Brain Stimulation 16 (2023) 1392–1400

Contents lists available at ScienceDirect

Brain Stimulation

journal homepage: www.journals.elsevier.com/brain-stimulation

A randomized trial comparing beam F3 and 5.5 cm targeting in rTMS treatment of depression demonstrates similar effectiveness

Nicholas T. Trapp [a,b,*,1], Benjamin D. Pace [a,1], Brandon Neisewander [a], Patrick Ten Eyck [c], Aaron D. Boes [a,b,d,e]

[a] Department of Psychiatry, University of Iowa, Iowa City, IA, United States
[b] Iowa Neuroscience Institute, University of Iowa, Iowa City, IA, United States
[c] Institute for Clinical and Translational Science, University of Iowa, Iowa City, IA, United States
[d] Department of Neurology, University of Iowa, Iowa City, IA, United States
[e] Department of Pediatrics, University of Iowa, Iowa City, IA, United States

ARTICLE INFO

Keywords:
TMS
Targeting
F3
Depression
Anxiety
5.5 cm rule
Clinical trial
Comparative effectiveness

ABSTRACT

Background: The Beam F3 and 5.5 cm methods are the two most common targeting strategies for localizing the left dorsolateral prefrontal cortex (DLPFC) treatment site in repetitive transcranial magnetic stimulation (rTMS) protocols. This prospective, randomized, double-blind comparative effectiveness trial assesses the clinical outcomes for these two methods in a naturalistic sample of patients with major depressive disorder (MDD) undergoing clinical rTMS treatment.

Methods: 105 adult patients with MDD (mean age = 43.2; range = 18–73; 66% female) were randomized to receive rTMS to the Beam F3 (n = 58) or 5.5 cm (n = 47) target. Between group differences from pre-to post-treatment were evaluated with the Patient Health Questionnaire-9 (PHQ-9) [primary outcome measure], Generalized Anxiety Disorder-7 (GAD-7), and clinician-administered Montgomery-Åsberg Depression Scale (MADRS). Primary treatment endpoint was completion of daily treatment series.

Results: Per-protocol analyses showed no statistically significant differences on any measure between the 5.5 cm and F3 groups (all $p \geq 0.50$), including percent improvement (PHQ-9: 39% vs. 39%; GAD-7: 34% vs. 27%; MADRS: 40% vs. 38%), response rate (PHQ-9: 37% vs. 43%; GAD-7: 27% vs. 30%; MADRS: 43% vs. 43%), and remission rate (PHQ-9: 22% vs. 21%; MADRS: 20% vs. 19%). Post hoc analysis of anxiety symptom change while controlling for depression severity suggested more favorable anxiolytic effects with 5.5 cm targeting ($p = 0.03$).

Conclusions: Similar antidepressant effects were observed with DLFPC rTMS using either the Beam F3 or 5.5 cm targeting method, supporting clinical equipoise in MDD patients with head circumference $\leq$ 60 cm. Comparison to MRI-based targeting and differential effects on anxiety symptoms require further investigation.

Trial registration: ClinicalTrials.gov identifier: NCT03378570.

1. Introduction

Repetitive transcranial magnetic stimulation (rTMS) applied to the left dorsolateral prefrontal cortex (left DLPFC) is an FDA-cleared treatment for major depressive disorder (MDD), with numerous randomized controlled trials and meta-analyses supporting its efficacy [1–3]. However, the cortical surface area defined as the left DLPFC is large (10–20 cm^2 or more depending on how it is defined [4]) and the optimal target site within the left DLPFC has been an active and evolving area of investigation for many years. The initial pivotal trials of rTMS for MDD utilized a scalp-based targeting method, here termed the "5 cm" method, whereupon the left DLPFC site was chosen as the site 5 cm anterior to the motor hand knob of the primary motor cortex [3]. Later studies suggested that this target resulted in premotor cortex stimulation in 9% of cases [5], which was a less efficacious target. More recent studies have suggested use of a "5.5 cm" or "6 cm" method to more reliably target the left DLPFC [6]. Other methods of targeting have since been introduced, which are not anchored with reference to the primary motor cortex.

* Corresponding author. University of Iowa Hospitals and Clinics, Department of Psychiatry, 200 Hawkins Drive, T223 General Hospital, Iowa City, IA 52242, United States.
E-mail address: nicholas-trapp@uiowa.edu (N.T. Trapp).
[1] Denotes both authors contributed equally to the work.

https://doi.org/10.1016/j.brs.2023.09.006
Received 10 July 2023; Received in revised form 8 September 2023; Accepted 8 September 2023
Available online 14 September 2023

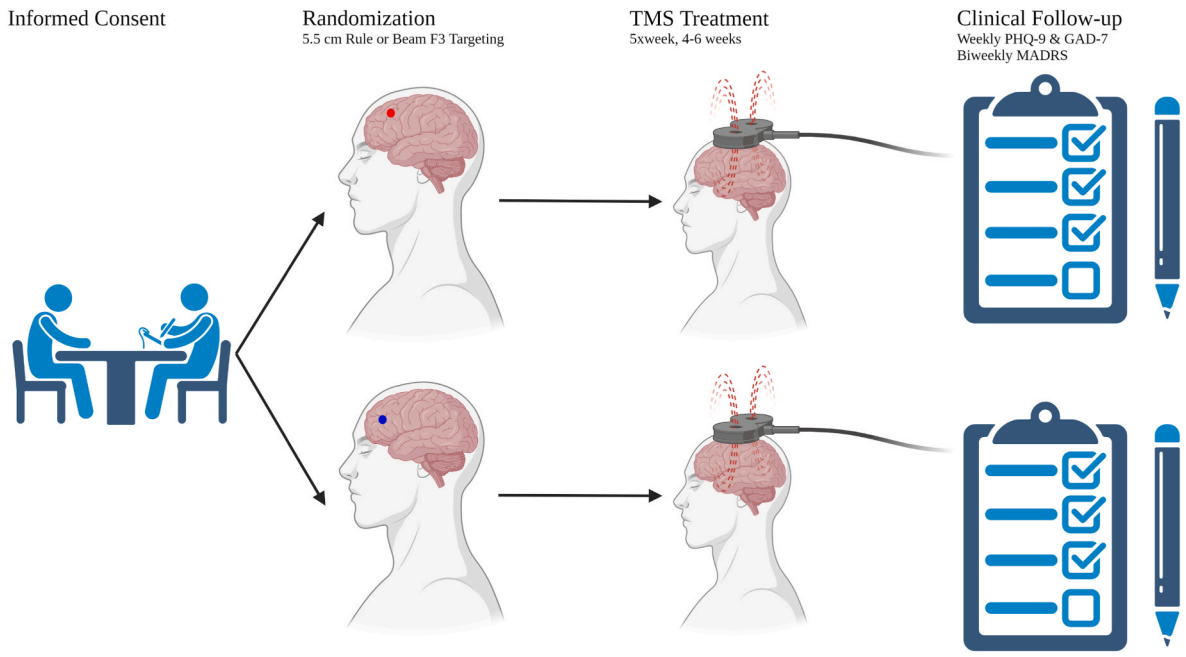

Fig. 1. Overview of study procedures.

