Chapter 5

No antidepressive effects of transcranial pulsed electromagnetic fields for treatment resistant depression – a replication randomized controlled trial

S.M. van Belkum, M.K. de Boer, E.M. Opmeer, R. Kortekaas, F. Woonings, H.J.R. Hoenders, H. Kamphuis, A. Aleman, R.A. Schoevers

Submitted.

Background
Noninvasive neurostimulation with transcranial Pulsed Electromagnetic Fields (tPEMF) is a promising method for the treatment of treatment resistant depression (TRD). An earlier RCT has shown substantial improvement of depressive symptoms in patients with TRD but this has not been replicated yet. Furthermore, there is no information on long-term antidepressive effects. The aim of this study was to investigate the short- and long-term efficacy of tPEMF in participants with TRD.

Methods
Eligible participants with TRD in this sham-controlled double-blind multicenter trial were randomly assigned to five weeks either daily active or sham tPEMF. Severity of depression and anxiety was assessed pre- and directly post-treatment and five and fifteen weeks post-treatment. Primary outcome was change on the 17-item Hamilton depression rating scale directly post-treatment. Secondary outcome was change on the Hamilton-17 during follow-up and change on the Inventory of Depressive Symptomatology Self-Report and the Beck Anxiety Index.

Results
Of the 55 included participants, 50 completed the treatment protocol. Depressive symptoms improved over time, independent of treatment type. The improvement continued after until the last follow-up measure. There was no difference in outcome between the active and the sham group on change in depression post-treatment or on any secondary measure.

Conclusion
Treatment with active tPEMF was not superior to sham in patients with TRD. This is in contrast to a previous study using a similar design and power calculation that reported improvement of depression after treatment with tPEMF compared to sham.

The trial was registered at the Dutch Trial Register (http://www.trialregister.nl), NTR3702.

1. Introduction

Treatment of depression is often challenging; up to one third of patients suffering from a severe major depressive disorder (MDD) do not respond to four consecutively prescribed antidepressants (Rush et al. 2006) and are suffering from Treatment Resistant Depression (TRD). TRD is the main cause for the large societal costs of depression (Greden 2001; Ivanova et al. 2010), making it paramount to improve treatment efficacy of MDD. To do so, new treatment possibilities are being investigated, of which non-invasive neurostimulation is of growing interest (Holtzheimer and Mayberg 2012).

Non-invasive neuromodulation for depression can be categorized into two broad categories: local or global modulation. Local modulation relies on modulation of local brain regions, for example using repetitive Transcranial Magnetic Stimulation (rTMS) (Allan, Herrmann, Ebmeier 2011). In rTMS, modulation of local brain regions is achieved by depolarization of the neuronal membrane by inducing electric currents in the brain (Barker, Jalinous, Freeston 1985), rTMS has become an established treatment for TRD and has been included in the NICE-guidelines (https://www.nice.org.uk/guidance/ipg542).

Global modulation of the brain refers to weak electromagnetic stimulation at multiple scalp sites simultaneously or with a more or less homogeneous magnetic field (Rohan et al. 2004; Rohan et al. 2013; van Belkum et al. 2016). An important development

in this field of research was the study conducted by Martiny et al. (Martiny, Lunde, Bech 2010). This research group adapted a magnetic stimulation method mostly used in orthopedics dubbed 'Pulsed Electro-Magnetic Fields' (PEMF) (Hannemann et al. 2014; Ryang We et al. 2013). Martiny et al. have applied this treatment transcranially (tPEMF) in patients with TRD and investigated the efficacy in a Randomized Controlled Trial (RCT). After five consecutive weeks of daily stimulation, depression severity (measured with the Hamilton Depression Rating Scale-17 (HAMD-17)) decreased significantly more in the active stimulation group compared to sham stimulation (Martiny, Lunde, Bech 2010).

Other RCTs have also shown a positive effect of global neuromodulation using pulsed electromagnetic fields, although different parameters with regard to pulse frequency, field strength and amount of coils were used (Leuchter et al. 2015; Rohan et al. 2013). For example, in one RCT the effect of a portable electromagnetic device producing rapidly oscillating electromagnetic fields was investigated. An immediate positive effect on depression severity in patients with a unipolar or bipolar depression was found after a single treatment-stimulus (Rohan et al. 2013). Another study has used a device with three rotating magnets (synchronized TMS or sTMS) and also showed some antidepressive effects (Leuchter et al. 2015).

