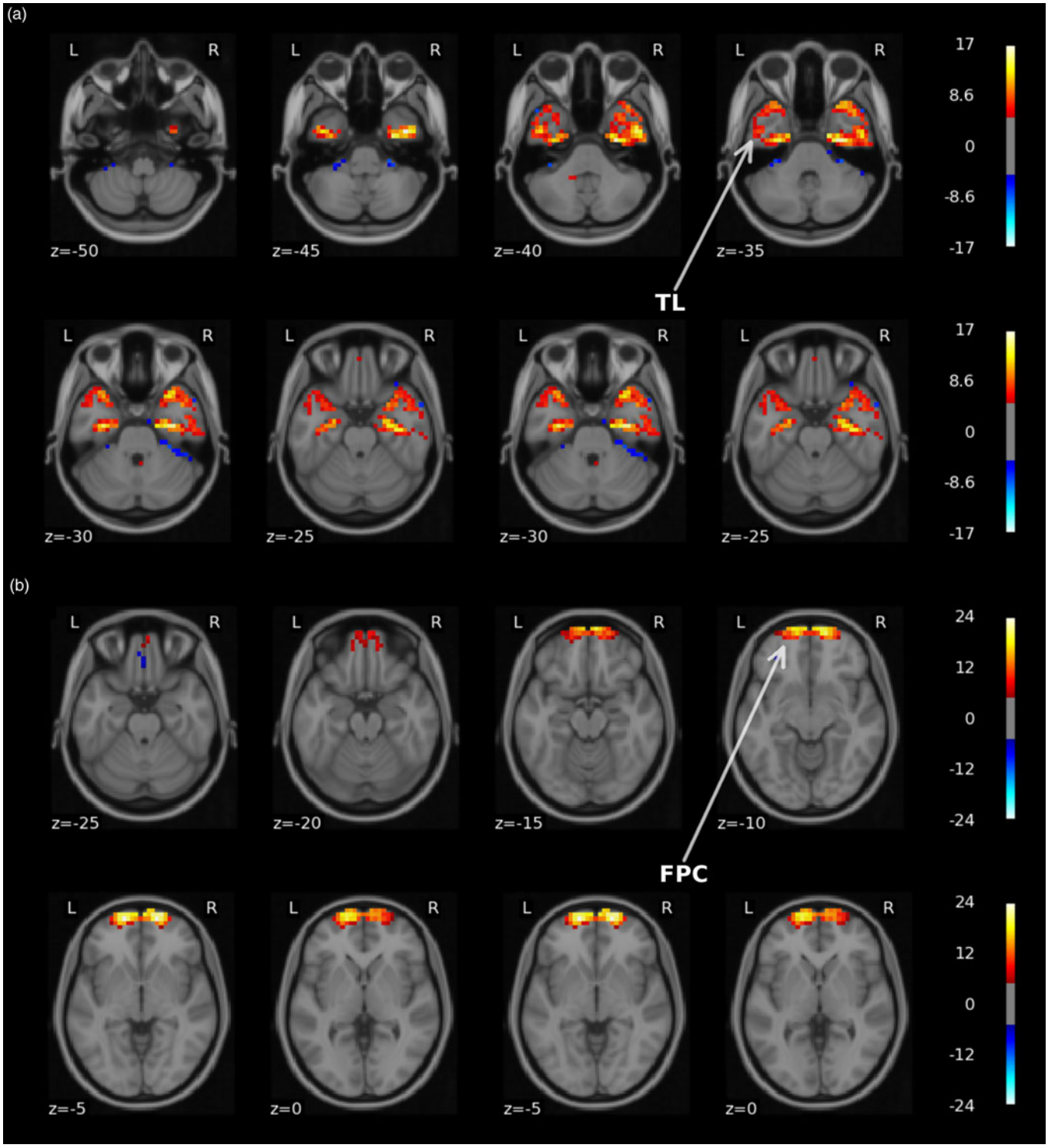

Figure 2. Visual representation of the two UK BioBank group ICA spatial components that led to AUC > 0.75 for either response or remission classification. Top panel *a* depicts a network (#42) centered around the temporal lobes (TL). The second panel *b* shows a network (#52) located in frontopolar cortex (FPC). Images are thresholded at $Z \geqslant 5$ and overlaid on a standard 2 mm MNI template. The figure was made with the nilearn package (http://nilearn.github.io).

Fig. S7) and multimodal (online Supplementary Fig. S8) models were widespread, significantly correlated ($r = 0.75$, $p < 0.001$) but showed less similarity to each other (dice coefficient = 0.32). These findings seem to indicate that most of the GM features were retained in the multimodal approach, whereas the similarity between important features used for the ICA-based FC was lower.

Discussion

The presented results show that neuroimaging data can provide a good prediction of ECT remission for individual patients across different centers. In line with recent meta-analyses, older age and presence of psychotic symptoms at baseline were associated

