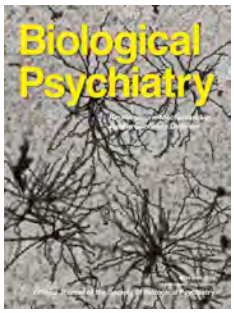


Evaluating Robustness of Brain Stimulation Biomarkers for depression: A Systematic Review of MRI and EEG Studies

Debby Klooster, Helena Voetterl, Chris Baeken, Martijn Arns

PII: S0006-3223(23)01569-X

DOI: https://doi.org/10.1016/j.biopsych.2023.09.009

Reference: BPS 15301

To appear in: *Biological Psychiatry*

Received Date: 12 May 2023

Revised Date: 30 August 2023

Accepted Date: 6 September 2023

Please cite this article as: Klooster D., Voetterl H., Baeken C. & Arns M., Evaluating Robustness of Brain Stimulation Biomarkers for depression: A Systematic Review of MRI and EEG Studies, *Biological Psychiatry* (2023), doi: https://doi.org/10.1016/j.biopsych.2023.09.009.

Evaluating Robustness of Brain Stimulation Biomarkers for depression: A Systematic Review of MRI and EEG Studies

Debby Klooster[1,2,3*], Helena Voetterl[4,5*], Chris Baeken[1,2,6] & Martijn Arns[4,5]

[1] Ghent Experimental Psychiatry Laboratory, Department of Head and Skin, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[2]4BRAIN Team, Department of Head and Skin, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[3] Center for Care and Cure, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

[4] Research Institute Brainclinics, Brainclinics Foundation, Nijmegen, The Netherlands

[5] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

[6] Vrije Universiteit Brussel (VUB), Universitair Ziekenhuis Brussel (UZ Brussel), Department of Psychiatry, Brussels, Belgium

*shared first authorship

Article type: Review

Word count: 3999

Keywords: fMRI; EEG; Depression; rTMS; systematic review; Robustness

Short title: EEG and MRI Biomarkers for Brain Stimulation

Figures: 4

Corresponding author:

Martijn Arns

Brainclinics Foundation

martijn@brainclinics.com

Abstract

Non-invasive brain stimulation (NIBS) treatments have gained considerable attention as a potential therapeutic intervention for psychiatric disorders. The identification of reliable biomarkers for predicting clinical response to NIBS has been a major focus of research in recent years. Neuroimaging techniques, such as electroencephalography (EEG) and (functional) magnetic resonance imaging (fMRI), have been used to identify potential biomarkers that could predict response to NIBS. However, identifying clinically actionable brain biomarkers requires robustness.

In this systematic review, we aimed to summarize the current state of brain biomarker research for NIBS in depression, focusing only on well-powered studies (N≥88) and/or studies that aimed at independently replicating prior findings, either successfully or unsuccessfully. A total of 220 studies were initially identified, of which 18 MRI studies and 18 EEG studies adhered to the inclusion criteria, all focused on repetitive transcranial magnetic stimulation treatment in depression.

After reviewing the included studies, we found the following MRI and EEG biomarkers to be most robust: 1) fMRI-based functional connectivity between the dorsolateral prefrontal cortex and subgenual anterior cingulate cortex, 2) fMRI-based network connectivity, 3) task-induced EEG frontal-midline theta, and 4) EEG individual alpha frequency.

Future prospective studies should further investigate the clinical actionability of these specific EEG and MRI biomarkers to bring biomarkers closer to clinical reality.

Introduction

The search for biomarkers of clinical response to non-invasive brain stimulation (NIBS) treatments has been a major focus of attention over the last decade. Since the introduction of the DSM-5 in 2013 an even stronger focus on biomarker research was ignited by the launch of the National Institute for Mental Health (NIMH) Research Domain Criteria (RDoC) project. A few years later, NIMH made RDoC inclusion mandatory for NIMH funded research, and the term 'personalized medicine' transitioned into the now more frequently used term 'precision psychiatry'. At the same time, some of the largest biomarker studies for major depressive disorder (MDD) emerged, such as the International Study to Predict Optimized Treatment in Depression (iSPOT-D) (1), EMBARC (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care) (2), or CAN-BIND (Canadian Biomarker Integration Network in Depression)(3). In parallel, a wider adoption of NIBS techniques emerged, such as repetitive transcranial magnetic stimulation (rTMS) for the treatment of MDD and other conditions such as obsessive-compulsive disorder (OCD) or addiction, with currently more than 24 FDA device approvals (4), as well as transcranial electrical stimulation (tES). Many NIBS studies have been complemented by imaging work (5–7). Since many NIBS applications have built upon neuroscientific knowledge (e.g., frontal asymmetry) and given the focus on interventional psychiatry and brain circuit therapeutics (8,9), identifying NIBS biomarkers is of great importance, both to improve clinical outcomes, and to validate hypothesized working mechanisms. We, therefore, aim to systematically review the current state of biomarker-driven precision psychiatry for NIBS.

Several prior reviews and meta-analyses have investigated biomarkers for depression focused on EEG (10) or MRI (11) and a critical meta-analysis questioned the usefulness of EEG biomarkers for guiding antidepressant response (12). This latter meta-analysis raised valid concerns about biomarker studies criticizing a lack of, particularly out-of-sample, replications, and demonstrating strong evidence for publication bias, with overrepresentation of studies with large effects and underrepresentation of null findings. This highlights the need for well-powered studies and out-of-sample validations to identify clinically actionable biomarkers. This systematic review, thus, focused on 1) adequately powered imaging studies and 2) studies that attempted

to (out-of-sample) replicate earlier findings. Concretely, biomarkers were considered robust if they showed the ability to predict treatment response in an adequately powered study and/or were replicated in multiple studies. The aim of this systematic review was to systematically extract robust biomarkers of NIBS treatment response.

Inclusion criteria

One of the main criticisms of Widge and colleagues (12) was that EEG biomarker studies suffered from low sample sizes (median N=25). Therefore, to prevent inclusion of underpowered studies and determine the right minimum sample size for inclusion, we first consulted power calculations from pivotal biomarker studies (see supplement). Given these pivotal trials yielded inconsistent sample-size justifications, we conducted a power calculation in GPower 3.1. (14) to determine a minimum sample size to define robust studies. We used a categorical outcome measure reflecting the difference in biomarker presence between responders and non-responders expressed as a medium effect size (Cohen's d=0.5) with an alpha level of p<0.05 and power of 0.7, resulting in a sample size of N=88. Furthermore, studies with smaller sample size could be included on the condition that subsequent replication studies were reported in an independent sample. Studies investigating pre-treatment biomarkers of any NIBS modality and protocol were included. Studies identifying treatment-emergent biomarkers *(biomarkers that reflect changes during treatment)*, were not taken into account in this review since such biomarkers would require high accuracy to justify stopping a treatment course halfway. Ideally, several studies found the same direction of effect in independent samples.

The exact search terms can be found in the supplement. Figure 1 visualizes the inclusion/exclusion and final selection of studies for EEG and MRI.

[Figure 1 here]

Results

Results are summarized below as well as in Figure 1, with biomarkers grouped thematically. Information on the included MRI and EEG studies are summarized in Table 1 and 2, respectively. The systematic review only yielded rTMS studies since no studies on other NIBS/TES modalities met our inclusion criteria. rTMS is a technique that can be used to non-invasively modulate brain activity, based on the principles of electromagnetic induction (16). In the specific case of depression treatment, mostly the left DLPFC is stimulated (17). When a different stimulation location was used or the biomarker was protocol-specific, this is explicitly stated. Additionally, a detailed description of technical terms used in this section can be found in Figure 2.

MRI biomarkers

[Figure 2 here]

Anatomical MRI: Cortical Thickness

In a first study, Boes and colleagues reported thinner rostral anterior cingulate (rACC) cortex at baseline to be associated with better clinical improvement (18). However, subsequent work failed to replicate this finding (19) albeit here accelerated intermittent theta burst stimulation (iTBS) was used, whereas Boes and colleagues used 10Hz rTMS.