Global stimulation adds an interesting branch to the expanding antidepressive neuromodulation tree, but only a few RCTs investigating global stimulation with weak electromagnetic fields have been reported, all differing in their stimulation parameters. Up until now only one RCT has investigated the specific effects of tPEMF on depression (Martiny, Lunde, Bech 2010) with no information on the long-term duration of the antidepressive effect.

Using a lightweight neurostimulator, that previously was found to be effective against experimental pain in healthy subjects (Kortekaas et al. 2013), we aimed to replicate the study of Martiny et al. (Martiny, Lunde, Bech 2010) of antidepressive effects in TRD. Moreover, we aimed to investigate long-term effects, and to evaluate the effect of tPEMF on the brain (reported separately).

2. Methods and materials

2.1. Study design

We included 55 depressed participants in a double blind, randomized controlled multicenter trial comparing active tPEMF treatment versus sham treatment in a 1:1 ratio, in three mental health institutions in the north of The Netherlands (Department of Psychiatry of the University Medical Center Groningen (UMCG), the mental healthcare provider GGZ Drenthe, and the Department of Psychiatry of the general hospital in Sneek). This study was approved by the Medical Ethical Committee of the UMCG, and at the study coordination center of each participating site. Written informed consent was obtained from each participant. The study was conducted according to the Declaration of Helsinki. The trial was registered at the Dutch Trial Register (http://www.trialregister.nl), part of the Dutch Cochrane Centre, under number NTR3702.

2.2. Study population

We recruited patients at major mental health care institutions (regular and academic mental health care institutions) in the northern part of the Netherlands and via media coverage. We included patients who met DSM-IV criteria for MDD and who were at the time in a first or recurrent depressive episode, assessed by the Mini-International Neuropsychiatric Interview (MINI) (Sheehan et al. 1998). Inclusion criteria were presence of at least a moderately severe depression (>17 on HAMD-17), non-responsiveness to one or more antidepressants, given for at least four weeks and in an adequate dose (i.e. the defined daily dose (DDD) (WHO Collaborating Centre for Drug Statistics Methodology 2016)) during the current episode, age between 18 and 80 years, and having a good understanding of the Dutch language (including writing skills). We included both in- and outpatients.

Exclusion criteria were presence of MDD with psychotic features, other major psychiatric disorders such as a primary psychotic disorder or an antisocial or borderline personality disorder, a neurological disorder such as dementia or epilepsy, visual or hearing problems that could not be corrected, suicidal thoughts (>2 on HAMD-17 for suicidal ideation) or a history of a serious suicide attempt, recent (past three months) alcohol or drug abuse or dependence, pregnancy, lactation, inability to comply with treatments and/or assessments, recent change (last four weeks) in antidepressant medication or requirement to change antidepressant medication during the course of the study, use of benzodiazepine(s) more than 2 mg lorazepam or equivalent per day within the last four weeks or during the course of the study, use of medication indicated for a somatic disease that may have affected mood within the last four

weeks, excessive use of coffee (>10 units per day) or alcohol (>5 units per day), or recent use (within four weeks) of cannabis or any other non-prescribed psychotropic drugs or unwillingness to abstain from these substances during the study. The use of antipsychotics and lithium was allowed. Because of the use of additional magnetic resonance imaging (MRI) (results will be reported elsewhere), there were additional exclusion criteria related to MRI incompatibility, for example the presence of metal implants in the body.

2.3. Treatment protocol

Eligible participants were randomly assigned to either five weeks active tPEMF stimulation or five weeks of sham stimulation. One of the authors (SvB) enrolled participants. Stimulation was administered by trained members of the research team who were present during the whole session, under medical supervision of one of the authors (SvB). Two identical tPEMF-stimulators were used for treatment at the different treatment sites and were moved if necessary. During a session, participants were seated next to the PEMF-stimulator in a quiet room while wearing the treatment cap. There were no restrictions for participants during these sessions and talking was allowed on their own initiative. Sessions took place every weekday for 30 minutes during office-hours, on the same time every day, with minimal deviations.