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

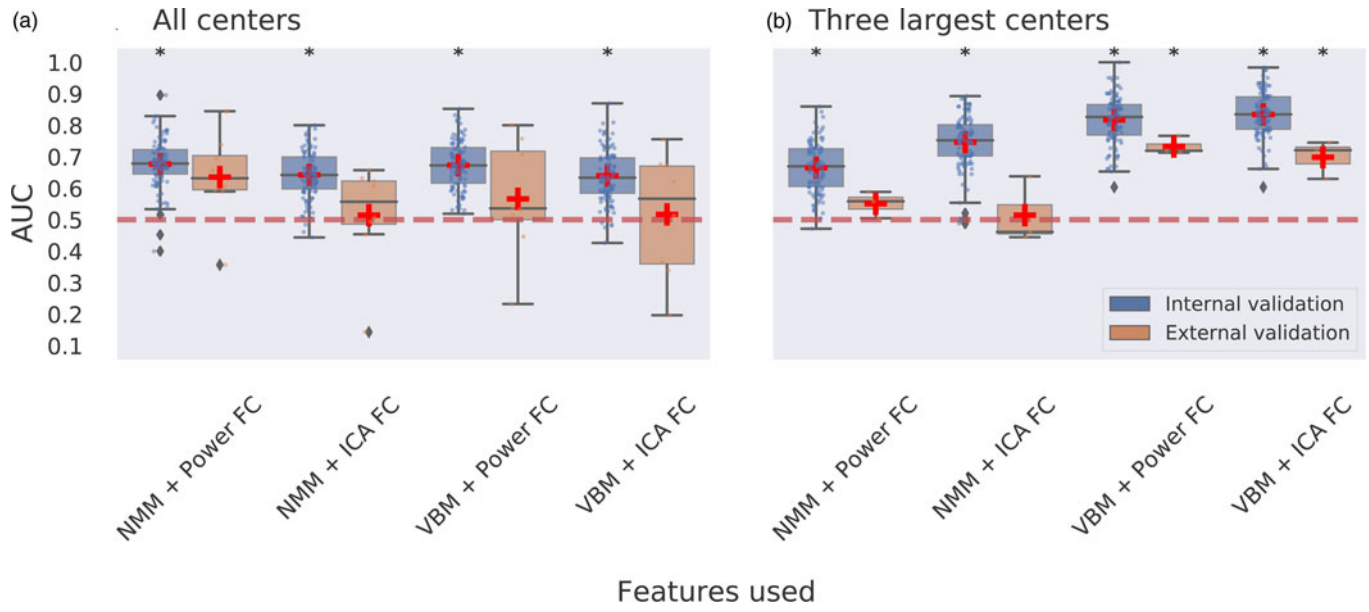

Figure 3. Multimodal multicenter predictions for ECT treatment remission. Panel *a* depicts classification performance using data from all centers and different combinations of features with internal validation (AUC is averaged over 100 stratified cross-validation splits) and external validation (leave-one-site-out cross-validation, scores are averaged across different centers left out for model testing). Panel *b* depicts classification performance using data from the three largest centers with internal and external validation. VBM, voxel-based morphometry; NMM, Neuromorphometrics atlas; FC, functional connectivity; ICA, group information guided independent component analysis. Red dashed line depicts chance-level performance (0.5 AUC). Asterisks indicate significant difference from chance level after permutation testing with false discovery rate correction for multiple comparisons ($p < 0.05$, corrected).

with better ECT outcome (van Diermen et al., 2018). However, our classification results show that this information is not sufficient for making individual predictions, highlighting the relevance of obtaining neuroimaging data for accurate predictions. Remission classification using a combination of voxel-wise GM and clinical data with either ICA-based FC or Power-atlas-based FC led to models with excellent discrimination when trained and tested on samples coming from each center (internal validation AUC > 0.8) and remained acceptable when validated on completely new data from other centers (external validation AUC > 0.7) (Hosmer et al., 2013). These results indicate that multimodal neuroimaging data may provide a biomarker that could be used to guide clinical decision making.

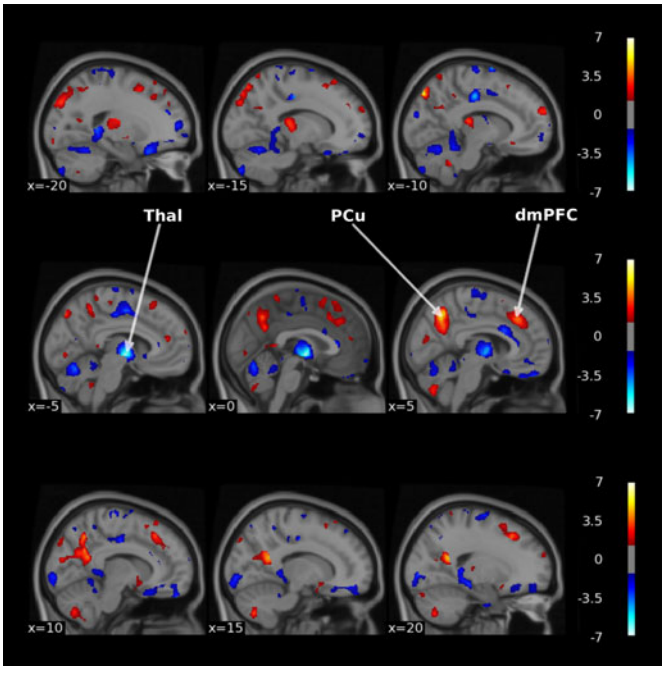

Figure 4. Thresholded-log(p) value maps characterizing the regions important for the treatment remission classification using voxel-wise GM data of the three largest centers (thresholded at $p < 0.05$ uncorrected). Hot colors indicate positive weights and cold colors indicate negative weights of the SVM. Thal, thalamus, PCu, Precuneus, dmPFC, dorsomedial prefrontal cortex. The figure was made with the nilearn package (http://nilearn.github.io).

Feature importances obtained for voxel-wise GM in the unimodal and multimodal approach were visually identical and more similar compared to those obtained for unimodal and multimodal ICA-based FC, which could indicate that the multimodal classification was mainly driven by GM features. Although the best performing multimodal model showed higher performance compared to the best unimodal model using voxel-wise GM, the observed differences were small. Nonetheless, the multimodal models showed higher overall performance with external validation, which could translate into more clinical utility when deployed on entirely unseen sites. Future research should investigate whether more advanced data fusion approaches could further improve the performance of multimodal MRI data over GM only. Given the costs and human labor associated with MRI acquisition, the added value compared to conventional diagnostic tools (e.g. structured interviews and questionnaires) needs to be further evaluated. By providing patients and clinicians a patient-specific prognosis, this could ultimately increase the success rate of ECT, avoid ineffective treatments and accompanying adverse effects, and increase the use of the most effective antidepressive treatment available.