These include scalp-based measurements that account for head circumference and size, such as the Beam F3 approach [7], as well as structural and functional neuronavigation targets. Evidence has suggested that these targets, which are often further anterior and lateral compared to the average 5 or 5.5 cm rule targets, may improve outcomes [8–12], although this has yet to be prospectively and convincingly demonstrated in a head-to-head trial. Although studies of functional connectivity-based targeting are revealing exciting early results [13–16], the current standard of practice as outlined in consensus recommendations documents on rTMS for MDD highlight the use of either the 5.5 cm rule method or the Beam F3 targeting method [6,17]. Each has unique advantages and disadvantages, and to date the two methods have not been compared head-to-head. For example, the 5.5 cm method has the distinction of being a variation of the targeting method most frequently studied in large clinical trials [2,3,18], yet the Beam F3 method is considered more reliable for identifying the same target site with repeated measurements, and also provides a closer approximation to the group average functional connectivity targets with promising early findings [8,19]. Finally, some studies suggest that both targeting methods may be effective, but for different symptom profiles within MDD [20].

In this double-blind (participant and rater-blinded), randomized comparative effectiveness trial, participants with MDD were randomized to receive either 5.5 cm targeting or Beam F3 targeting as part of their standard clinical rTMS course. We investigated whether these two targeting methods demonstrate similar treatment effectiveness. The alternative hypothesis was that Beam F3 targeting would be more effective than the 5.5 cm targeting method and lead to greater improvement in depression symptoms.

2. Methods

2.1. Participants

All participants were recruited from the University of Iowa Hospitals and Clinics (UIHC) Interventional Psychiatry Service between May 2018 and May 2022. One hundred twenty-three individuals provided written, informed consent to participate in this trial as approved by the Institutional Review Board at the University of Iowa and pre-registered on ClinicalTrials.gov (NCT03378570). Eligibility for inclusion required a primary psychiatric diagnosis of MDD and age 18 years–90 years. Diagnosis was made using DSM-5 criteria by clinical interview with a board-certified psychiatrist with additional clinical training in administration of rTMS therapy. Exclusion criteria included ferromagnetic implants in the head or neck; diagnosis of seizure disorder or epilepsy; age <18 or >90; head circumference >60 cm; or a primary psychiatric diagnosis other than MDD. Patients remained on prescribed medications, and adjustments were made as clinically indicated throughout the treatment course.

2.2. Randomization and blinding

After enrollment by a member of the research team and prior to any treatment or study-related activities, each participant was randomized in a 1:1 allocation ratio to either the 5.5 cm rule (adapted from the 5 cm rule described in Ref. [2]) or the Beam F3 [7] target (see Fig. 1 for visual workflow of trial, created with Biorender.com) using a random number generator (randomizer.org). Primary research staff, participants, and clinical assessors remained blinded to the randomization scheme for the duration of each participant's treatment course. The randomization schedule was generated and kept by the unblinded TMS technician in a locked cabinet inaccessible to members of the research team or clinical assessors. Research team members and clinical assessors were not present during treatments and were instructed not to ask participants details about their treatment targeting during evaluations. TMS technicians were instructed to provide no details of the targeting or randomization to the participants during the trial. The description of targeting methods during the consent process was intentionally vague and lacked the details necessary for participants to distinguish between the methods being applied. If desired, participants were unblinded to the targeting group assignment at the completion of the treatment course.

2.3. rTMS protocol

Subjects received either 10 Hz rTMS or intermittent theta burst stimulation (iTBS) treatment protocols. The treatment protocol (10 Hz versus iTBS) was selected independent of the participant's involvement in this study and depended on factors such as time of enrollment (six

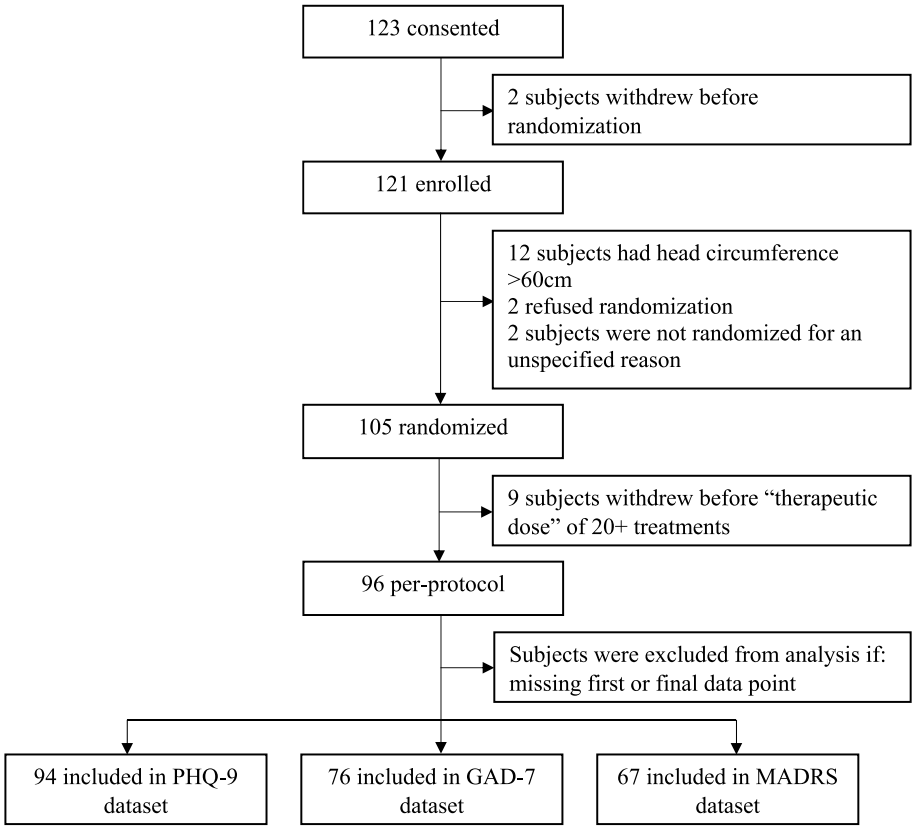

Fig. 2. CONSORT diagram.

subjects enrolled prior to FDA clearance of iTBS) and insurance policies. Trained clinical staff determined a resting motor threshold by visual observation of movement in participants' right abductor pollicis brevis from single pulses of TMS to the left motor cortex. The lowest intensity of TMS machine output that elicited three out of five responses determined the motor threshold. Weekly motor threshold testing was performed throughout the course of treatment, or more frequently if clinically indicated (e.g., medication changes). Patients receiving 10 Hz stimulation received 3000 pulses at stimulus intensity of 120% of motor threshold, 4-s pulse trains, with either an 11-s or 26-s intertrain interval. Patients receiving iTBS received 600 pulses at stimulus intensity of 120% of motor threshold, 50 Hz triplet bursts in a 5 Hz carrier frequency, 2-s pulse trains with 8-s intertrain intervals. Treatment was delivered with the MagVenture MagPro X100 using a Cool-B65 Figure 8 Butterfly Coil with Active Cooling (Magventure, Alpharetta, GA). Subjects typically received treatment 5 days a week for 4–6 weeks, occasionally followed by a taper period of 1–3 weeks.