Functional MRI: DLPFC-sgACC functional connectivity

In an influential 2012 study, Fox and colleagues suggested that the DLPFC (as part of the Central Executive Network) should only be seen as an entry point to a network relevant to the pathophysiology of depression (20). They demonstrated that clinical benefit of rTMS for depression was related to intrinsic functional connectivity (FC) of the respective DLPFC target to the sgACC (as part of the Default Mode Network), as determined by resting-state functional MRI (rs-fMRI). This functional connectivity was indexed as 'anti-correlation' of the sgACC to prefrontal cortical areas, and suggestive of a way to individualize prefrontal rTMS sites for MDD treatment by selecting the most sgACC-anti-correlated prefrontal site.

Several studies have attempted to replicate this finding, with successful conceptual replications by Weigand and colleagues (21), Cash (22), Siddiqi (23) and Elbau and colleagues (24). However, studies in which a whole brain FC analysis was performed, using the sgACC as seed-region showed no relationship between functional anti-correlations between the seed and stimulation targets in the left DLPFC and response (25–30). These non-replication studies are all based on individual rs-fMRI data. Hopman and colleagues even suggested an inverse relationship, i.e., stronger connectivity between the sgACC and stimulation site was related to improved clinical response (27).

The studies by Fox et al. (20) and Weigand et al. (21) employed a normative functional connectome to derive FC. Cash et al. reasoned that using individual rs-fMRI data instead of a normative functional connectome may potentially improve TMS-personalization (22). Besides replication of previous results based on the normative connectome, this study showed that the relation between functional anti-correlation and clinical response was preserved when individual rs-fMRI data were used instead of group connectome data.

In 2021, Cash et al. introduced new insights into the relationship between FC and clinical responses (31). Instead of the direct FC between the stimulation site in the left DLPFC and the sgACC, the proximity between the clinically applied stimulation site and the rs-fMRI-personalized target in the left DLPFC was found to be related to clinical response. This relationship was not significant when personalized targets were replaced by a group average target derived from a normative functional connectome, arguing for the first time for the advantages of using individual rs-fMRI data. Siddiqi et al. (23) confirmed the importance of distance and even reported a response rate of 100% for patients whose stimulated target was within 25 mm of the personalized target.

Recently, Elbau et al. published the largest study (N=295), focusing on the potential of sgACC connectivity to infer TMS coil positions, as of now (24). Although an association between FC between the sgACC and left DLPFC target and clinical response was observed, this association

was much weaker (r=-.16) compared to previous studies (e.g. r=-.355 (20)), with low explained variance (3%). Only when subject-specific TMS-induced electric field simulations were performed and a weighted seedmap method was used to derive the time series of the sgACC, the weak but robust correlation was found. Of note, this relation was stronger in a subgroup of patients with strong global signal fluctuations due to burst breathing patterns (24). It was suggested that this weaker relationship could potentially be attributed to the relatively low-resolution of the rs-fMRI data (voxel size 5x5x5mm) (32). Indeed, better data quality could lead to better predictions and nowadays more sophisticated scanning sequences such as multi-echo and multi-band sequences, are available (33). Moreover, studies that showed stronger relations between anti-correlations and clinical responses based on high(er) resolution rs-fMRI data used strong smoothing parameters, effectively lowering the spatial resolution.

FC between the sgACC and the left DLPFC has been studied extensively in relation to clinical response to rTMS treatment in MDD. This information can be used to define personalized coil-positions and might in the future become a robust MRI-derived biomarker. However, optimal methodology to compute FC needs further investigation, and future prospective studies are warranted to demonstrate utility of this approach on the individual level.

Functional MRI: Lesion network mapping

In addition to using functional connections between specific brain regions as potential biomarkers, connectivity of stimulation sites with brain networks have also been related to clinical responses. A general depression network was identified by studying FC profiles from the normative connectome of 14 independent datasets including data on brain lesions, TMS, or deep brain stimulation (DBS), representing different sources of causal effects (8). Correlations between the individual connectivity maps of the TMS stimulation site and the depression network predicted the efficacy of the stimulation target. Cash et al. used a comparable approach to derive a network related to aberrant emotional processing in MDD patients, using coordinate network mapping of spatially heterogeneous coordinates (34). Of note, this emotional network resembles the depression network by Siddiqi et al. ($\boldsymbol{\rho}$=0.47, p=0.00)(8). Closer proximity between the

stimulation target and the emotional-network-derived personalized targets was associated with better clinical response (34).

These findings suggest that in the future, effective rTMS stimulation sites could be derived from correlations between individual connectivity profiles and the depression network.

Functional MRI: Machine learning-derived biotype approach

Using FC as input to machine learning (ML) approaches, Drysdale et al. (35) identified four clusters, called biotypes, which in a subsequent validation showed differential sensitivity to response to rTMS over the dorsomedial prefrontal cortex (dmPFC). Subtype 1, represented by reduced connectivity in a fronto-amygdalar network and reduced connectivity to anterior cingulate and orbitofrontal areas, showed a high partial response rate of 83% (25%, 61% and 30% for subtypes 2, 3 and 4, respectively). Of note, partial response was defined as a >25% reduction in Hamilton depression rating scale (HDRS), albeit results were similar when using the more traditional >50% cut-off for response, but predicted full-response was lower (e.g. ~63% for biotype 1).

Later work by Dinga (36) failed to replicate these findings in a more heterogeneous sample of 187 patients with depression and anxiety. Their analysis led to three instead of four clusters. Neither the canonical correlates nor the clusters were statistically significant. Potential methodological explanations for this non-replication are overfitting of the nonregularized canonical correlation analysis and arbitrary definitions of the subtypes (37). Also, variations in the clinical sample characteristics might explain the non-replication (38).

[Figure 3 here]

EEG biomarkers

EEG Frequency band power: Theta Power

EEG biomarker studies have traditionally focused on frequency band power (e.g. theta or alpha), however few sufficiently powered biomarkers have been found and replicated for NIBS.

An early study by Arns (39) in 90 MDD patients reported that high frontocentral theta power, low prefrontal delta and beta cordance and high P300 amplitude at baseline were associated with non-response to 10Hz rTMS over DLPFC. However, in a replication attempt the findings for theta and P300 could not be replicated by the same group (40).

Frontal-midline theta power and *change* in frontal theta power, measured after a rostral ACC-engaging cognitive task demonstrated predictive potential in a small pilot study (41). The findings were replicated in an independent sample and moreover it was shown that the obtained predictor was specific to 10Hz rTMS since it could not predict response to iTBS treatment (42). In both studies, response was evaluated after 10 treatment sessions - a low number to assess clinical improvement. The final sample size was small (N=24 in the pilot and N=35 per treatment arm in the replication), however, the concept of independent-sample replication strengthens the findings, and the differential prediction for iTBS vs 10Hz rTMS suggests potential for future treatment stratification.

[Figure 4 here]

EEG Machine-Learning and Source-Reconstruction

Wu and colleagues reported on ML applied to the alpha band, where response to sertraline – but not placebo – could be specifically predicted in the EMBARC dataset (43). When this alpha-signature of response to rTMS was prospectively tested, it predicted change on the anxiety subscale of the DASS (Depression, Anxiety and Stress Scale) after 1Hz TMS treatment. Notably, the predictive effect was specific to 1Hz treatment (and not 10Hz), and opposite that of sertraline, offering potential for stratification. However, since no effects for depressive symptoms were reported (neither BDI nor DASS-depression), this analysis cannot be considered a true out-of-sample validation. Moreover, when another group inferred the data points reported for the sertraline finding, and calculated the ROC curve, model performance was rather weak with an AUC= 0.67 (for a detailed critique about the methodology, see (44)).

A novel approach which conceptually resembles the previously mentioned rs-fMRI biotype analyses, applied independent component analysis to source-reconstructed EEG frequency band data. An EEG signature was identified that was associated with the polygenic risk scores for antidepressant response (45). Subsequent application of this signature to new samples yielded an association with response to both antidepressants and rTMS in men, but not women. As selecting EEG biomarkers using genetic data is a novel technique, this study should rather be viewed as a proof-of-concept that could aid in future biomarker development but requires further replication and comparison of the obtained networks with other known rs-fMRI or EEG networks.