Our findings show that the prediction of treatment response was poor, while prediction of remission was good (see online Supplementary Results). This indicates that ECT outcome

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

prediction is limited to remission and that the remission group can be best differentiated from the other patients. Remission may provide a better outcome criterion than response and has become the gold standard for depression treatment, because patients who do not remit have a poorer prognosis and greater chance of relapse and recurrence than those who do (McIntyre & O'Donovan, 2004; van Diermen et al., 2018). Remission is also associated with a lower full symptomatic recurrence rate compared with achieving treatment response (McIntyre & O'Donovan, 2004). Furthermore, while unimodal and multimodal models performed comparable for remission classification using data from the largest centers with internal validation, only the multimodal classifications remained acceptable with external validation on different centers. We speculate that multimodal data may increase the probability that either the structural or functional MRI data overlaps across centers.

Previous monocenter studies using neuroimaging data to predict ECT outcome with either structural or functional MRI were able to obtain up to 0.83 AUC (Cohen et al., 2021). Here, we achieved similar classification performance in a multicenter setting. Using data from different samples involves many additional sources of technological (e.g. different MR hardware and scanner protocols) and clinical (e.g. different ECT protocols, patient cohort, and recruitment procedures) variability (Schnack & Kahn, 2016). These additional sources of variability may decrease prediction accuracy of MRI measurements for ECT outcome (Schnack & Kahn, 2016). Conversely, a multicenter study avoids cohort-specific solutions and so helps test generalizability of the results across different samples, increasing the likelihood that features identified as discriminatory between remitters and non-remitters reflect generic properties related to treatment outcome across datasets. Our results showed that generalizability to new samples came at the cost of lower accuracy, as best performing classifications with internal validation (AUC $\approx$ 0.83) outperformed those using external validation (AUC $\approx$ 0.73). Additionally, we found that using a subsample of the data containing three centers with $N \geqslant 20$ each ($N = 109$) led to better model performance compared to using all seven centers ($N = 189$). This improvement could not be attributed solely to reduced clinical heterogeneity, as differences in sample demographics and clinical characteristics between the three largest centers were found to be similar to those seen in the entire sample (online Supplementary Tables S5 and S6). We therefore hypothesize that the exclusion of smaller centers ensured that the model had sufficient examples per center for training. With regard to clinical heterogeneity specifically, our sample included patients with unipolar (UP) and bipolar (BP) depression. We evaluated whether the classification performance was similar between UP and BP patients for our best performing model. Both patient groups showed good classification performance: metrics calculated for BP only resulted in a balanced accuracy of 0.64 and AUC of 0.74, whereas results for MDD showed a balanced accuracy of 0.74 and AUC of 0.84. We expect that the higher performance for MDD is due to a better representation of this group in the training set. As ECT is indicated for both mood disorders (Bayes & Parker, 2018; Parker, Graham, & Tavella, 2017), we chose to develop a predictive model that includes both indications.

Brain regions that contributed most to remission classification using sMRI data included the dmPFC, precuneus, and thalamus. The dmPFC is involved in the top-down process that regulates many emotional and cognitive functions (Bai et al., 2019; Zhang et al., 2021). Clinical and preclinical studies on depression have consistently reported functional, structural, and system-level abnormalities that span many PFC regions (Pizzagalli & Roberts, 2022; Price & Drevets, 2012). Interestingly, the dorsomedial part of the PFC is integrated in the three large scale functional networks that constitute the triple network model implicated in depression (i.e. central executive, default mode, and salience network) (Menon, 2011; Sheline, Price, Yan, & Mintun, 2010; van Waarde et al., 2015). The dmPFC cluster found important for remission classification (displayed in Fig. 4) partly overlaps with the salience network, and dysfunction in this network is associated with emotional dysregulation and negatively biased information processing in depression (Hamilton et al., 2016; Menon, 2011). Neuroimaging studies investigating changes in brain structure and connectivity following ECT have reported significant correlations between the dmPFC and an overall reduction in depressive symptom scores (Bai et al., 2019; Dukart et al., 2014; Li et al., 2022; Perrin et al., 2012; Zhang et al., 2021), as well as associations between the dmPFC and remission specifically, in line with our findings (Abbott et al., 2013; van Waarde et al., 2015). Our results also pointed to an important role of the precuneus for remission classification. The precuneus is a posterior region of the medial parietal cortex with widespread connections and is regarded as a major association area important for complex cognition and behavior including visuospatial imagery, episodic memory retrieval, and self-related processing (Cavanna & Trimble, 2006; Mulders et al., 2016; Utevsky, Smith, & Huettel, 2014). Importantly, this region (together with the posterior cingulate cortex) is considered the core functional hub of the aforementioned default mode network, and dysregulations within this network have been associated with symptom severity, and in particularly with rumination, a core characteristic of depression (Bai et al., 2019; Li et al., 2022; Peng et al., 2015; Utevsky et al., 2014; Zhong, Pu, & Yao, 2016). Preliminary evidence links changes in precuneus structure, network connectivity, and cerebral blood flow with ECT treatment outcome (Leaver et al., 2019; Mulders et al., 2016, 2020). Previous studies using structural or functional MRI to predict ECT outcome have also implicated the precuneus (Jiang et al., 2018; Li et al., 2022). Our results also pointed to the thalamus, which is a structure embedded within the cortico-striatal-thalamo-cortical (CSTC) pathways that mediate several cognitive, affective, and motivational processes (Sherman, 2016; Takamiya et al., 2019). It acts as a central hub within the wider limbic-cortical-striatal-pallidal-thalamic circuit (overlapping with the salience network) that is thought to play an important role in emotion dysregulation (Drevets, Price, & Furey, 2008) and shares direct anatomical connections with regions implicated in mood disorders like the hippocampus (also associated with the default mode network) (Hamilton, Farmer, Fogelman, & Gotlib, 2015; Leaver, Espinoza, Wade, & Narr, 2022; Price & Drevets, 2012; Sherman, 2016). There is evidence of decreased thalamic volume and hyperactivity during rest and emotional processing in depression (Arnone et al., 2016; Bora, Harrison, Davey, Yucel, & Pantelis, 2012; Palmer, Crewther, Carey, & Team, 2014; Price & Drevets, 2012), and it has been suggested that abnormal thalamic connectivity may lead to disruptions in higher-order cortico-cortical connectivity (Gallo et al., 2023). The thalamus is also considered an important structure in seizure physiology, and several studies have associated pre- *v.* post-ECT changes in thalamic volume, functional connectivity and cerebral blood flow with ECT efficacy (Leaver et al., 2016, 2019; Sun et al., 2020; Takamiya et al., 2019). A recent