2.4. Clinical assessments

Three well-validated scales were used to assess symptoms longitudinally over the course of treatment. This included two self-report scales, the Patient Health Questionnaire-9 (PHQ-9) [21] and General Anxiety Disorder-7 (GAD-7) [22], along with the 10-item clinician-administered Montgomery-Åsberg Depression Rating Scale (MADRS) [23]. Baseline scores for the PHQ-9 and GAD-7 were collected at the first treatment visit, and staff psychiatrists administered the baseline MADRS during the initial outpatient TMS consultation visit. While self-report assessments occurred on a weekly basis, the MADRS was administered approximately every ten treatments, or biweekly. All psychiatrists were trained in the administration of the scale, and when possible, the same physician administered the MADRS throughout an individual treatment course. The same team of physicians performed assessments for both treatment arms while remaining blinded. If an outcome assessment was not performed immediately after the final treatment in the acute series, analysis was performed with the outcome measure on the visit closest to the end of the acute treatment series (prior to taper period). PHQ-9, GAD-7, and MADRS scores were used to measure change in depression symptoms in the 5.5 cm rule and Beam F3 target groups. The primary outcome of interest was the percent change in scores on the PHQ-9 from baseline to end of acute treatment series. Secondary outcomes of interest included percent change on the GAD-7 and MADRS from baseline to end of treatment course, as well as response and remission rates for each outcome measure. GAD-7 and MADRS data collection were emphasized as secondary outcomes later in the recruitment phase based on new data suggesting that different prefrontal TMS targets may differentially affect anxiety symptoms [20] and to promote data quality and validity with partially redundant measures. Response for the PHQ-9, GAD-7, and MADRS was defined as >50% reduction in scale score relative to baseline [24–26], and remission for the PHQ-9 was defined as a score <5 and for the MADRS a score <10 [27,28]. Additional pre-specified outcome metrics including functional and structural MRI changes, NIH Toolbox Emotion Battery scores, and Montreal Cognitive Assessment scores will be analyzed, reported, and published separately.

2.5. Data analysis

A power analysis was performed prior to data collection using an effect size of 0.6 from a clinical trial that used similar self-report and clinician-administered outcome scales to compare scalp targeting to neuronavigation targeting [11]. A total sample size of 144 would have 80% power when testing for differences between target groups at alpha = 0.05. Accounting for potential attrition, the study conservatively planned to enroll 200 participants, with a planned interim analysis after 100 participants were randomized. Stopping guidelines at the interim analysis included terminating the trial if updated power calculations

Table 1
Subject demographics. Demographics for the two groups after randomization.

	5.5 (n = 47)	F3 (n = 58)
Age; mean (SD)	41.3 (15.8)	44.7 (15.1)
Sex		
Female	30 (63.8%)	39 (67.2%)
Male	16 (34.0%)	19 (32.8%)
Female to Male	1 (2.1%)	0 (0%)
Race		
White	45 (95.7%)	56 (96.6%)
American Indian/Alaska Native	0 (0%)	1 (1.7%)
Multi-racial	2 (4.3%)	1 (1.7%)
Ethnicity		
Hispanic	0 (0%)	1 (1.7%)
Non-Hispanic	47 (100%)	57 (98.3%)
Comorbidities		
Anxiety disorders	26 (55.3%)	24 (41.4%)
OCD	2 (4.3%)	4 (6.9%)
PSTD	8 (17.0%)	7 (12.1%)
ADHD	3 (6.4%)	12 (20.7%)
Borderline Personality Disorder	7 (14.9%)	5 (8.6%)
Psychiatric Medications		
Antidepressants	43 (91.5%)	55 (94.8%)
Antipsychotics	18 (38.3%)	16 (27.6%)
Benzodiazepines	17 (36.2%)	33 (56.9%)
Stimulants	9 (19.1%)	16 (27.6%)
Mood Stabilizers	8 (17.0%)	12 (20.7%)
Anxiolytics/Hypnotics	21 (44.7%)	19 (32.8%)
Number of Medication Trials (SD)	10.0 (5.4)	9.8 (4.9)
TMS Treatment Protocol		
10 Hz	10 (21.3%)	20 (34.5%)
iTBS	37 (78.7%)	38 (65.5%)
Number of TMS Treatments (SD)	31.7 (8.7)	33.6 (7.7)
Baseline PHQ-9 Score (SD)	18.8 (4.3)	17.8 (4.9)
Baseline GAD-7 Score (SD)	14.3 (4.7)	13.5 (5.2)
Baseline MADRS Score (SD)	29.1 (6.9)	29.8 (5.9)

revealed a revised sample size that would render further enrollment and trial completion impractical.

Two-tailed, independent sample t tests were employed to compare differences between groups in percent change across the PHQ-9 (primary outcome), GAD-7, and MADRS. Likewise, two-tailed $\chi 2$ tests of independence were used to compare differences between groups in response rates as determined by the PHQ-9, GAD-7, and MADRS and remission rates as determined by the PHQ-9 and MADRS. The null hypothesis of this study was that there would be no differences between targets across all measures. These analyses were conducted with SPSS (Version 28.2) and R Studio (Version 2021.09.0, Build 351) [29]. The primary analysis was a per-protocol analysis focused on participants who received at least 20 treatments in their acute treatment series, with the primary endpoint at the completion of the acute treatment series, prior to taper. Secondary analyses included a modified intent-to-treat analysis of group by timepoint interactions for each rating scale using linear mixed effect models with fixed effect predictors including categorical or continuous timepoint measurement, subject treatment group, and the interaction between the two measures; unique study ID was specified as a random effect for all models to account for repeated measures on study subjects. Modified intent-to-treat analyses were conducted on the sample of all randomized subjects having received at least one TMS treatment and having a baseline rating scale score. Post hoc analysis was performed investigating group by time interactions for GAD-7 anxiety change while controlling for depression symptom severity with PHQ-9 measures, and vice versa (assessing PHQ-9 depression symptoms while controlling for anxiety symptoms with GAD-7). There was no correction for multiple comparisons on the secondary analyses due to their exploratory nature. All tests used a p-value of <0.05 to determine statistical significance.