Individual alpha peak frequency

One of the most heritable and reproducible aspects of the EEG is the individual alpha peak frequency (iAF) - the exact frequency of the alpha oscillations (46–48). Initial findings for iAF were mixed. Some studies reported an association between slow alpha and non-response to DLPFC rTMS (39,49) which could not be replicated by the same group (40) or by Widge (50). Adding iAF to a predictive model of non-linear EEG features of the alpha band, on the other hand, improved model prediction albeit in a rather small group of non-responders (N=20)(51).

More recent work shed light on these contradictory results by showing a predictive effect of iAF that was specific to 10Hz rTMS treatment outcome (with no such effect for 1Hz R-DLPFC rTMS) and could only be found using an average reference (indexing more focal activity than the linked-ears montage used in the studies mentioned above) (52). Furthermore, the association between iAF and symptom improvement turned out to be a quadratic instead of the previously assumed linear effect, demonstrating that the *distance* of iAF to 10Hz was negatively correlated with symptom improvement after 10Hz rTMS (6). These results were successfully replicated (52) in the same sample by Krepel et al. (40), where previous findings (using linked ears reference) could initially not be replicated. This emphasizes the importance of *exact* methodological replications

and a uniform way to preprocess and analyze EEG data. These differential predictive results for iAF in 10Hz and 1Hz rTMS were recently further validated in a treatment stratification approach using iAF-based Brainmarker-I including multiple blinded out-of-sample validations for 1Hz rTMS and ECT (53).

EEG Cordance

A study investigating prefrontal theta cordance found that baseline cordance could predict response to 1Hz rTMS with high accuracy (54) although this could not be replicated in another study where only 1-week change in theta cordance at central electrode sites predicted differences in response but not baseline or prefrontal cordance (55).

Two ML studies investigated pretreatment frontal cordance to predict outcome to 25Hz rTMS in the same dataset of 147 subjects, using artificial neural networks (56,57). High classification accuracies were obtained, albeit in the first study only a 6-fold cross validation was conducted but models were not tested in an external validation set which is considered necessary to prevent over-fitting (58). The second study in 2016 included a separate sample of 36 subjects for external validation, achieving high accuracy (AUC=.807-918). However, another ML study that used minimal-redundancy-maximal-relevance feature selection to test response prediction with frontal and prefrontal baseline cordance found no differences between responders and non-responders (59). Thus, no conclusions can be drawn about the predictive value of baseline cordance.

EEG Functional Connectivity

Zhang and colleagues used ML to identify differences in beta connectivity in frontal and posterior regions during eyes-open recordings which could distinguish two clinical subtypes that responded differentially to psychotherapy in posttraumatic stress disorder and SSRI treatment in MDD (60). However, no such differences between subtypes were found for rTMS, suggesting little relevance for rTMS prediction, but possible relevance for stratification between SSRI and rTMS treatment. Another ML model, built on 54 EEG features, such as baseline and week-1 alpha and

theta connectivity (and other features such as power, iAF and cordance), demonstrated high predictive accuracy of response (86.6%) (61), which could not be replicated in an independent sample (62). The discovery analysis was based on only 12 responders compared to 128 responders in the replication sample. One important caveat of the replication analysis was the strong differences in EEG processing that can lead to different results (52).
Findings regarding FC are, thus, inconclusive with different processing and modelling approaches hampering robust findings.

Discussion

The aim of this systematic review was to assess the progress regarding EEG and MRI-biomarkers for NIBS techniques. To improve upon previous criticisms, particularly the lack of replications as highlighted by Widge and colleagues(12), we focused on robustness in this review. To achieve this, we included only studies with a sample size of N≥88 or those that attempted to replicate biomarkers in independent samples, in order to identify robust biomarkers that can be used clinically to predict response to NIBS techniques.

Eighteen MRI and 18 EEG biomarker studies were included (visualized in figure 1). All studies focused on rTMS while no relevant imaging biomarker studies were found for TES.

MRI Biomarkers

The most robust rs-fMRI based metric predicting clinical response supported by a large sample (N=295) (24) as well as several independent replications, is the anti-correlation between the stimulation target (within the left DLPFC) and sgACC (20). This anti-correlation was shown to be related to response to various rTMS protocols, such as iTBS and 10Hz rTMS. However, replication in the largest sample yielded only weak effects (24), potentially suggesting reduced utility in clinical practice. Thus, prospective studies targeting the personalized location in the DLPFC with the highest anti-correlation with the sgACC should demonstrate if this connection has true biomarker potential.

A newer method based on network mapping also demonstrated biomarker potential of the connection between the stimulation site and a general depression network or an emotional network. Even though these findings are based on a large study using data from 151 TMS stimulation sites (four merged TMS datasets) (8) and was independently replicated (34), more and prospective research is warranted to demonstrate clinical value.

EEG Biomarkers

For EEG biomarkers, frontal-midline-elicited theta power after an rACC-activating task and iAF emerged as the most promising and robust EEG biomarkers. Frontal-midline theta power has been extensively described in the literature as a biomarker for treatment prediction and is thought to reflect rACC theta (for review see Pizzagalli (63)), supported by the finding that an rACC-engaging task can elicit this frequency (41,42). Interestingly, rACC activation was found to be predictive across imaging modalities, including EEG and fMRI (63). However, this was true for both sertraline *and* placebo response (13). Thus, despite successful replication, future studies should further investigate whether this finding is specific to 10Hz rTMS vs. iTBS or should rather be considered a non-specific predictor of response, including placebo.

The iAF finding emerged from two well-powered studies (N=143; N=153) by two independent groups. Interestingly, this result was specific to 10Hz rTMS (proximity of iAF to 10Hz was associated with better clinical response, suggesting 'synchronisation' effects of rTMS to the endogenous iAF rhythm). Recent work indicated promise of the iAF-based Brainmarker-I to stratify between 10Hz L-DLPFC and 1Hz R-DLPFC rTMS to enhance clinical outcomes (53), providing additional clinical merit of this biomarker.

Lessons learned: The devil is in the detail.

There are many methods to derive seed regions and compute prefrontal-sgACC FC. Even though earlier work used circles or weighted cone models to derive seed region time-series, currently more advanced methods such as individual TMS-induced electric field simulations and weighted seedmap methods are used. These methodological details have shown to be highly influential

since Elbau et al. (24) only found a relation between the FC between stimulation site and sgACC and clinical response when the stimulated area was derived from the simulated electric field distribution and the sgACC time-series were derived using a seedmap approach. Four of the papers included in this review demonstrate clinical value of using individual rs-fMRI data compared to group connectome data (23,24,34,64). Future research needs to compare biomarkers derived from these different connectomes and answer the question whether baseline individualized rs-fMRI data collection should be added to treatment protocols.

In the case of the iAF EEG biomarker, initial findings were mixed, even though several well-powered studies were used to examine the effect (e.g. N=180 (50) or N=90-106 (39,40), and replication analyses were conducted. Later work actually led to consistent and robust findings (6,52), showing that the crucial factors were: 1) Use of the correct EEG montage: initial studies used the less focal linked-ears reference, while Roelofs (52) demonstrated that the main result critically depended on the average reference montage; 2) protocol-specific effects for 10Hz TMS and no such effect for 1Hz TMS, meaning effects could average out when combined, and 3) a quadratic association between TMS response and iAF as opposed to the presumed linear association (i.e. lower iAF predicts worse TMS-response).

The actual predictive value of clinical response of these MRI- and EEG-derived metrics depends on the preprocessing pipelines used. Future research is necessary to investigate if the content of these metrics is related to core brain mechanisms or reflect other sources of signal fluctuations such as respiration or cardiac patterns.

Artificial intelligence (AI) and machine learning (ML)

The present review reveals a limited biomarker potential for AI and ML-techniques in both EEG and rs-fMRI studies.