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

review proposes a circuit-level model for the mechanisms underlying ECT, in which repeated seizure therapy improves symptoms by correcting or resetting the disrupted limbic-cortical-striatal-pallidal-thalamic circuit in depression (Leaver et al., 2022). It is thought that seizure propagation between distant brain regions through cortical–thalamocortical and direct cortical–cortical connections is pivotal for ECT effectiveness (Fink & Ottosson, 1980; Leaver et al., 2016, 2019; McNally & Blumenfeld, 2004; Singh & Kar, 2017; Takamiya et al., 2018). Previous classification studies using pre-treatment rs-fMRI to predict ECT outcome have reported thalamic connectivity as an important predictor in line with these findings (Sun et al., 2020; Takamiya et al., 2019). Notably, the systematic review of Enneking, Leehr, Dannlowski, and Redlich (2020) compared studies on biomarkers of response for the most common antidepressive treatments, namely antidepressive pharmacotherapy (AD), electroconvulsive therapy, and cognitive-behavioral therapy (CBT) (Enneking et al., 2020). According to this review, pre-treatment GM volume in the thalamus is associated with ECT treatment outcome (in line with our findings) but not for AD and CBT. The precuneus, which was also found to be important for our GM classifications, was not associated with outcome for any of the treatments. However, the anterior cingulate cortex (ACC), which is closely located to dmPFC, emerged as a predictive region for outcome prediction in all three treatments (Enneking et al., 2020). These findings suggest that GM volume in regions surrounding the ACC might indicate broader treatment responsiveness, while specific regions such as the thalamus may be more indicative of treatment outcome to ECT specifically.

Our results further indicated an important role for remission classification using ICA spatial resting-state components centered around the anterior temporal lobes and frontopolar cortex that both resulted in AUC > 0.7. It is important to acknowledge that although using these components for classification yielded acceptable discrimination with above chance-level performance ($p_{uncorrected} < 0.05$), they did not remain significant after applying multiple comparison correction. Consequently, these results should be interpreted with caution. The frontopolar cortex (Brodmann Area 10) plays an important role in integrating cognitive, social, and emotional processes. Its medial parts are mostly associated with affective processing such as emotional and social cognition and its lateral parts with working memory and perception. (Bludau et al., 2014; Gilbert et al., 2006). Previous studies have reported that depression is associated to reduced medial frontal pole volume (Bludau et al., 2016), and that decreased frontal pole volume and FC following ECT was related to therapeutic efficacy (Xu et al., 2018). The medial temporal lobes have been consistently implicated in ECT neuroimaging research and include the hippocampus and amygdala, which have shown to undergo structural changes in volume, functional connectivity, and perfusion following ECT (Leaver et al., 2016; Mulders et al., 2020; Ousdal et al., 2020; Redlich et al., 2016; Takamiya et al., 2018; Wilkinson, Sanacora, & Bloch, 2017). The temporal lobes also show the highest magnitude of electrical current in right unilateral stimulation (Fridgeirsson, Deng, Denys, van Waarde, & van Wingen, 2021), and increased electrical field strength has been associated with increased right hippocampal neuroplasticity and improved antidepressant outcomes (Deng et al., 2021). Smaller hippocampal volumes, and to a lesser extend the amygdala, are apparent in MDD patients and support the current hypothesis that mood disorders consist of dysfunction in neural circuits important for processing and integrating emotional and cognitive events (Kempton et al., 2011; Schmaal et al., 2016). Previous classification studies have also highlighted anterior lateral temporal lobe volume, hippocampal and amygdala gray matter (Abbott et al., 2014; Jiang et al., 2018; Takamiya et al., 2020), and temporal cortex functional connectivity (Leaver et al., 2018) as important predictors for ECT outcome.

Altogether, these results provide evidence for the importance of dmPFC, thalamic and precuneus structure and fronto-temporal FC for both depression and ECT-related clinical response. Notably, the identification of brain regions contributing most to the classification resulted from a multivariate analysis, and the localization of these regions should therefore be interpreted with caution as these regions may not only be related to treatment outcome but also contribute to denoising during the classification process (Haufe et al., 2014).

Several limitations have to be taken into account when interpreting our findings. First, our models were trained on both medicated and unmedicated patients. Medication was usually tapered before ECT or kept stable during ECT, but was not consistently registered to enable medication-specific analyses. The current study design also did not include another treatment (as control condition), and therefore we do not know whether the predictive markers are specific for ECT. Future studies on ECT prediction may explore the effects of concurrent medication use and whether models do (not) share features with predictive models for other treatments for treatment resistant depression such as ketamine or deep brain stimulation. Second, although the sample size in this study is higher than those typically seen in previous neuroimaging studies predicting ECT outcome, it is likely that including more patients for classifier training would increase the robustness and performance of the models (see learning curve; online Supplementary Fig. S2). In addition, our classifiers were trained using a relatively high ratio of features to participants that could lead to potential model overfitting, referred to as the 'curse of dimensionality'. We employed regularization to avoid overfitting by tuning the 'C' parameter of the SVM (through grid-search). However, it is possible that some overfitting might still occur, and only larger sample sizes can guarantee a lower bias towards overfitting. Nonetheless, our best performing model showed significant classification performance with external validation, indicating that the model performs above chance level when applied to data unseen centers and is unlikely to overfit on data from a single center. Future studies using even larger samples should further investigate the feasibility of using MRI data to predict ECT outcome. Next, we used a retrospectively pooled sample from existing data across the world, without harmonized protocols for scanning, inclusion criteria or demographic and clinical characteristics. Not surprisingly, we found significant differences in sample demographics and clinical characteristics between the different data collection centers. These sources of heterogeneity may limit classification performance but also provide an opportunity for model development using independent data sets and the discovery of generalizable biomarkers that are reproducible across centers. However, classification performance might be improved by using standardized acquisition parameters for possible future clinical utility. Finally, it should be mentioned that artificial dichotomization of post-ECT scores to remission and response rates leads to some loss of information. For example, patients that show partial remission or response (i.e. HAM-D scores of 7% or 49% reduction in symptoms) are considered the same as those patients that do not improve at all. From a clinical perspective, regression-based approaches that allow for continuous