Table 2
Outcomes data. This table summarizes the outcomes data for the cohort of subjects who was randomized to one of the two intervention arms.

Measure & Timepoint	Total Randomized (# 5.5 cm/# Beam F3)	5.5 cm Rule	Beam F3	p-value
PHQ-9				
First Visit	104 (46/58)	18.8 (4.3)	17.8 (4.9)	0.875
Week 1	*101 (44/57)*	*15.7 (6.2)*	*16.2 (5.9)*	*0.084*
Week 2	99 (44/55)	14.6 (6.3)	14.4 (6.3)	0.277
Week 3	97 (43/54)	14.2 (6.7)	12.6 (6.4)	0.686
Week 4	94 (41/53)	13.2 (6.3)	13.0 (6.5)	0.369
Week 5	93 (42/51)	12.5 (6.3)	11.8 (7.1)	0.806
Week 6	88 (37/51)	10.7 (6.7)	11.2 (7.5)	0.231
GAD-7				
First Visit	85 (37/48)	14.3 (4.7)	13.5 (5.2)	0.772
Week 1	82 (36/46)	12.3 (5.1)	12.6 (5.2)	0.194
Week 2	80 (36/44)	10.9 (5.4)	11.3 (5.7)	0.127
Week 3	78 (34/44)	10.7 (5.7)	10.8 (5.6)	0.196
Week 4	*76 (33/43)*	*9.7 (4.9)*	*10.6 (6.4)*	*0.079*
Week 5	75 (33/42)	9.6 (5.5)	10.0 (6.5)	0.187
Week 6	*72 (30/42)*	*9.0 (5.6)*	*10.0 (6.6)*	*0.051*
MADRS				
Evaluation	99 (42/57)	29.1 (6.9)	29.8 (5.9)	0.570
Week 2-3 Follow-up	*85 (40/45)*	*24.2 (8.1)*	*22.0 (7.2)*	*0.063*
Week 4–5 Follow-up	79 (37/42)	18.4 (9.3)	19.6 (8.7)	0.792
Week 6–7 Follow-up	81 (35/46)	15.3 (8.8)	16.4 (9.7)	0.963

Table #2. Weekly Outcomes Data. The second column shows the total number of participants contributing data at each timepoint, including the number in each group. The third and fourth columns list the mean score and standard deviation for each group at each timepoint. The final column lists the p-value of the group by time interaction for each timepoint using a linear mixed effects model as described in the Methods section. Italicized rows indicate $p < 0.10$.

3. Results

3.1. Participant demographics

One hundred five participants were randomized (mean age = 43.2; range = 18–73; 66% female; n = 58 for Beam F3 target, n = 47 for 5.5 cm rule target) and 94 had data for the primary analysis (n = 52 for Beam F3 target, n = 42 for 5.5 cm target; CONSORT diagram shown in Fig. 2). Baseline demographics are shown in Table 1. The groups were similar with relation to age, gender, and most demographic factors. Secondary outcome assessments with GAD-7 and MADRS were added on at a later stage of data collection and thus have smaller samples available for analysis (n = 76 and 67, respectively). The average number of treatments delivered was 32.7 ± 7.8. All participants in the study were TMS-naïve and maintained the same total number of daily stimuli (e.g., 3000 pulses for 10 Hz, 600 pulses for iTBS) throughout the trial.

3.2. Tolerability & retention

As demonstrated in the CONSORT diagram, nine participants dropped out of the study (8.6%, Fig. 2). The drop-out rate was similar for

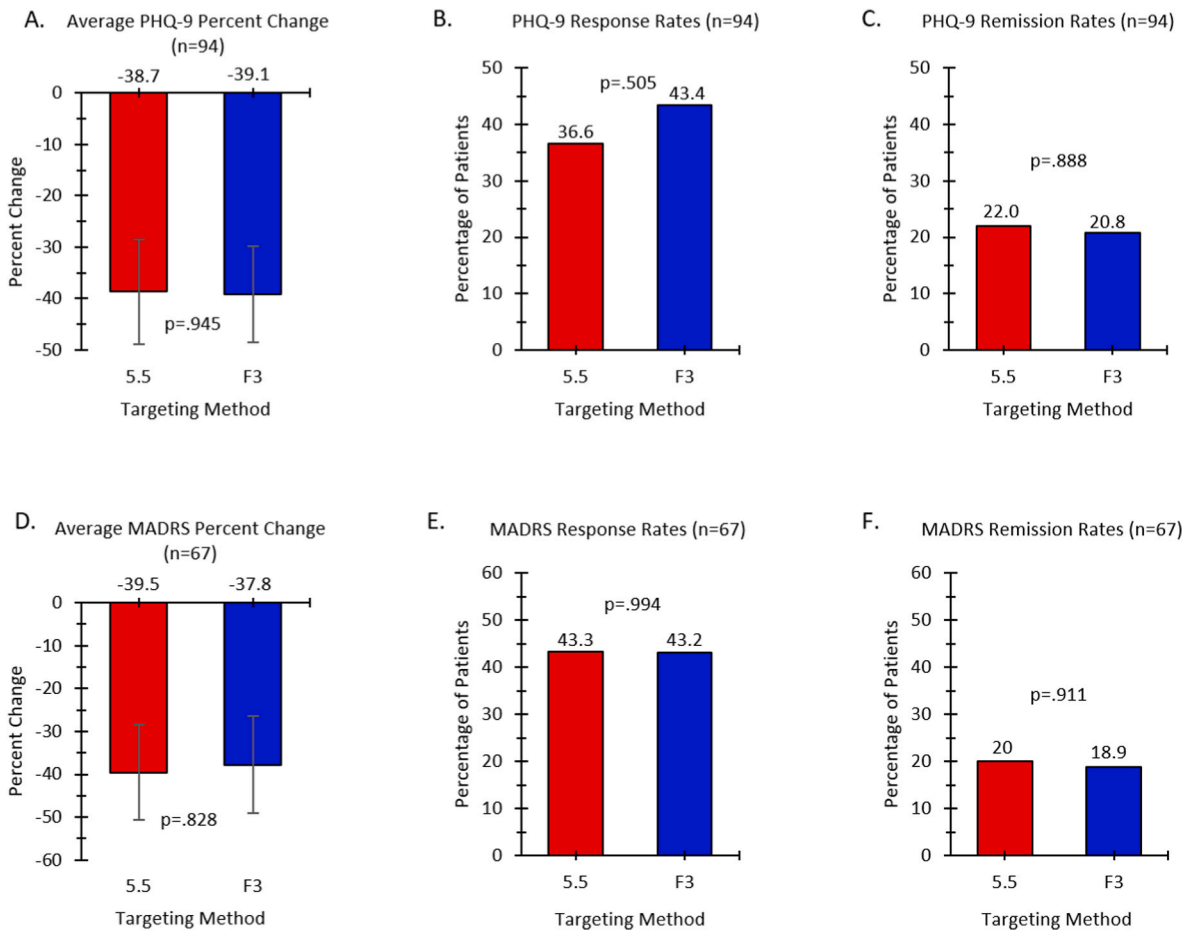

Fig. 3. Depression outcomes compared between groups. This includes comparison of the primary outcome, the Patient Health Questionnaire-9 (PHQ-9 – A,B,C) and a secondary outcome scale, the Montgomery Åsberg Depression Rating Scale (MADRS – D,E,F). There were no statistically significant differences in outcomes between the two groups on either depression outcome measure.

both the Beam F3 and 5.5 cm method (10.6% and 6.9%, respectively, p = 0.496), suggesting similar tolerability between targeting methods. No seizures occurred in either treatment group.