Although large – and often multiple - samples were used and results seemed promising at first glance, some studies lacked external validation samples (56) which are needed to prevent overfitting (58); some out-of-sample validation results were only significant for different

measures than the discovery measure (e.g. anxiety instead of depression) (43); and some could not be replicated, possibly due to overfitting (62). Finally, it is important to use consistent definitions for response and remission (e.g. not 'partial response'), in order to keep outcomes comparable.

Future directions

It remains to be investigated whether the biomarkers described in this review generalize to multiple rTMS protocols. If not, this might at least partly explain some of the unsuccessful replication attempts. Moreover, especially for biomarkers with weaker effects, the cost/benefit ratio needs to be assessed.

Although predicting NIBS outcome in other disorders would be highly relevant, the present manuscript only discusses robust biomarkers for MDD. Future research is needed to determine if these are also predictive of treatment response in other disorders.

Finally, prospective studies, similar to van der Vinne et al (65), will be necessary to test treatment individualization in daily clinical practice.

Conclusion

This systematic review has identified four robust neuroimaging biomarkers that have reached a sufficient level for testing in prospective trials to evaluate their feasibility and clinical actionability. Some of those biomarkers show promise for treatment stratification (e.g. stratification between 10Hz vs. 1Hz rTMS protocols using iAF (53)) which might be a more realistic and feasible approach for clinical practice opposed to precision psychiatry (66).

Overall, a limited number of studies met our inclusion criteria, highlighting the need for improvements in the quality of imaging biomarker research for rTMS. Nevertheless, the identification of four robust biomarkers over the past decade presents a promising outlook and justifies large trials, similar to iSPOT-D and EMBARC for antidepressant medication, but then aimed at rTMS and NIBS.

Acknowledgements

We would like to thank Mark Koppenberg for the design of figures and tables.

Disclosures

MA is unpaid chairman of the Brainclinics Foundation, holds equity/stock in neurocare, served as consultant to neurocare and Roche; Brainclinics Foundation received research and consultancy support from neurocare and equipment support from Deymed and MagVenture. DK, HV and CB report no biomedical financial interests or potential conflicts of interest.

References

1. Williams LM, Rush AJohn, Koslow SH, Wisniewski SR, Cooper NJ, Nemeroff CB, *et al.* (2011): International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: Rationale and protocol. *Trials* 12: 1–17.

2. Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, *et al.* (2016): Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *J Psychiatr Res* 78: 11–23.

3. Lam RW, Milev R, Rotzinger S, Andreazza AC, Blier P, Brenner C, *et al.* (2016): Discovering biomarkers for antidepressant response: protocol from the Canadian biomarker integration network in depression (CAN-BIND) and clinical characteristics of the first patient cohort. *Bmc Psychiatry* 16: 105.

4. Cohen SL, Bikson M, Badran BW, George MS (2022): A visual and narrative timeline of US FDA milestones for Transcranial Magnetic Stimulation (TMS) devices. *Brain Stimul* 15: 73–75.

5. Blumberger DM, Vila-Rodriguez F, Thorpe KE, Feffer K, Noda Y, Giacobbe P, *et al.* (2018): Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. *Lancet* 391: 1683–1692.

6. Corlier J, Carpenter LL, Wilson AC, Tirrell E, Gobin AP, Kavanaugh B, Leuchter AF (2019): The relationship between individual alpha peak frequency and clinical outcome with repetitive Transcranial Magnetic Stimulation (rTMS) treatment of Major Depressive Disorder (MDD). *Brain Stimul*. https://doi.org/10.1016/j.brs.2019.07.018

7. Dijk H van, Wingen G van, Denys D, Olbrich S, Ruth R van, Arns M (2022): The two decades brainclinics research archive for insights in neurophysiology (TDBRAIN) database. *Sci Data* 9: 333.

8. Siddiqi SH, Schaper FLWVJ, Horn A, Hsu J, Padmanabhan JL, Brodtmann A, *et al.* (2021): Brain stimulation and brain lesions converge on common causal circuits in neuropsychiatric disease. *Nat Hum Behav* 1–10.

9. Spellman T, Liston C (2020): Toward Circuit Mechanisms of Pathophysiology in Depression. *Am J Psychiat* 177: 381–390.

10. Olbrich S, Dinteren R van, Arns M (2015): Personalized Medicine: Review and Perspectives of Promising Baseline EEG Biomarkers in Major Depressive Disorder and Attention Deficit Hyperactivity Disorder. *Neuropsychobiology* 72: 229–40.

11. Wingen G van, Bockting C, Zantvoord J, Wezenberg B, Cohen S (2020): Magnetic resonance imaging for individual prediction of treatment response in major depressive disorder: a systematic review and meta-analysis. *Medrxiv* 2020.06.27.20141465.

12. Widge AS, Bilge MT, Montana R, Chang W, Rodriguez CI, Deckersbach T, *et al.* (2019): Electroencephalographic biomarkers for treatment response prediction in major depressive illness: a meta-analysis. *Am J Psychiatry* 176: 44–56.

13. Pizzagalli DA, Webb CA, Dillon DG, Tenke CE, Kayser J, Goer F, *et al.* (2018): Pretreatment Rostral Anterior Cingulate Cortex Theta Activity in Relation to Symptom Improvement in Depression: A Randomized Clinical Trial. *Jama Psychiat* 75: 547.

14. Faul F, Erdfelder E, Lang A-G, Buchner A (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39: 175–191.

15. Bollen, K., Cacioppo, T. J, Kaplan, M. R, *et al.* (2015): *Report of the Subcommittee on the Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, & Economic Sciences.* Retrieved from https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

16. Baeken C, Brem A-K, Arns M, Brunoni AR, Filipčić I, Ganho-Ávila A, *et al.* (2019): Repetitive transcranial magnetic stimulation treatment for depressive disorders. *Curr Opin Psychiatr* 32: 409–415.

17. O'Reardon JP, Solvason HB, Janicak PG, Sampson S, Isenberg KE, Nahas Z, *et al.* (2007): Efficacy and Safety of Transcranial Magnetic Stimulation in the Acute Treatment of Major Depression: A Multisite Randomized Controlled Trial. *Biol Psychiatry* 62: 1208–1216.

18. Boes AD, Uitermarkt BD, Albazron FM, Lan MJ, Liston C, Pascual-Leone A, *et al.* (2018): Rostral anterior cingulate cortex is a structural correlate of repetitive TMS treatment response in depression. *Brain Stimul* 11: 575–581.

19. Baeken C, Beek V van, Vanderhasselt M, Duprat R, Klooster D (2021): Cortical Thickness in the Right Anterior Cingulate Cortex Relates to Clinical Response to Left Prefrontal Accelerated Intermittent Theta Burst Stimulation: An Exploratory Study. *Neuromodulation Technology Neural Interface*. https://doi.org/10.1111/ner.13380

20. Fox MD, Buckner RL, White MP, Greicius MD, Pascual-Leone A (2012): Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biol Psychiatry* 72: 595–603.

21. Weigand A, Horn A, Caballero R, Cooke D, Stern AP, Taylor SF, *et al.* (2018): Prospective Validation That Subgenual Connectivity Predicts Antidepressant Efficacy of Transcranial Magnetic Stimulation Sites. *Biol Psychiat* 84: 28–37.

22. Cash RFH, Zalesky A, Thomson RH, Tian Y, Cocchi L, Fitzgerald PB (2019): Subgenual Functional Connectivity Predicts Antidepressant Treatment Response to Transcranial Magnetic Stimulation: Independent Validation and Evaluation of Personalization. *Biol Psychiat* 86: e5–e7.

23. Siddiqi SH, Weigand A, Pascual-Leone A, Fox MD (2021): Identification of Personalized Transcranial Magnetic Stimulation Targets Based on Subgenual Cingulate Connectivity: An Independent Replication. *Biol Psychiat* 90: e55–e56.