https://doi.org/10.1017/S0033291723002040 Published online by Cambridge University Press

predictions of symptom reduction might provide more clinical utility in comparison to binary classification (Gartner et al., 2021). We investigated whether this was the case for the data in our study, and although we obtained a significant correlation between the predicted and true post-ECT HAM-D scores, other regression performance metrics like the R2 score and mean absolute error indicated poor overall performance. Future studies could further investigate the feasibility of using other models that provide continuous predictions (e.g. using deep learning) for ECT outcome in larger, multi-samples.

Taken together, this study suggests that ECT remission can be predicted with acceptable discrimination using MRI data in a large, ecologically valid, multicenter sample of patients receiving ECT, indicating that future development of a clinical decision support tool might be feasible. MRI could easily be incorporated during decision making, as ECT is typically provided in a hospital setting. And as MRI is inexpensive compared to ECT, the additional costs are expected to outweigh the costs of unsuccessful treatments.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291723002040.

Acknowledgements. We would like to thank the logistic and academic support of the entire GEMRIC consortium. The full overview of the GEMRIC board members can be found here: https://mmiv.no/gemric/. In addition, we would like to acknowledge Louise Emsell for contributing data that was considered but not used for the final analysis.

Author contributions. W. B. B., G. A. v. W. & A. D. conceived the research question, contributed to the design, and wrote the first draft of the manuscript. W. B. B. performed data analysis and developed MRI processing pipelines. G. A. v. W. and P. Z. made substantial contributions to data analysis. L. O. coordinated the GEMRIC consortium. W. B. B., L. O., H. B., C. C. A., M. A., T. B., J. A. C., S. C., R. E., P. C. R. M., K. L. N., M. L. O., D. R., F. t. D., I. T., P. v. E., E. v. E., M. v. V., J. v. W., A. D. & G. A. v. W. contributed to data acquisition. W. B. B., P. Z., L. O., H. B., C. C. A., M. A., T. B., J. A. C., S. C., R. E., P. C. R. M., K. L. N., M. L. O., D. R., F. t. D., I. T., P. v. E., E. v. E., M. v. V., B. W., J. v. W., A. D., and G. A. v. W. substantially revised the manuscript. All authors critically revised the manuscript for important intellectual content, approved the final draft and had final responsibility for the decision to submit for publication.

Financial support. This work was supported by the Netherlands Organization for Scientific Research (NWO/ZonMW Vidi 917.15.318, Dr van Wingen), Western Norway Regional Health Authority (Grant No. 91223, Dr Oltedal), NARSAD Young Investigator Grant (No. 27786 to BW), a K99 Pathway to Independence Award (Grant No. MH119314 to BW), and the National Institute of Mental Health (Grant No. MH092301 and MH110008 for Dr Narr and Dr Espinoza; MH111826 and MH125126 for Dr Abbott; MH119616 for Dr Argyelan and R01MH112737 for Dr Camprodon).

Conflict of interest. Dr van Wingen has received research grant support from Philips. Dr Camprodon serves in the Scientific Advisory Board of Hyka Therapeutics and Feelmore Labs and has been a consultant for Neuronetics. All other individually named co-authors in the GEMRIC working group declared no biomedical financial interests or potential conflicts of interest. A preprint version of this manuscript has been posted on medRxiv (https://doi.org/10.1101/2021.07.29.21261206); ID: 2021.07.29.21261206).

Role of the funding source. The funders of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data sharing statement. Individual participant data cannot be made available publicly because there is no consent or ethical approval for this and the data cannot be anonymized. The data are stored on a secure centralized server at the University of Bergen, Norway. Participating GEMRIC sites have access to the raw data according to specific data policy and safety rules of the consortium and in accord with the approval from the ethical committee. The GEMRIC consortium welcomes new members who are interested in the neuroimaging research of ECT. For more about the application process, please visit https://mmiv.no/how-to-join-gemric/ or write to Leif Oltedal (leif.oltedal@uib.no). General information about the consortium can be found on the following website: https://mmiv.no/gemric/.

Code availability. The code that support the findings of this study are available from the corresponding authors upon reasonable request.

References

Abbott, C. C., Jones, T., Lemke, N. T., Gallegos, P., McClintock, S. M., Mayer, A. R., … Calhoun, V. D. (2014). Hippocampal structural and functional changes associated with electroconvulsive therapy response. *Translational Psychiatry*, *4*, e483. doi:10.1038/tp.2014.124.

Abbott, C. C., Lemke, N. T., Gopal, S., Thoma, R. J., Bustillo, J., Calhoun, V. D., & Turner, J. A. (2013). Electroconvulsive therapy response in major depressive disorder: A pilot functional network connectivity resting state FMRI investigation. *Frontiers in Psychiatry*, *4*, 10. doi:10.3389/fpsyt.2013.00010.

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage*, *147*, 736–745. doi:10.1016/j.neuroimage.2016.10.045.

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., … Smith, S. M. (2018). Image processing and Quality Control for the first 10000 brain imaging datasets from UK Biobank. *Neuroimage*, *166*, 400–424. doi:10.1016/j.neuroimage.2017.10.034.

Arnone, D., Job, D., Selvaraj, S., Abe, O., Amico, F., Cheng, Y., … McIntosh, A. M. (2016). Computational meta-analysis of statistical parametric maps in major depression. *Human Brain Mapping*, *37*(4), 1393–1404. doi:10.1002/hbm.23108.

Bai, T., Wei, Q., Zu, M., Xie, W., Wang, J., Gong-Jun, J., … Wang, K. (2019). Functional plasticity of the dorsomedial prefrontal cortex in depression reorganized by electroconvulsive therapy: Validation in two independent samples. *Human Brain Mapping*, *40*(2), 465–473. doi:10.1002/hbm.24387.