3.3. Depression outcomes

The depression and anxiety outcome data are summarized in Table 2. The primary outcome measure was the percent change in PHQ-9, with secondary outcomes measured with MADRS. For both the PHQ-9 and MADRS, there were no statistically significant differences between the 5.5 cm rule and Beam F3 groups in terms of percent change from baseline (Fig. 3A,D). For the PHQ-9, rates of improvement were 38.7% for the 5.5 cm group and 39.1% for the Beam F3 group (p = 0.945). Likewise, the clinician administered MADRS had similar rates of improvement for the 5.5 cm (39.5%) and the Beam F3 group (37.8%) (p = 0.828). Analysis of the modified intent-to-treat sample revealed similar outcomes (5.5 cm rule and Beam F3% change on PHQ-9 were 36.8% and 36.6%, respectively [n = 101, p = 0.968]; on MADRS were 36.6% and 37.3%, respectively [n = 91, p = 0.914]).

Similarly, this study found no significant differences between groups for PHQ-9 or MADRS-defined response and remission rates (Fig. 3B,C,E,F). The response rates on the PHQ-9 in the 5.5 cm rule group were 36.6% compared to 43.4% for the Beam F3 group (p = 0.505). Remission rates in the 5.5 cm rule group and Beam F3 group were 22% and 20.8%, respectively (p = 0.888). Meanwhile, the MADRS response rates were 43.3% for the 5.5 cm rule group and 43.2% for the Beam F3 group (p = 0.994) with remission rates for the 5.5 cm rule group and Beam F3 group of 20% and 18.9%, respectively (p = 0.911).

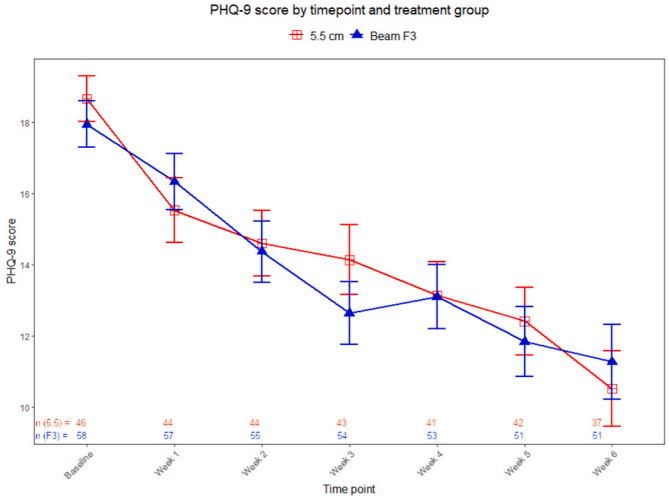

Fig. 4. PHQ-9 mean scores at each timepoint for the 5.5 cm and Beam F3 groups. This includes the data collected for all randomized subjects during the acute treatment period, with the sample size denoted along the x-axis for each group. There were no significant group by time interactions. Standard error of the mean is denoted by error bars for each score.

Linear mixed effect models did not detect any group by time interaction for the PHQ-9 (p = 0.993, Fig. 4) or MADRS (p = 0.617, Fig. 5). As part of an exploratory analysis to investigate for between-group

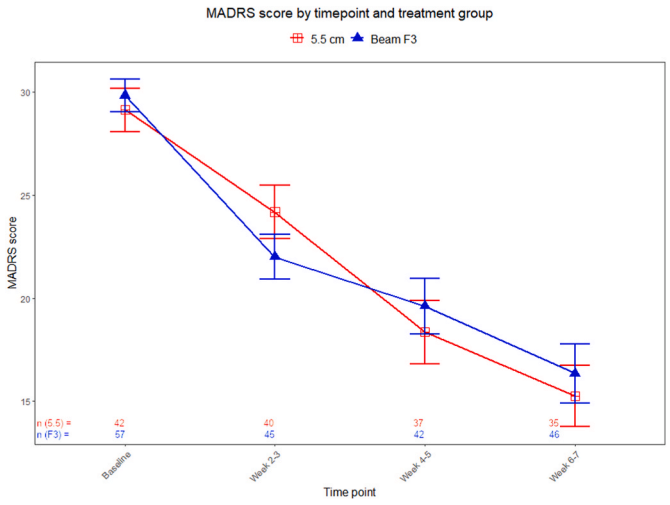

Fig. 5. MADRS mean scores at each timepoint for the 5.5 cm and Beam F3 groups. This includes the data collected for all randomized subjects during the acute treatment period, with the sample size denoted along the x-axis for each group. Due to less frequent administration of the MADRS compared to the PHQ-9 and GAD-7, the follow-up timepoints were binned based on treatment number. Week 2–3 timepoints included all MADRS scores obtained between treatment 6–15; week 4–5 timepoints included all MADRS scores obtained between treatment 16–25; and week 6–7 timepoints included all MADRS scores obtained between treatment 26–36. If two MADRS values were obtained in the same treatment window, which only occurred for the week 6–7 timepoint, the score obtained during the acute treatment series was included, and any timepoints during the taper phase of the treatment were excluded. There were no significant group by time interactions, although Beam F3 improvement at Week 2–3 was marginally significant ($p = 0.062$). Standard error of the mean is denoted by error bars for each score.

differences in the trajectory of improvement, linear mixed effects models were run on the full dataset and estimates of group differences from baseline to intermediate timepoints were analyzed for group by time interactions. This revealed a marginally significant group by time interaction on the MADRS at the week 2–3 follow-up timepoint (MADRS obtained between treatment 6 and treatment 15), favoring greater early improvement in the Beam F3 group ($p = 0.062$). No other timepoints showed significance.