24. Elbau IG, Lynch CJ, Downar J, Vila-Rodriguez F, Power JD, Solomonov N, *et al.* (2023): Functional Connectivity Mapping for rTMS Target Selection in Depression. *Am J Psychiat* 180: 230–240.

25. Baeken C, Marinazzo D, Wu G-R, Schuerbeek PV, Mey JD, Marchetti I, *et al.* (2014): Accelerated HF-rTMS in treatment-resistant unipolar depression: Insights from subgenual anterior cingulate functional connectivity. *World J Biological Psychiatry* 15: 286–297.

26. Baeken C, Duprat R, Wu G-R, Raedt RD, Heeringen K van (2017): Subgenual Anterior Cingulate–Medial Orbitofrontal Functional Connectivity in Medication-Resistant Major Depression: A Neurobiological Marker for Accelerated Intermittent Theta Burst Stimulation Treatment? *Biological Psychiatry Cognitive Neurosci Neuroimaging* 2: 556–565.

27. Hopman HJ, Chan SMS, Chu WCW, Lu H, Tse C-Y, Chau SWH, *et al.* (2021): Personalized prediction of transcranial magnetic stimulation clinical response in patients with treatment-refractory depression using neuroimaging biomarkers and machine learning. *J Affect Disorders* 290: 261–271.

28. Persson J, Struckmann W, Gingnell M, Fällmar D, Bodén R (2020): Intermittent theta burst stimulation over the dorsomedial prefrontal cortex modulates resting-state connectivity in depressive patients: A sham-controlled study. *Behav Brain Res* 394: 112834.

29. Ge R, Downar J, Blumberger DM, Daskalakis ZJ, Vila-Rodriguez F (2020): Functional connectivity of the anterior cingulate cortex predicts treatment outcome for rTMS in treatment-resistant depression at 3-month follow-up. *Brain Stimulation* 13. https://doi.org/10.1016/j.brs.2019.10.012

30. Salomons TV, Dunlop K, Kennedy SH, Flint A, Geraci J, Giacobbe P, Downar J (2013): Resting-state cortico-thalamic-striatal connectivity predicts response to dorsomedial prefrontal rTMS in major depressive disorder. *Neuropsychopharmacol Official Publ Am Coll Neuropsychopharmacol* 39: 488–98.

31. Cash RFH, Cocchi L, Lv J, Fitzgerald PB, Zalesky A (2020): Functional Magnetic Resonance Imaging–Guided Personalization of Transcranial Magnetic Stimulation Treatment for Depression. *Jama Psychiat*. https://doi.org/10.1001/jamapsychiatry.2020.3794

32. Siddiqi SH, Philip NS (2023): Hitting the Target of Image-Guided Psychiatry? *Am J Psychiat* 180: 185–187.

33. Lynch CJ, Power JD, Scult MA, Dubin M, Gunning FM, Liston C (2020): Rapid Precision Functional Mapping of Individuals Using Multi-Echo fMRI. *Cell Reports* 33: 108540.

34. Cash RFH, Müller VI, Fitzgerald PB, Eickhoff SB, Zalesky A (2023): Altered brain activity in unipolar depression unveiled using connectomics. *Nat Mental Health* 174–185.

35. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, *et al.* (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23: 28–38.

36. Dinga R, Schmaal L, Penninx B, Tol MJ van, Veltman DJ, Velzen L van, *et al.* (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017). *Neuroimage Clin* 22: 101796.

37. Dinga R, Schmaal L, Marquand AF (2020): A Closer Look at Depression Biotypes: Correspondence Relating to Grosenick et al. (2019). *Biological Psychiatry Cognitive Neurosci Neuroimaging* 5: 554–555.

38. Grosenick L, Liston C (2020): Reply to: A Closer Look at Depression Biotypes: Correspondence Relating to Grosenick et al. (2019). *Biological Psychiatry Cognitive Neurosci Neuroimaging* 5: 556.

39. Arns M, Drinkenburg WH, Fitzgerald PB, Kenemans JL (2012): Neurophysiological predictors of non-response to rTMS in depression. *Brain Stimul* 5: 569–76.

40. Krepel N, Sack AT, Kenemans JL, Fitzgerald PB, Drinkenburg WH, Arns M (2018): Non-replication of neurophysiological predictors of non-response to rTMS in depression and neurophysiological data-sharing proposal. *Brain Stimul* 11: 639–641.

41. Li C-T, Hsieh J-C, Huang H-H, Chen M-H, Juan C-H, Tu P-C, *et al.* (2016): Cognition-Modulated Frontal Activity in Prediction and Augmentation of Antidepressant Efficacy: A Randomized Controlled Pilot Study. *Cereb Cortex* 26: 202–210.

42. Li C-T, Cheng C-M, Juan C-H, Tsai Y-C, Chen M-H, Bai Y-M, *et al.* (2021): Task-Modulated Brain Activity Predicts Antidepressant Responses of Prefrontal Repetitive Transcranial Magnetic Stimulation: A Randomized Sham-Control Study. *Chronic Stress* 5: 24705470211006856.

43. Wu W, Zhang Y, Jiang J, Lucas MV, Fonzo GA, Rolle CE, *et al.* (2020): An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol* 38: 439–447.

44. Nilsonne G, Harrell FE (2020): EEG-based model and antidepressant response. *Nat Biotechnol* 1–1.

45. Meijs H, Prentice A, Lin BD, Wilde BD, Hecke JV, Niemegeers P, *et al.* (2022): A polygenic-informed approach to a predictive EEG signature empowers antidepressant treatment prediction: A proof-of-concept study. *Eur Neuropsychopharm* 62: 49–60.

46. Posthuma D, Neale MC, Boomsma DI, Geus EJC de (2001): Are Smarter Brains Running Faster? Heritability of Alpha Peak Frequency, IQ, and Their Interrelation. *Behav Genet* 31: 567–579.

47. Smit DJA, Posthuma D, Boomsma DI, Geus EJC de (2005): Heritability of background EEG across the power spectrum. *Psychophysiology* 42: 691–697.

48. Beijsterveldt CEM van, Baal GCM van (2002): Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biol Psychol* 61: 111–138.

49. Martijn A, Desirée S, B. F Paul (2010): Potential differential effects of 9 Hz rTMS and 10 Hz rTMS in the treatment of depression. *Brain Stimul* 3: 124–126.

50. Widge AS, Avery DH, Zarkowski P (2013): Baseline and Treatment-Emergent EEG Biomarkers of Antidepressant Medication Response Do Not Predict Response to Repetitive Transcranial Magnetic Stimulation. *Brain Stimul* 6: 929–931.

51. Arns M, Cerquera A, Gutiérrez RM, Hasselman F, Freund JA (2014): Non-linear EEG analyses predict non-response to rTMS treatment in major depressive disorder. *Clin Neurophysiol* 125: 1392–1399.

52. Roelofs C, Krepel N, Corlier J, Carpenter LL, Fitzgerald PB, Daskalakis ZJ, *et al.* (2020): Individual alpha frequency proximity associated with repetitive transcranial magnetic stimulation outcome: An independent replication study from the ICON-DB consortium. *Clin Neurophysiol* S1388-2457: 30532–0.

53. Voetterl H (2023): Utility of alpha peak frequency-based Brainmarker-I in antidepressant treatment stratification: Blinded out-of-sample validations for pharmacotherapy and brain stimulation. Retrieved from https://doi.org/10.1016/j.brs.2023.01.687

54. Bares M, Brunovsky M, Novak T, Kopecek M, Stopkova P, Sos P, Höschl C (2015): QEEG Theta Cordance in the Prediction of Treatment Outcome to Prefrontal Repetitive Transcranial Magnetic Stimulation or Venlafaxine ER in Patients With Major Depressive Disorder. *Clin Eeg Neurosci* 46: 73–80.

55. Hunter AM, Nghiem TX, Cook IA, Krantz DE, Minzenberg MJ, Leuchter AF (2017): Change in Quantitative EEG Theta Cordance as a Potential Predictor of Repetitive Transcranial Magnetic Stimulation Clinical Outcome in Major Depressive Disorder. *Clin Eeg Neurosci* 49: 306–315.