Bayes, A. J., & Parker, G. B. (2018). Comparison of guidelines for the treatment of unipolar depression: A focus on pharmacotherapy and neurostimulation. *Acta Psychiatrica Scandinavica*, *137*(6), 459–471. doi:10.1111/acps.12878.

Bludau, S., Bzdok, D., Gruber, O., Kohn, N., Riedl, V., Sorg, C., … Eickhoff, S. B. (2016). Medial prefrontal aberrations in major depressive disorder revealed by cytoarchitectonically informed voxel-based morphometry. *American Journal of Psychiatry*, *173*(3), 291–298. doi:10.1176/appi.ajp.2015.15030349.

Bludau, S., Eickhoff, S. B., Mohlberg, H., Caspers, S., Laird, A. R., Fox, P. T., … Amunts, K. (2014). Cytoarchitecture, probability maps and functions of the human frontal pole. *Neuroimage*, *93*, 260–275. doi:10.1016/j.neuroimage.2013.05.052.

Bora, E., Harrison, B. J., Davey, C. G., Yucel, M., & Pantelis, C. (2012). Meta-analysis of volumetric abnormalities in cortico-striatal-pallidal-thalamic circuits in major depressive disorder. *Psychological Medicine*, *42*(4), 671–681. doi:10.1017/S0033291711001668.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(Pt 3), 564–583. doi:10.1093/brain/awl004.

Cohen, S. E., Zantvoord, J. B., Wezenberg, B. N., Bockting, C. L. H., & van Wingen, G. A. (2021). Magnetic resonance imaging for individual prediction of treatment response in major depressive disorder: A systematic review and meta-analysis. *Translational Psychiatry*, *11*(1), 168. doi:10.1038/s41398-021-01286-x.

Deng, Z. D., Argyelan, M., Miller, J., Quinn, D. K., Lloyd, M., Jones, T. R., … Abbott, C. C. (2021). Electroconvulsive therapy, electric field, neuroplasticity, and clinical outcomes. *Molecular Psychiatry*, *27*(3), 1676–1682. 10.1038/s41380-021-01380-y.

Drevets, W. C., Price, J. L., & Furey, M. L. (2008). Brain structural and functional abnormalities in mood disorders: Implications for neurocircuitry models of depression. *Brain Structure and Function*, *213*(1-2), 93–118. doi:10.1007/s00429-008-0189-x.

Du, Y., & Fan, Y. (2013). Group information guided ICA for fMRI data analysis. *Neuroimage*, *69*, 157–197. doi:10.1016/j.neuroimage.2012.11.008.

Dukart, J., Regen, F., Kherif, F., Colla, M., Bajbouj, M., Heuser, I., … Draganski, B. (2014). Electroconvulsive therapy-induced brain plasticity determines therapeutic outcome in mood disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), 1156–1161. doi:10.1073/pnas.1321399111.

Enneking, V., Leehr, E. J., Dannlowski, U., & Redlich, R. (2020). Brain structural effects of treatments for depression and biomarkers of response: A systematic review of neuroimaging studies. *Psychological Medicine*, *50*(2), 187–209. doi:10.1017/S0033291719003660.

Fink, M., & Ottosson, J. O. (1980). A theory of convulsive therapy in endogenous depression: Significance of hypothalamic functions. *Psychiatry Research*, *2*(1), 49–61. doi:10.1016/0165-1781(80)90006-2.

Fridgeirsson, E. A., Deng, Z. D., Denys, D., van Waarde, J. A., & van Wingen, G. A. (2021). Electric field strength induced by electroconvulsive therapy is associated with clinical outcome. *NeuroImage: Clinical*, *30*, 102581. doi:10.1016/j.nicl.2021.102581.

Gallo, S., El-Gazzar, A., Zhutovsky, P., Thomas, R. M., Javaheripour, N., Li, M., … van Wingen, G. (2023). Functional connectivity signatures of major depressive disorder: Machine learning analysis of two multicenter neuroimaging studies. *Molecular Psychiatry*, 1–10. 10.1038/s41380-023-01977-5.

Gaonkara, B., Shinohara, R. T., & Davatzikos, C. (2016). Interpreting support vector machine models for multivariate group wise analysis in neuroimaging. *Medical Image Analysis*, *1848*, 3047–3054. doi:10.1016/j.bbamem.2015.02.010.Cationic.

Gartner, M., Ghisu, E., Herrera-Melendez, A. L., Koslowski, M., Aust, S., Asbach, P., … Bajbouj, M. (2021). Using routine MRI data of depressed patients to predict individual responses to electroconvulsive therapy. *Experimental Neurology*, *335*, 113505. doi:10.1016/j.expneurol.2020.113505.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): A meta-analysis. *Journal of Cognitive Neuroscience*, *18*(6), 932–948. doi:10.1162/jocn.2006.18.6.932.

Hamilton, J. P., Farmer, M., Fogelman, P., & Gotlib, I. H. (2015). Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biological Psychiatry*, *78*(4), 224–230. doi:10.1016/j.biopsych.2015.02.020.

Hamilton, J. P., Glover, G. H., Bagarinao, E., Chang, C., Mackey, S., Sacchet, M. D., & Gotlib, I. H. (2016). Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. *Psychiatry Research: Neuroimaging*, *249*, 91–96. doi:10.1016/j.pscychresns.2016.01.016.

Haq, A. U., Sitzmann, A. F., Goldman, M. L., Maixner, D. F., & Mickey, B. J. (2015). Response of depression to electroconvulsive therapy: A meta-analysis of clinical predictors. *Journal of Clinical Psychiatry*, *76*(10), 1374–1384. doi:10.4088/JCP.14r09528.

Haufe, S., Meinecke, F., Gorgen, K., Dahne, S., Haynes, J. D., Blankertz, B., & Biessmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, *87*, 96–110. doi:10.1016/j.neuroimage.2013.10.067.