3.4. Anxiety outcomes

Anxiety was primarily measured with the GAD-7 self-report measure. Between the Beam F3 and 5.5 cm method groups, there was no significant difference in percent change from baseline to completion of treatment in the per protocol sample (Fig. 6A). The GAD-7% change from baseline was 33.6% for the 5.5 cm group and 27.5% for the Beam F3 group ($p = 0.503$). Analysis of the modified intent-to-treat sample revealed similar findings (33.2% vs. 25.8% improvement in 5.5 cm group vs. Beam F3 group, respectively [$n = 82$, $p = 0.390$]). Similarly, this study found no significant differences between groups for GAD-7 response defined by >50% improvement from baseline (Fig. 6B). Response rates were 27.3% for the 5.5 cm rule group and 30.2% for the Beam F3 group ($p = 0.778$). Remission rates were not calculated for the GAD-7 as there is no standard for this in the literature and anxiety disorder was not the primary diagnosis being treated. Interestingly, the 5.5 cm rule group had lower GAD-7 scores at every timepoint after the baseline visit (Fig. 6C). Linear mixed modeling analysis showed no significant group by time interaction ($p = 0.123$); however, an exploratory, post hoc secondary analysis controlling for depression symptom severity as measured by PHQ-9 showed a significant group by time interaction favoring anxiety reduction in the 5.5 cm rule group compared to the Beam F3 group ($p = 0.030$). A similar and opposite association was seen in an exploratory, post hoc analysis of the PHQ-9 depression improvement when controlling for anxiety severity using the GAD-7, with marginally greater improvement in depressive symptoms with the Beam F3 targeting method ($p = 0.090$).

3.5. Futility analysis

A planned futility analysis was conducted using interim results to provide revised power calculations for testing study hypotheses. This included calculating the projected power at the final recruitment goal of $n = 200$, if the study were continued to completion, for each of the primary outcome measures, and also recalculating the estimated sample size needed to achieve 80% power using the interim findings. These calculations indicated the study would be underpowered when testing for differences between the two targeting strategies at the planned enrollment of $n = 200$ (Supplemental Table S1), and reaching revised enrollment goals to achieve adequate power was unfeasible in a reasonable time period. Thus, the study was terminated following the interim futility analysis.

For the purposes of promoting further investigation into the non-inferiority or equivalence of Beam F3 to 5.5 cm targeting, we present in Supplemental Table S2 mean differences and confidence intervals for the modified intent-to-treat and per-protocol samples.

4. Discussion

The current consensus standard for rTMS targeting of the left DLPFC includes either the Beam F3 or 5.5 cm targeting method, which are two of the most common methods employed in the United States. This study is the first to compare these two approaches head-to-head in a randomized clinical trial. In doing so, our results support clinical equipoise between these targeting strategies. Any detectable difference in clinical effectiveness between the Beam F3 targeting approach and the 5.5 cm targeting approach in the treatment of major depressive disorder with standard daily treatment protocols is minor, and unlikely to be clinically meaningful. Indeed, the mean absolute difference between the Beam F3 and 5.5 cm groups on the PHQ-9 total score was 0.48 points with a 95% confidence interval from −1.87 to 2.83 (see again Supplemental Table S2); this is within commonly accepted standards for equivalence on this scale [26]. The mean absolute difference was similarly small for the MADRS (mean difference 0.22, 95% confidence interval −3.48 to 3.04). The differences in the GAD-7 outcomes were greater, suggesting further investigation would be beneficial and likely require a large, multisite investigation with a larger sample size. Our planned interim futility analysis suggested that the study would be significantly underpowered ($1\text{-}\beta = 0.05$ to 0.37 depending on outcome) if carried out to enrollment target of $n = 200$ participants and was thus terminated early due to futility. Furthermore, based on these preliminary results, the number of participants needed to achieve 80% power was determined to be 1624 for the primary outcome of percent change in PHQ-9. Secondary outcomes were projected to require recruiting between 500 to over 9 million participants, depending on the measure (see again Supplemental Table S1). Although it is possible that a difference could be detected between groups with an exceptionally large sample size, this would be unlikely to translate into a clinically meaningful difference between targeting strategies.

It is interesting to interpret these results in the context of emerging neuroimaging-based targeting approaches being developed [13–15,30]. Some studies have suggested that the Beam F3 targeting approach may approximate group average MRI-guided targets within 1.36 cm in 95% of subjects [19], which is within the 2–5 cm^2 area of activation suggested by some electrical field modeling of standard Figure of 8 TMS coils [31–33]. If Beam F3 is assumed to be a closer approximation of MRI-guided targeting than the 5.5 cm rule, our results do not show this as positively impacting treatment efficacy. However, there may be other advantages to a navigated strategy not considered here, such as greater