56. Erguzel TT, Ozekes S, Gultekin S, Tarhan N, Sayar GH, Bayram A (2014): Neural Network Based Response Prediction of rTMS in Major Depressive Disorder Using QEEG Cordance. *Psychiatry Investig* 12: 61–65.

57. Erguzel TT, Tarhan N (2016): Machine Learning Approaches to Predict Repetitive Transcranial Magnetic Stimulation Treatm ent Response in Major Depressive Disorder. *SAI Intelligent Systems Conference* 391–401.

58. Ho SY, Phua K, Wong L, Goh WWB (2020): Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. *Patterns* 1: 100129.

59. Hasanzadeh F, Mohebbi M, Rostami R (2019): Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal. *J Affect Disord* 256: 132–142.

60. Zhang Y, Wu W, Toll RT, Naparstek S, Maron-Katz A, Watts M, *et al.* (2020): Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nat Biomed Eng* 5: 309–323.

61. Bailey N, Hoy K, Rogasch N, Thomson R, McQueen S, Elliot D, *et al.* (2019): Differentiating responders and non-responders to rTMS treatment for depression after one week using resting EEG connectivity measures. *J Affect Disorders* 242: 68–79.

62. Bailey NW, Krepel N, Dijk H van, Leuchter AF, Vila-Rodriguez F, Blumberger DM, *et al.* (2021): Resting EEG theta connectivity and alpha power to predict repetitive transcranial magnetic stimulation response in depression: A non-replication from the ICON-DB consortium. *Clin Neurophysiol* 132: 650–659.

63. Pizzagalli DA (2011): Frontocingulate Dysfunction in Depression: Toward Biomarkers of Treatment Response. *Neuropsychopharmacol* 36: 183–206.

64. Cash RFH, Weigand A, Zalesky A, Siddiqi SH, Downar J, Fitzgerald PB, Fox MD (2021): Using Brain Imaging to Improve Spatial Targeting of Transcranial Magnetic Stimulation for Depression. *Biol Psychiat* 90: 689–700.

65. Vinne N van der, Vollebregt MA, Rush AJ, Eebes M, Putten MJAM van, Arns M (2021): EEG biomarker informed prescription of antidepressants in MDD: a feasibility trial. *Eur Neuropsychopharm* 44: 14–22.

66. Arns M, Dijk H van, Luykx JJ, Wingen G van, Olbrich S (2022): Stratified psychiatry: Tomorrow's precision psychiatry? *Eur Neuropsychopharm* 55: 14–19.

Figure legends:

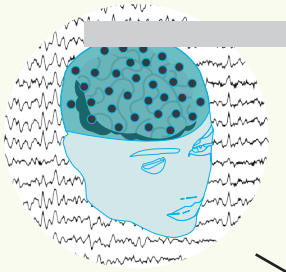
Figure 1: Flow-diagram of total studies identified, excluded and included in the systematic review for EEG-biomarkers (left) and MRI-biomarkers (right), as well as all identified and the most robust biomarkers that emerged from this systematic review (1, 2 for EEG and 3, 4 for MRI).
Records were excluded on basis of the abstract if they turned out to be non-human studies, no original research, pertain to another pathology than MDD, or another biomarker than EEG/MRI, or another treatment than NIBS.
Prespecified exclusion criteria were: 1. Treatment-emergent biomarker, and 2. Sample size <88 and no replication (iAF= individual alpha frequency; rACC=rostral anterior cingulate cortex; DLPFC=dorsolateral prefrontal cortex, sgACC=subgenual anterior cingulate cortex).

Figure 2: Overview of study details on the included MRI studies based on sample size (N≥88; highlighted in green) or based on replication-work (highlighted in blue). Strength of finding reports the area under the receiver-operator characteristic curve (AUC), effect size, correlation coefficient or another measure of effect size, depending on what was reported in article. Total N refers to the full sample size used to compute the biomarker while Group N is the sample size of the group in which the biomarker was tested for rTMS.
(a)iTBS= (accelerated) intermittent theta burst stimulation, dmPFC = dorsomedial prefrontal cortex, HDRS= Hamilton Depression Rating Scale, rTMS = repetitive transcranial magnetic stimulation, QIDS = Quick Inventory of Depressive Symptomatology, DLPFC = dorsolateral prefrontal cortex, (sg)ACC= (subgenual) anterior cingulate cortex, BDI = Beck Depression Inventory, MADRS = Montgomery-Åsberg Depression Rating Scale, DBS = deep brain stimulation, ANOVA = analysis of variance, (rs) FC = (resting-state) functional connectivity.

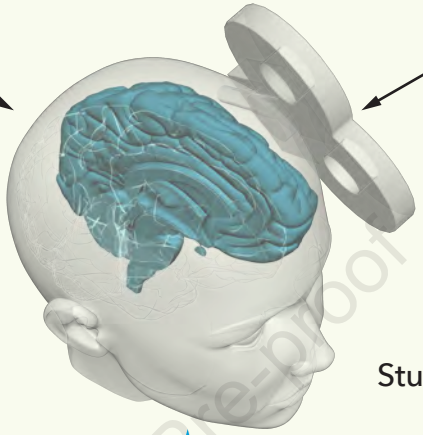
Figure 3: Glossary of terms used throughout the article.

Figure 4: Overview of study details of included EEG studies based on sample size (N≥88; highlighted in green) or based on replication-work (highlighted in blue). Strength of finding reports the area under the receiver-operator characteristic curve (AUC), effect size, correlation coefficient or another measure of effect size, depending on what was reported in article. Total N refers to the full sample size used to compute the biomarker while Group N is the sample size of the group in which the biomarker was tested for rTMS.

BDI = Beck Depression Inventory, iAF= individual alpha peak frequency, IDS(-SR) = Inventory of Depressive Symptomatology (self-rated), rTMS = repetitive transcranial magnetic stimulation, HDRS= Hamilton Depression Rating Scale, ML= machine learning, PRS= polygenic risk score, fICA= functional independent component analysis, (DL)PFC = (dorsolateral) prefrontal cortex, DASS = Depression Anxiety and Stress Scale, EO= eyes open, EC = eyes closed, MADRS = Montgomery-Åsberg Depression Rating Scale, CGI-I= Clinical Global Impressions Scale (Improvement), iTBS= intermittent theta burst stimulation, RECT= rostral anterior cingulate cortex engaging cognitive task



EEG		MRI	
N			N
135	Records identified	Records identified	85
94	*Excluded on basis of abstract*	*Excluded on basis of abstract*	*37*
41	Studies assessed in detail	Studies assessed in detail	48
23	*Excluded based on prespecified exclusion criteria*	*Excluded based on prespecified exclusion criteria*	*30*
18	Studies included	Studies included	18