Heijnen, W. T., Birkenhager, T. K., Wierdsma, A. I., & van den Broek, W. W. (2010). Antidepressant pharmacotherapy failure and response to subsequent electroconvulsive therapy: A meta-analysis. *Journal of Clinical Psychopharmacology*, *30*(5), 616–619. doi:10.1097/JCP.0b013e3181ee0f5f.

Hosmer D. W. Jr, Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. *398*). New York, NY: John Wiley & Sons.

Jiang, R., Abbott, C. C., Jiang, T., Du, Y., Espinoza, R., Narr, K. L., … Calhoun, V. D. (2018). SMRI Biomarkers predict electroconvulsive treatment outcomes: Accuracy with independent data sets. *Neuropsychopharmacology*, *43*(5), 1078–1087. doi:10.1038/npp.2017.165.

Kempton, M. J., Salvador, Z., Munafo, M. R., Geddes, J. R., Simmons, A., Frangou, S., & Williams, S. C. (2011). Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Archives of General Psychiatry*, *68*(7), 675–690. doi:10.1001/archgenpsychiatry.2011.60.

Kottas, M., Kuss, O., & Zapf, A. (2014). A modified Wald interval for the area under the ROC curve (AUC) in diagnostic case-control studies. *BMC Medical Research Methodology*, *14*, 26. doi:10.1186/1471-2288-14-26.

Leaver, A. M., Espinoza, R., Pirnia, T., Joshi, S. H., Woods, R. P., & Narr, K. L. (2016). Modulation of intrinsic brain activity by electroconvulsive therapy in major depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(1), 77–86. doi:10.1016/j.bpsc.2015.09.001.

Leaver, A. M., Espinoza, R., Wade, B., & Narr, K. L. (2022). Parsing the network mechanisms of electroconvulsive therapy. *Biological Psychiatry*, *92*(3), 193–203. doi:10.1016/j.biopsych.2021.11.016.

Leaver, A. M., Vasavada, M., Joshi, S. H., Wade, B., Woods, R. P., Espinoza, R., & Narr, K. L. (2019). Mechanisms of antidepressant response to electroconvulsive therapy studied with perfusion magnetic resonance imaging. *Biological Psychiatry*, *85*(6), 466–476. doi:10.1016/j.biopsych.2018.09.021.

Leaver, A. M., Wade, B., Vasavada, M., Hellemann, G., Joshi, S. H., Espinoza, R., & Narr, K. L. (2018). Fronto-Temporal connectivity predicts ECT outcome in major depression. *Frontiers in Psychiatry*, *9*, 92. doi:10.3389/fpsyt.2018.00092.

Li, Y., Yu, X., Ma, Y., Su, J., Li, Y., Zhu, S., … Wang, J. (2022). Neural signatures of default mode network in major depression disorder after electroconvulsive therapy. *Cerebral Cortex*, *33*(7), 3840–3852. 10.1093/cercor/bhac311.

McIntyre, R. S., & O'Donovan, C. (2004). The human cost of not achieving full remission in depression. *Canadian Journal of Psychiatry*, *49*(3 Suppl 1), 10S–16S. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15147032.

McNally, K. A., & Blumenfeld, H. (2004). Focal network involvement in generalized seizures: New insights from electroconvulsive therapy. *Epilepsy and Behavior*, *5*(1), 3–12. doi:10.1016/j.yebeh.2003.10.020.

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506. doi:10.1016/j.tics.2011.08.003.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., … Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1–73. doi:10.7326/M14-0698.

Mulders, P. C., Llera, A., Beckmann, C. F., Vandenbulcke, M., Stek, M., Sienaert, P., … Tendolkar, I. (2020). Structural changes induced by electroconvulsive therapy are associated with clinical outcome. *Brain Stimulation*, *13*(3), 696–704. doi:10.1016/j.brs.2020.02.020.

Mulders, P. C., van Eijndhoven, P. F., Pluijmen, J., Schene, A. H., Tendolkar, I., & Beckmann, C. F. (2016). Default mode network coherence in treatment-resistant major depressive disorder during electroconvulsive therapy. *Journal of Affective Disorders*, *205*, 130–137. doi:10.1016/j.jad.2016.06.059.

Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classier performance. *Journal of Machine Learning Research*, *11*, 1833–1863. doi:10.1109/ICDM.2009.108.

Oltedal, L., Bartsch, H., Sørhaug, O. J. E., Kessler, U., Abbott, C., Dols, A., … Oedegaard, K. J. (2017). The Global ECT-MRI Research Collaboration (GEMRIC): Establishing a multi-site investigation of the neural mechanisms underlying response to electroconvulsive therapy. *NeuroImage: Clinical*, *14*, 422–432. doi:10.1016/j.nicl.2017.02.009.

Oltedal, L., Narr, K. L., Abbott, C., Anand, A., Argyelan, M., Bartsch, H., … Dale, A. M. (2018). Volume of the human hippocampus and clinical response following electroconvulsive therapy. *Biological Psychiatry*, *84*(8), 574–581. doi:10.1016/j.biopsych.2018.05.017.

Ousdal, O. T., Argyelan, M., Narr, K. L., Abbott, C., Wade, B., Vandenbulcke, M., … Sienaert, P. (2020). Brain changes induced by electroconvulsive therapy are broadly distributed. *Biological Psychiatry*, *87*(5), 451–461. doi:10.1016/j.biopsych.2019.07.010.

Palmer, S. M., Crewther, S. G., Carey, L. M., & Team, S. P. (2014). A meta-analysis of changes in brain activity in clinical depression. *Frontiers in Human Neuroscience*, *8*, 1045. doi:10.3389/fnhum.2014.01045.

Parker, G. B., Graham, R. K., & Tavella, G. (2017). Is there consensus across international evidence-based guidelines for the management of bipolar

disorder? *Acta Psychiatrica Scandinavica*, *135*(6), 515–526. doi:10.1111/acps.12717.

Peng, D., Liddle, E. B., Iwabuchi, S. J., Zhang, C., Wu, Z., Liu, J., … Fang, Y. (2015). Dissociated large-scale functional connectivity networks of the precuneus in medication-naive first-episode depression. *Psychiatry Research*, *232*(3), 250–256. doi:10.1016/j.pscychresns.2015.03.003.