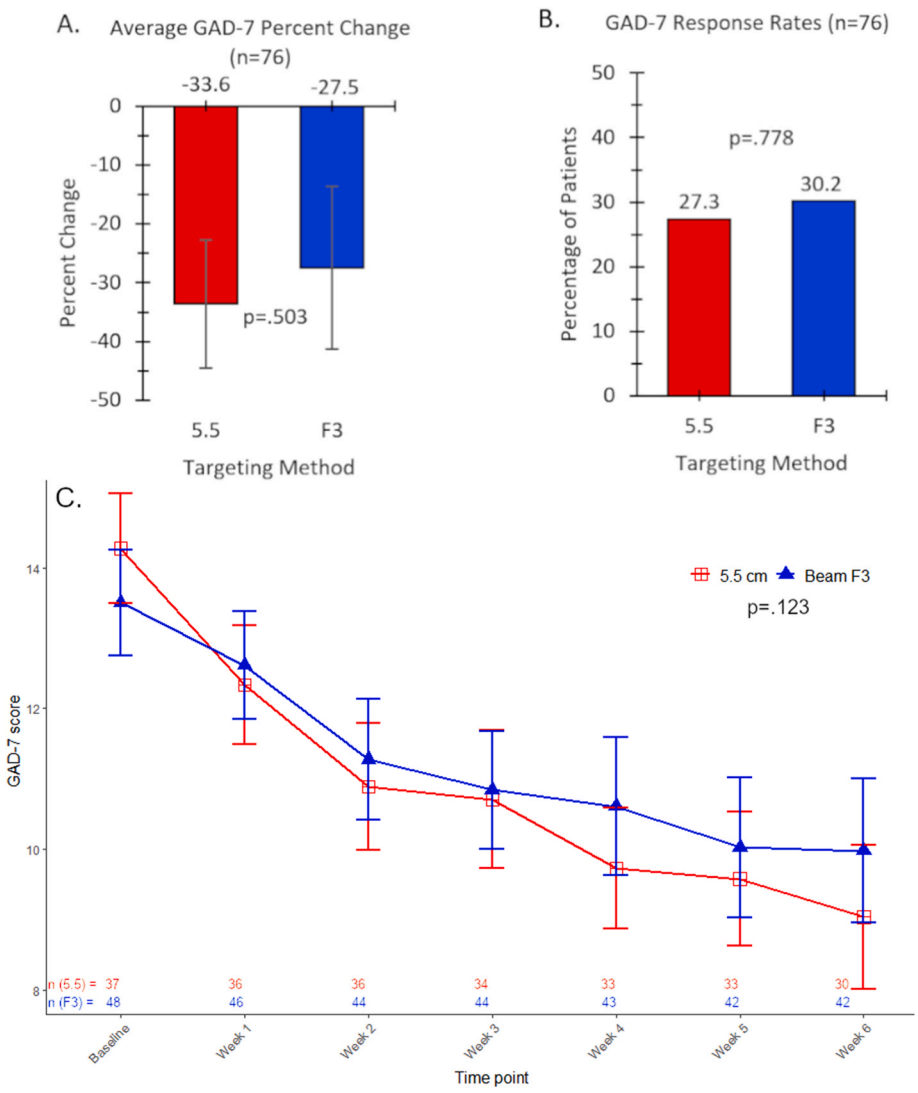

Fig. 6. Anxiety outcomes compared between groups. This includes comparison of percent change (A) and response rate (B) on the Generalized Anxiety Disorder Screener-7 (GAD-7), as well as a comparison of GAD-7 mean scores at baseline and each follow-up timepoint during the acute treatment series (C). There were no statistically significant differences in outcome between the two groups on any of the primary anxiety outcome measures ($p = 0.123$); when controlling for depression severity, anxiety improvement favors the 5.5 cm target ($p = 0.030$). In (C), the sample size is shown along the x-axis for each group, and standard error of the mean is denoted by error bars for each score.

consistency of reaching the same target with each treatment session. Caution is also warranted, as the spatial specificity of TMS remains unclear, and this limits our ability to infer whether being within 1.36 cm of a treatment target is close enough. Studies in non-human primates and pre-surgical TMS cortical mapping suggesting a much smaller region of activation (~2–5 mm^2), as validated by single unit recordings or direct cortical stimulation [33–35]. Additionally, the Beam F3 target may be further away from individualized functional connectivity-based DLPFC targets proposed in other recent trials [13].

An intriguing area for future research should include further investigation of differential effects of DLPFC targeting on anxiety symptoms. This study lends insight into the utility of targeting methods for different brain circuits that may be associated with specific symptom clusters of depression. In general, the 5.5 cm targeting approach appears to identify a target near a network node associated with preferential improvement in anxious and somatic symptoms, whereas the Beam F3 targeting approach approximates a target more associated with cognitive and dysphoric symptoms of depression [8,20]. When controlling for depressive symptom burden, our analysis showed a significant group by time interaction favoring the 5.5 cm target for improvement in anxiety symptoms, as would be expected based on the work by Siddiqi and colleagues referenced above [20]. This difference was small, the analysis was exploratory and not corrected for multiple comparisons, and the study here was not powered to look at this specific question. Nonetheless, the results are worthy of follow-up study, especially in light of the fact that individualized targeting was not utilized, and the 5.5 cm method may not be the best strategy for approximating the "anxiosomatic" circuit of interest. Further studies on rTMS targeting of symptom-specific depression networks are underway (e.g., NCT04604210).

This study had several limitations. As this was designed to focus on comparative clinical effectiveness in a real-world sample, the study included many people with significant psychiatric comorbidities and did not standardize or control medication changes during the trial. Although in general our clinical team tends to minimize medication changes during an rTMS treatment course, this variable may have limited the study's ability to detect rTMS-specific differences in outcomes between groups. The study sample was also predominantly white and non-Hispanic, and thus it is unclear if these results would generalize to other demographic populations.

Similarly, subjects were not randomized into iTBS or 10 Hz, and assignment to iTBS or 10 Hz stimulation was determined often by variables beyond the control of the clinical or research team, such as insurance policies limited to the 10 Hz protocol for some patients. Although this introduced another uncontrolled variable, prior large-scale studies have suggested that iTBS is non-inferior to 10 Hz [36], and prior published analysis of our own data has also suggested no difference in outcomes between the two treatment approaches [37].

Another limitation was lack of assessment of blinding. Participants were not told their randomization arm, but it is possible that some participants could have independently read up on the differences between targeting methods and thus compromised the blind. Although not systematically assessed, the authors have no knowledge of this occurring, and participants were educated about the lack of known superiority between targeting methods during the consent process, which may have reduced bias towards a particular method.

Additionally, our study excluded participants with a head circumference >60 cm. This was due to the research and clinical team's concern that patients with larger head circumferences were at greater risk for receiving premotor cortex treatment with the 5.5 cm approach [5]. This excluded a small sample of subjects (n = 12), most who went on to receive Beam F3 targeting as clinically indicated. However, this exclusion could have biased the primary study outcome toward the null hypothesis by excluding subjects who potentially stood most to benefit from Beam F3 targeting as opposed to 5.5 cm targeting. Thus, our conclusions cannot be applied to patients with head circumference >60 cm. To further evaluate this, a post hoc analysis of the PHQ-9, MADRS, and GAD-7 outcomes for the subjects who were excluded from the randomization due to head size demonstrated similar findings to the randomized groups (34.1%, 37.9%, and 22.8% improvement in PHQ-9, MADRS, and GAD-7, respectively). Indeed, these changes on average are similar to but lower than the improvements noted in either randomized group, despite this sample being given Beam F3 targeting to account for head size. Adjusting DLPFC targeting to account for head size, at least in this small subset of patients with larger head circumferences, therefore did not seem to have a large impact on outcomes.

Finally, due to the naturalistic sample of interest, a cut-off of having received at least 20 treatments was used for inclusion in the primary analysis. This number was chosen to try to maximize the included sample while excluding patients who were likely "underdosed" by dropping out of the study early. Nonetheless, evidence suggests there may be a dose-response relationship for rTMS in depression, and therefore inclusion of participants receiving 20 treatments may represent "underdosing" by current standards. This likely explains the somewhat lower response and remission rates relative to other studies reported from our group when a later timepoint was used for the final outcome measure [37] and others reported in the literature [18,36]. This may have biased our primary results towards a null hypothesis. However, secondary analysis of the sub-set of patients who received 30 sessions or more still did not demonstrate a statistically significant difference between groups, and analyses conducted on the intent-to-treat sample with all randomized subjects also showed non-significant findings (p = 0.390–0.968).

In conclusion, this study presents the results of a randomized, double-blind trial comparing rTMS outcomes in MDD for Beam F3 and 5.5 cm rule targeting of the left DLPFC. The study demonstrates that Beam F3 and 5.5 cm rule targets each achieve remarkably similar antidepressant effects as measured with both self-report and clinician-administered outcome measures. Additional research should focus on comparing these targeting methods to individualized functional connectivity-based targeting of the left DLPFC, and also on whether there are symptom specific changes unique to different targeting methods or subregions of stimulation, as may be suggested by the differential changes in anxiety scores here.