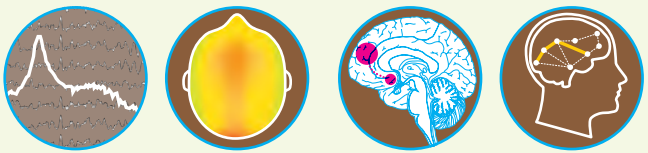
Identified biomarkers

EEG cordance

EEG functional connectivity

EEG machine learning source reconstruction

1 2 3 4



iAF

Frontal-midline rACC theta

DLPFC-sgACC functional connectivity

Network mapping

Robust biomarkers

MRI cortical thickness

MRI machine learning derived biotype approaches

Biomarker Category	Study	Total N (Group N)	rTMS protocol	Outcome measure	Strength of finding	Positive Finding	Alternative Finding
Cortical thickness	Boes, 2018	48 (48)	10 Hz rTMS	BDI/HDRS-24	r N.S., p<0.001	Thinner cortex in the rostral part [illegible] related with better clinical response	
	Baeken, 2021	50 (21)	Accelerated iTBS	HDRS-17	r = 0.51, p = 0.02		Thicker cortex in the right caudal part of the ACC at baseline was associated with better clinical response 3 days after stimulation
DLPFC-sgACC functional connectivity	Fox, 2012	149 (27)	10 Hz rTMS	MADRS	r = –0.355, p<.05	DLPFC sites with higher clinical efficacy showed higher anti-correlations with the sgACC	
	Baeken, 2014	20 (12)	Accelerated 10 Hz rTMS	HDRS-17	ANOVA F-value = 3.62		Responders showed significantly stronger rs FC anti-correlation between the sgACC and parts of the left superior medial prefrontal cortex
	Salomons, 2014	25 (25)	10 Hz rTMS (dmPFC)	HDRS-17	Peak z-score 3.6		Higher baseline connectivity between sgACC and dlPFC was associated with better clinical response.
	Weigand, 2017	25 (25) 16 (12)	10 Hz rTMS or 20 Hz rTMS 10 Hz rTMS	BDI MADRS	r = –0.55, p < 0.005 r = –0.52, p < 0.05	Better clinical response was related to higher anti-correlations between the stimulation site in the DLPFC and the sgACC	
	Baeken, 2017	50 (44)	Accelerated iTBS	HDRS-17	p<0.01		FC between sgACC and medial orbito-frontal cortex at baseline could distinguish aiTBS responders from non-responders
	Cash, 2019	47 (24)	10 Hz rTMS	MADRS	r = –0.61, p=0.001	Better clinical response was related to higher anti-correlations between the stimulation site in the DLPFC and the sgACC	
	Cash, 2020	26 (26)	10 Hz rTMS	MADRS	r = –0.54, p = 0.002	Better clinical response was related to higher anti-correlations between the stimulation site in the DLPFC and the sgACC	
	Ge, 2020	50 (50) 50 (32)*	10 Hz rTMS or iTBS	HDRS-17	AUC = 0.87, p<.001, r = –0.62, p<.001 AUC = 0.79, p<.001, r = –0.49, p=0.001		Clinical response (post-treatment and 12 weeks after rTMS) was related to lower functional connectivity between the sgACC and the right DLPFC
	Persson, 2020	30 (20)	iTBS	MADRS	p = 0.021, T = 6.75		Baseline functional connectivity between the sgACC and the precuneus is negatively correlated with clinical response
	Hopman, 2021	70 (61 long-term, 63 short-term)	10 Hz rTMS	MADRS	Effect size: .26 – .30 depending on area cluster (for long-term responders vs non-responders)		Stronger DLPFC-sgACC connectivity was associated with symptom improvement. Long-term responders showed higher connectivity between sgACC and frontal pole, superior parietal lobule, and occipital cortex and between the left DLPFC and the central opercular cortex
	Siddiqi, 2021	25 (25)	10 Hz rTMS	BDI	r = –0.6, p < 0.005	Better clinical response was related to higher anti-correlations between the stimulation site in the DLPFC and the sgACC	
	Elbau, 2023	414 (295)	10 Hz rTMS or iTBS	QIDS-SR	r = –0.16, p=.006	Clinical response was related to higher anti-correlations between the stimulation site in the DLPFC and the sgACC	
Network Mapping	Siddiqi, 2021	713 (151)	10 Hz or 20 Hz rTMS	BDI/MADRS/ HDRS-24	Weighted mean r = 0.22, p < 0.001	Circuits derived from lesions, rTMS, and DBS stimulation sites are similar and connectivity to this circuit predicts efficacy of rTMS treatment	
	aCash, 2023	26 (26)	10 Hz rTMS	MADRS	r = –0.41, p = 0.018	Closer proximity between actual and emotional network-specific TMS targets is associated with better clinical outcome	
Machine Learning	Drysdale, 2017	1188 (154)	10 Hz or iTBS (dmPFC)	HDRS-17	χ^2 = 25.7, p<.001	Four distinct biotypes, characterised by dysfunctional connectivity in limbic and frontostriatal networks predicted clinical response to dmPFC rTMS	
	Dinga et al. 2019	187 (187)	Not applicable	IDS	ns		Biotypes could not be replicated

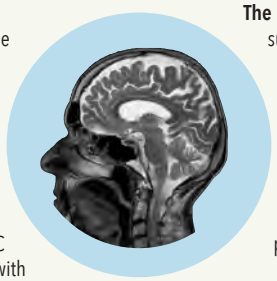
* *12-week Follow up*
Group N denotes the treatment group tested for effect
ns = not significant; N.S= not specified; IDS = inventory of depressive symptomatology

Sample size ≥88
Replication studies, sample size potentially <88
Robust biomarker found

MRI

connectivity map derived from rs-fMRI scans from many individuals, also called functional group connectome or human connectome. This connectome represents the average wiring diagram of the brain's functional connections. The advantage of a normative functional connectome is that the signal-to-noise ratio is higher compared to individual rs-fMRI data. However, inter-individual differences in functional connectivity are discarded.

TMS-induced electric field simulations can provide insight in the distribution of the TMS effects within the brain. When a TMS pulse is applied to the brain, a secondary electric field is induced in the superficial layers of the cortex. The exact distribution of this TMS-induced electric field depends on the shape of the TMS coil used as well as on the individual's gyral folding pattern.



The weighted seedmap method, introduced by Cash et al (1), is an alternative method to compute the time-series in the sgACC combining knowledge from the normative functional connectome with the individual rs-fMRI data. According to the weighted seedmap approach the time-series of the sgACC is computed as the weighted spatial average of the time-series in the gray matter voxels of the individual rs-fMRI data, excluding the DLPFC region. The weights are derived from the connectivity strength between the sgACC and the gray matter voxels in the normative functional connectome.

Global signal is the mean of the voxel time-series within the brain. Particularly in the work of Elbau et al. (2), the global signal is relevant since it was shown to reflect burst breathing patterns. Especially the subset of patients showing global signal patterns related to burst breathing showed strong negative correlations between sgACC-stimsite FC and clinical response.

er focal brain regions but is also sensitive to networks connected to those regions. At first, network mapping used lesions to seek convergence for symptoms caused by lesions in different non-overlapping brain regions (3). Network mapping has since been expanded to contain other (causal) sources of information such as TMS stimulation sites (TMS network mapping) (4) or coordinates related to abnormal brain functioning (coordinate network mapping) (5).

The emotional network, identified by Cash et al. (5), involves the subgenual cingulate cortex, pregenual anterior cingulate cortex, left DLPFC, cingulum, and superior frontal gyrus including the pre-supplementary motor area.

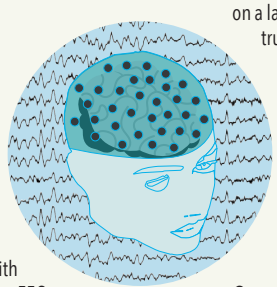
The depression network, derived by Siddiqi et al. (4), contains positive peaks in the DLPFC, frontal eye fields, inferior frontal gyrus, intraparietal sulcus and extrastriate visual cortex and negative peaks in the subgenual cingulate cortex and ventromedial prefrontal cortex. .

Canonical correlation analysis (CCA) is a well-established method used to identify the association between two sets of variables. Drysdale et al. (6) used CCA to select a low-dimensional representation of FC features that were related to weighted combinations of clinical symptoms. **Regularized CCA** is based on a subset of features. This prevents overfitting of CCA as might be the case in **nonregularized CCA.**

EEG

Frequency band power is most commonly calculated for the 5 standard frequency bands delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (30-100 Hz) although the frequency ranges are not standardized and often differ between studies. The power spectrum within a frequency band is usually calculated by Fast-Fourier Transform (FFT), an algorithm that transforms a signal from a time or space domain to a frequency domain.

The P300 is an event-related potential (ERP), which can be observed in the EEG in response to an infrequent tone in a row of frequent tones. It denotes a positive deflection approximately 300ms following the stimulus and is assumed to be involved in attention and memory processes.



Independent component analysis (ICA) is a computational method to filter a multivariate signal into its distinct subcomponents. ICA was here applied to data which had been source reconstructed with **LORETA (Low Resolution Brain Electromagnetic Tomography)**, an EEG method for 3D imaging brain activity to estimate where signals come from in the brain.