Perrin, J. S., Merz, S., Bennett, D. M., Currie, J., Steele, D. J., Reid, I. C., & Schwarzbauer, C. (2012). Electroconvulsive therapy reduces frontal cortical connectivity in severe depressive disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5464–5468. doi:10.1073/pnas.1117206109.

Pizzagalli, D. A., & Roberts, A. C. (2022). Prefrontal cortex and depression. *Neuropsychopharmacology*, *47*(1), 225–246. doi:10.1038/s41386-021-01101-7.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678. doi:10.1016/j.neuron.2011.09.006.

Price, J. L., & Drevets, W. C. (2012). Neural circuits underlying the pathophysiology of mood disorders. *Trends in Cognitive Sciences*, *16*(1), 61–71. doi:10.1016/j.tics.2011.12.011.

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage*, *112*, 267–277. doi:10.1016/j.neuroimage.2015.02.064.

Redlich, R., Opel, N., Grotegerd, D., Dohm, K., Zaremba, D., Burger, C., … Dannlowski, U. (2016). Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry*, *73*(6), 557–564. doi:10.1001/jamapsychiatry.2016.0316.

Schmaal, L., Veltman, D. J., van Erp, T. G., Samann, P. G., Frodl, T., Jahanshad, N., … Hibar, D. P. (2016). Subcortical brain alterations in major depressive disorder: Findings from the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry*, *21*(6), 806–812. doi:10.1038/mp.2015.69.

Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Frontiers in Psychiatry*, *7*, 50. doi:10.3389/fpsyt.2016.00050.

Sheline, Y. I., Price, J. L., Yan, Z., & Mintun, M. A. (2010). Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 11020–11025. doi:10.1073/pnas.1000446107.

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature Neuroscience*, *19*(4), 533–541. doi:10.1038/nn.4269.

Singh, A., & Kar, S. K. (2017). How electroconvulsive therapy works?: Understanding the neurobiological mechanisms. *Clinical Psychopharmacology and Neuroscience*, *15*(3), 210–221. doi:10.9758/cpn.2017.15.3.210.

Slade, E. P., Jahn, D. R., Regenold, W. T., & Case, B. G. (2017). Association of electroconvulsive therapy with psychiatric readmissions in US hospitals. *JAMA Psychiatry*, *74*(8), 798–804. doi:10.1001/jamapsychiatry.2017.1378.

Sun, H., Jiang, R., Qi, S., Narr, K. L., Wade, B. S., Upston, J., … Sui, J. (2020). Preliminary prediction of individual response to electroconvulsive therapy using whole-brain functional magnetic resonance imaging data. *NeuroImage: Clinical*, *26*, 102080. doi:10.1016/j.nicl.2019.102080.

Takamiya, A., Chung, J. K., Liang, K. C., Graff-Guerrero, A., Mimura, M., & Kishimoto, T. (2018). Effect of electroconvulsive therapy on hippocampal and amygdala volumes: Systematic review and meta-analysis. *British Journal of Psychiatry*, *212*(1), 19–26. doi:10.1192/bjp.2017.11.

Takamiya, A., Kishimoto, T., Liang, K. C., Terasawa, Y., Nishikata, S., Tarumi, R., … Mimura, M. (2019). Thalamic volume, resting-state activity, and their association with the efficacy of electroconvulsive therapy. *Journal of Psychiatric Research*, *117*, 135–141. doi:10.1016/j.jpsychires.2019.08.001.

Takamiya, A., Liang, K. C., Nishikata, S., Tarumi, R., Sawada, K., Kurokawa, S., … Kishimoto, T. (2020). Predicting individual remission after electroconvulsive therapy based on structural magnetic resonance imaging: A machine learning approach. *The Journal of ECT*, *36*(3), 205–210. doi:10.1097/YCT.0000000000000669.

Utevsky, A. V., Smith, D. V., & Huettel, S. A. (2014). Precuneus is a functional core of the default-mode network. *Journal of Neuroscience*, *34*(3), 932–940. doi:10.1523/JNEUROSCI.4227-13.2014.

van Diermen, L., van den Ameele, S., Kamperman, A. M., Sabbe, B. C. G., Vermeulen, T., Schrijvers, D., & Birkenhager, T. K. (2018). Prediction of electroconvulsive therapy response and remission in major depression: Meta-analysis. *British Journal of Psychiatry*, *212*(2), 71–80. doi:10.1192/bjp.2017.28.

van Waarde, J. A., Scholte, H. S., van Oudheusden, L. J., Verwey, B., Denys, D., & van Wingen, G. A. (2015). A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Molecular Psychiatry*, *20*(5), 609–614. doi:10.1038/mp.2014.78.

Wilkinson, S. T., Sanacora, G., & Bloch, M. H. (2017). Hippocampal volume changes following electroconvulsive therapy: A systematic review and meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(4), 327–335. doi:10.1016/j.bpsc.2017.01.011.

Xu, J., Wei, Q., Xu, Z., Hu, Q., Tian, Y., Wang, K., … Wang, J. (2018). Electroconvulsive therapy modulates the structural and functional architecture of frontal pole in major depressive disorder. *Neuropsychiatry*, *08*(01), 213–223. 10.4172/Neuropsychiatry.1000342.

Zhang, T., He, K., Bai, T., Lv, H., Xie, X., Nie, J., … Tian, Y. (2021). Altered neural activity in the reward-related circuit and executive control network associated with amelioration of anhedonia in major depressive disorder by electroconvulsive therapy. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *109*, 110193. doi:10.1016/j.pnpbp.2020.110193.

Zhong, X., Pu, W., & Yao, S. (2016). Functional alterations of fronto-limbic circuit and default mode network systems in first-episode, drug-naive patients with major depressive disorder: A meta-analysis of resting-state fMRI data. *Journal of Affective Disorders*, *206*, 280–286. doi:10.1016/j.jad.2016.09.005.