Funding

This research was supported by grants from the National Institutes of Mental Health (1K23MH125145 and 5T32-MH019113 to NTT), the National Institute of Neurological Disease and Stroke (R01 NS114405-02 to ADB), the Brain and Behavior Research Foundation (Young Investigator Grant 31275 to NTT), the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number (UM1TR004403), the Adamfest Research Award (NTT), the University of Iowa Department of Psychiatry Pilot Grant (NTT), and the Roy J. Carver Trust (ADB).

CRediT authorship contribution statement

Nicholas T. Trapp: Writing – original draft, Conceptualization, Investigation, Formal analysis. **Benjamin D. Pace:** Data curation, Writing – original draft. **Brandon Neisewander:** Writing – review & editing. **Patrick Ten Eyck:** Methodology, Formal analysis. **Aaron D. Boes:** Conceptualization, Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to acknowledge all the participants who volunteered their time to participate in this research. We would also like to acknowledge Christopher Sanborn, Brandt Uitermarkt, and Lara Fegley who assisted with the randomization process and patient treatments. We would like to acknowledge Eric Tsang and Mark Bowren for their consultation on statistical methods.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brs.2023.09.006.

References

[1] Dalhuisen I, et al. The influence of sample and treatment characteristics on the efficacy of repetitive transcranial magnetic stimulation in depression: a meta-analysis and meta-regression. Brain Stimul 2021;14(6):1671.
[2] George MS, et al. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. Arch Gen Psychiatr 2010;67(5):507–16.
[3] O'Reardon JP, et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. Biol Psychiatr 2007;62(11):1208–16.
[4] Donahue CJ, et al. Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. Proc Natl Acad Sci U S A 2018;115(22):E5183–92.
[5] Johnson KA, et al. Prefrontal rTMS for treating depression: location and intensity results from the OPT-TMS multi-site clinical trial. Brain Stimul 2013;6(2):108–17.
[6] McClintock SM, et al. Consensus recommendations for the clinical application of repetitive transcranial magnetic stimulation (rTMS) in the treatment of depression. J Clin Psychiatry 2018;79(1).
[7] Beam W, et al. An efficient and accurate new method for locating the F3 position for prefrontal TMS applications. Brain Stimul 2009;2(1):50–4.
[8] Trapp NT, et al. Reliability of targeting methods in TMS for depression: Beam F3 vs 55 cm. Brain Stimul 2020;13(3):578–81.
[9] Rusjan PM, et al. Optimal transcranial magnetic stimulation coil placement for targeting the dorsolateral prefrontal cortex using novel magnetic resonance image-guided neuronavigation. Hum Brain Mapp 2010;31(11):1643–52.
[10] Herbsman T, et al. More lateral and anterior prefrontal coil location is associated with better repetitive transcranial magnetic stimulation antidepressant response. Biol Psychiatr 2009;66(5):509–15.
[11] Fitzgerald PB, et al. A randomized trial of rTMS targeted with MRI based neuronavigation in treatment-resistant depression. Neuropsychopharmacology 2009;34(5):1255–62.
[12] Fitzgerald PB, et al. Exploring the optimal site for the localization of dorsolateral prefrontal cortex in brain stimulation experiments. Brain Stimul 2009;2(4):234–7.
[13] Cole EJ, et al. Stanford neuromodulation therapy (SNT): a double-blind randomized controlled trial. Am J Psychiatr 2022;179(2):132–41.
[14] Siddiqi SH, et al. Identification of personalized transcranial magnetic stimulation targets based on subgenual cingulate connectivity: an independent replication. Biol Psychiatr 2021;90(10):e55–6.
[15] Cash RFH, et al. Functional magnetic resonance imaging-guided personalization of transcranial magnetic stimulation treatment for depression. JAMA Psychiatr 2021;78(3):337–9.
[16] Cash RFH, et al. Using brain imaging to improve spatial targeting of transcranial magnetic stimulation for depression. Biol Psychiatr 2021;90(10):689–700.

[17] Perera T, et al. The clinical TMS society consensus review and treatment recommendations for TMS therapy for major depressive disorder. Brain Stimul 2016;9(3):336–46.
[18] Carpenter LL, et al. Transcranial magnetic stimulation (TMS) for major depression: a multisite, naturalistic, observational study of acute treatment outcomes in clinical practice. Depress Anxiety 2012;29(7):587–96.
[19] Mir-Moghtadaei A, et al. Concordance between BeamF3 and MRI-neuronavigated target sites for repetitive transcranial magnetic stimulation of the left dorsolateral prefrontal cortex. Brain Stimul 2015;8(5):965–73.
[20] Siddiqi SH, et al. Distinct symptom-specific treatment targets for circuit-based neuromodulation. Am J Psychiatr 2020;177(5):435–46.
[21] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–13.
[22] Spitzer RL, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092–7.
[23] Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–9.
[24] Kroenke K, et al. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatr 2010;32(4):345–59.
[25] Duru G, Fantino B. The clinical relevance of changes in the Montgomery-Asberg Depression Rating Scale using the minimum clinically important difference approach. Curr Med Res Opin 2008;24(5):1329–35.
[26] Lowe B, et al. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004;42(12):1194–201.
[27] Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ (Can Med Assoc J) 2012;184(3):E191–6.
[28] Hawley CJ, et al. Defining remission by cut off score on the MADRS: selecting the optimal value. J Affect Disord 2002;72(2):177–84.
[29] Team RCR. A language and environment for statistical computing. 2021. Available from: https://www.R-project.org/.
[30] Siddiqi SH, et al. Repetitive transcranial magnetic stimulation with resting-state network targeting for treatment-resistant depression in traumatic brain injury: a randomized, controlled, double-blinded Pilot study. J Neurotrauma 2019;36(8):1361–74.
[31] Deng ZD, et al. Electric field modeling for transcranial magnetic stimulation and electroconvulsive therapy. Brain and Human Body Modeling: Computational Human Modeling at EMBC 2018 2019:75–84.
[32] Deng ZD, Lisanby SH, Peterchev AV. Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. Brain Stimul 2013;6(1):1–13.
[33] Romero MC, et al. Neural effects of transcranial magnetic stimulation at the single-cell level. Nat Commun 2019;10(1):2642.
[34] Paiva WS, et al. Cortical mapping with navigated transcranial magnetic stimulation in low-grade glioma surgery. Neuropsychiatric Dis Treat 2012;8:197–201.
[35] Krieg SM, et al. Utility of presurgical navigated transcranial magnetic brain stimulation for the resection of tumors in eloquent motor areas. J Neurosurg 2012;116(5):994–1001.
[36] Blumberger DM, et al. Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. Lancet 2018;391(10131):1683–92.
[37] Spitz NA, et al. Similar outcomes in treating major depressive disorder with 10 Hz repetitive transcranial magnetic stimulation (rTMS) versus intermittent theta burst stimulation (iTBS): a naturalistic observational study. J Psychiatr Pract 2022;28(2):98–107.