Polygenic risk scores (PRS) estimate a person's genetic predisposition to develop certain traits or disorders, based on their genetic profile and genome-wide association study data.

The individual alpha peak frequency (iAF) is the frequency at which an individual's alpha oscillations are most pronounced. It is calculated by determining the power spectrum within the alpha frequency band (see above) and identifying the highest (modal) peak in that spectrum.

Brainmarker-I is an iAF-based biomarker which has been age- and sex-normalized on a large dataset (>4000 individuals) by employing the biological ground truth that the iAF matures (speeds up) during childhood and adolescence (7).

Cordance is an EEG measure, originally developed by Leuchter and colleagues (8) that combines both absolute and relative power within a specific frequency band, with negative values reflecting increased slow-wave and decreased fast activity. This pattern was termed discordance and is assumed to reflect low perfusion and metabolism.

Cross validation is a statistical method used in machine learning to evaluate model performance. Ideally, an **external validation dataset** is used to test model predictions. Often, cross validation is done on a segment basis, meaning all data segments from all participants are merged and some segments are kept for later validation. This can lead to high prediction accuracy, so-called overfitting, since the model is predicting the participant instead of the signal.

1. Cash RFH, Cocchi L, Lv J, Wu Y, Fitzgerald PB, Zalesky A (2021): Personalized connectivity-guided DLPFC-TMS for depression: Advancing computational feasibility, precision and reproducibility. *Hum Brain Mapp* 42: 4155–4172.
2. Elbau IG, Lynch CJ, Downar J, Vila-Rodriguez F, Power JD, Solomonov N, *et al.* (2023): Functional Connectivity Mapping for rTMS Target Selection in Depression. *Am J Psychiat* 180: 230–240.
3. Boes AD, Prasad S, Liu H, Liu Q, Pascual-Leone A, Caviness VS, Fox MD (2015): Network localization of neurological symptoms from focal brain lesions. *Brain* 138: 3061–3075.
4. Siddiqi SH, Schaper FLWVJ, Horn A, Hsu J, Padmanabhan JL, Brodtmann A, *et al.* (2021): Brain stimulation and brain lesions converge on common causal circuits in neuropsychiatric disease. *Nat Hum Behav* 1–10.
5. Cash RFH, Müller VI, Fitzgerald PB, Eickhoff SB, Zalesky A (2023): Altered brain activity in unipolar depression unveiled using connectomics. *Nat Mental Health* 174–185.
6. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, *et al.* (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23: 28–38.
7. Voetterl H, Wingen G van, Michelini G, Griffiths KR, Gordon E, DeBeus R, *et al.* (2022): Brainmarker-I differentially predicts remission to various attention-deficit/hyperactivity disorder treatments: a blinded discovery, transfer and validation study. *Biological Psychiatry Cognitive Neurosci Neuroimaging.* https://doi.org/10.1016/j.bpsc.2022.02.007
8. Leuchter AF, Cook IA, Lufkin RB, Dunkin J, Newton TF, Cummings JL, *et al.* (1994): Cordance: A New Method for Assessment of Cerebral Perfusion and Metabolism Using Quantitative Electroencephalography. *Neuroimage* 1: 208–219.

Biomarker Category	Study	Total N (Group N)	rTMS protocol	Outcome measure	AUC (/effect size)	Positive Finding	Alternative Finding
Theta power & iAF	Arns, 2012 (39)	90 (90)	10 Hz or 1 Hz	BDI	.814	Non-response characterised by increased fronto-central theta, slower iAF, larger P300 amplitude in response to auditory oddball task, decreased prefrontal delta and beta cordance	
	Krepel, 2018 (40)	106 (106)	10 Hz or 1 Hz	BDI	ns		Non-replication of Arns, 2012
iAF	Widge, 2013 (50)	180 (86)	10 Hz	HDRS-17	ns		All variables were non-significant (non-replication of iAF)
	Arns, 2014 (51)	90 (90)	10 Hz or 1 Hz	BDI	.697 (for alpha), .793 (for iAF)	Decrease in Lempel-Ziv complexity (LZC) from minute 1 to minute 2 in non-responders, increase in responders and controls; predictive accuracy improved when LZC was calculated on iAF range	
	Corlier, 2019 (6)	147 (68)	10 Hz, 5 Hz or sequential bilateral (10 Hz and 1 Hz)	IDS-30 SR (response ≥40%)	r = –0.305 (adj p = .045)	A higher iAF and lower iAF distance to 10 Hz were significantly correlated with symptom improvement to 10 Hz but not to 5 Hz or bilateral rTMS	
	Roelofs, 2021 (52)	153 (59)	10 Hz or 1 Hz	BDI	r = –0.250 (p = .028)	Significant negative correlation between distance of iAF to 10 Hz and BDI percent change for 10 Hz but not 1 Hz rTMS	
ML & theta cordance	Erguzel, 2014 (56)	147 (147)	25 Hz	HDRS-17	.904 (using genetic algorithm)	ML algorithm based on delta and theta cordance can classify responders and nonresponders with high accuracy	
	Erguzel, 2016 (57)	147 (147)	25 Hz	HDRS-17	.807 - .918	Erguzel, 2014 was replicated in same sample but with added external validation and assessing different classifiers	
ML & source reconstruction	Wu, 2020 (43)	177 (152)	10 Hz or 1 Hz	DASS	rTMS: p = .004; effect size N.S. Sertraline: AUC = 0.67 (taken from Nilsonne and Harrell, 2020)	Values of SELSER algorithm below median predict better outcome to 1 Hz rTMS in anxiety subscale of DASS	
	Meijs, 2022 (45)	193 (95)	10 Hz or 1 Hz	BDI	.719 (model with baseline BDI and age)	PRS-informed fICA EEG component, reflecting delta and theta power in left DLPFC, inversely correlated with delta power in right anterior PFC, distinguishes response/non-response	
EEG functional connectivity	Zhang, 2020 (60)	179 (179)	10 Hz or 1 Hz	BDI	ns		Identified subtypes based on beta functional connectivity could not distinguish response/nonresponse for rTMS
	Bailey, 2019 (61)	71 (42)	10 Hz initially; later unilateral 10 Hz or 1 Hz or sequential bilateral	HDRS-17	Balanced accuracy: 86.6%	Responders showed higher theta connectivity (averaged across EO and EC) than controls; ML model based on 54 alpha and theta power, connectivity, iAF and theta cordance features can classify responders/non-responders with high accuracy	
	Bailey, 2021 (62)	193 (193)	10 Hz or 1 Hz or sequential bilateral	BDI	ns (Cohen's d = 0.25241)		Non-replication of Bailey, 2019: no difference between responders/non-responders in all measured variables
Theta cordance	Bares, 2015 (54)	50 (25)	1 Hz	MADRS	.82	Baseline cordance and decrease in cordance after week 1 of treatment predictive of response	
	Hunter, 2018 (55)	18 (18)	10 Hz (with potential switch to bilateral after session 10)	IDS-SR, CGI-I	Baseline ns (p = .15)		Central cordance change at week-1 but not at baseline was significantly associated with treatment outcome
	Hasanzadeh, 2021 (59)	46 (46)	10 Hz initially; last 6 sessions unilateral 10 Hz or 1 Hz or sequential bilateral	BDI, HRSD	Accuracy: 91.3% beta, 76.1% cordance		Cordance features were not significantly different between responders/non-responders
Frontal-midline theta	Li, 2016 (41)	36 (24)	10 Hz	HDRS-17	.799	Responders showed significant increase of frontal theta after RECT	
	Li, 2021 (42)	105 (70)	10 Hz or iTBS	HDRS-17	.800 (for 10 Hz), .549 (for iTBS)	Replication of Li, 2016: post-RECT frontal theta predictive of 10 Hz rTMS response but not of response to iTBS	

Group N denotes the treatment group tested for effect
ns = not significant
N.S= not specified

Sample size ≥88
Replication studies, sample size potentially <88
Robust biomarker found