Properties of the magnetic stimulator have previously been published in detail (Kortekaas et al. 2013). A laptop computer (Dell Latitude D610, Round Rock, Texas,

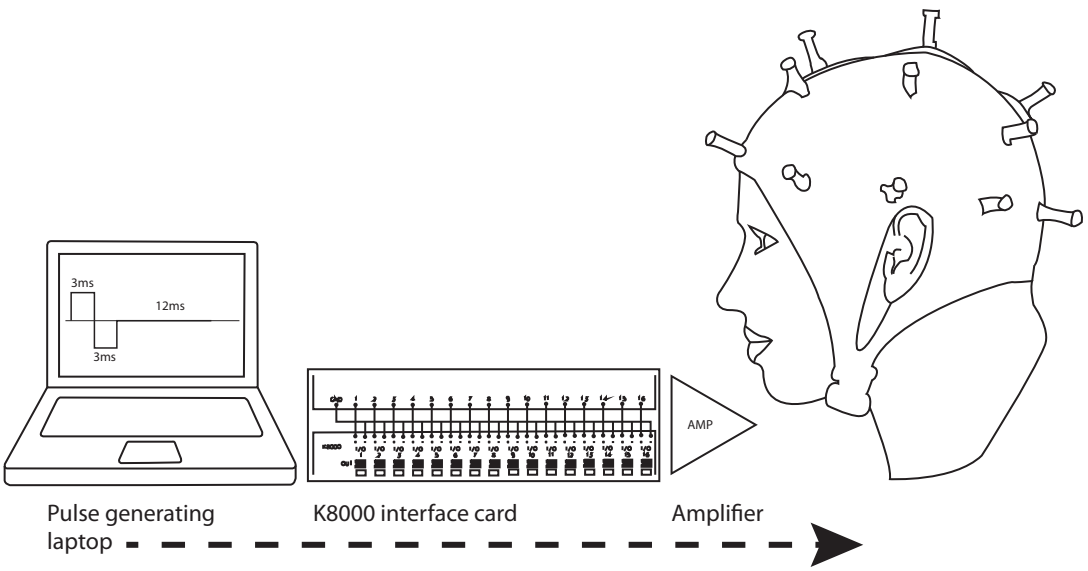


Figure 5.1: Schematic overview of the hardware. The interface card translates digital values into values. The amplifier in turn increases power.

USA) and interface card (K8000, Velleman, Gavere, Belgium) were used as a waveform generator (see figure 1). The computer ran Xubuntu Linux (www.xubuntu.org). A low voltage DC coupled amplifier with a medical power supply and an isolation unit as additional safety features was used to increase the output power. For the active condition, alternating bipolar square pulses of 7 V were used as input, equal to the bipolar pulses used in the stimulation set of Martiny et al. (Martiny, Lunde, Bech 2010). The stimulation pattern consisted of 3 ms north and 3 ms south and 12 ms pause, thus lasting 18 ms in total. For the sham condition no pulse was generated; only a signal filled with zeroes. The stimulator did not produce sound, heat, nor skin sensations. It was thus impossible for participants and the research team to distinguish between the active and sham condition.

The electromagnets of the cap consisted of 25 mm long, 9 mm thick reed relays (Reed Relay 275–232, Radio Shack, Fort Worth, TX, USA) of which the reed switch was replaced by a steel bolt, transforming them into iron core electromagnets. Nineteen of these electromagnets were radially attached to a regular EEG cap with a chinstrap (SU-60 and KR, MedCaT, Erica, The Netherlands) using non-metallic nuts on the inside of the cap. Electromagnets were positioned according to the international 10/20 system for EEG electrodes.

2.4. Stratification, randomization and blinding

Participants were stratified by duration of the depressive episode (less or more than one year) and depression severity (HAMD-17 baseline score between 18 and 25 or 25 and higher), resulting in four strata. We did not adopt a minimization procedure (Pocock and Simon 1975). One of the authors (SvB) assigned each participant a unique subject code, composed of two parts: the first indicating the number of the stratum (1 – 4) and the second indicating the sequence of enrollment. Thus, the first subject in stratum 1 received the number 'stratum 1 participant 1', or, in abbreviated form: s1p1. These unique codes corresponded to the names of different data files on the pulse generating PC. Each data file contained a description of one of the two treatment waves that were offered: an active wave or a sham 'wave'. These data files were a direct copy of either the active master file or the sham master file and were automatically and randomly created in the preparation phase of the study by a computerized random number generator under responsibility of one of the authors (RK). Due to the sequential numbering of the data files, which were identical in appearance and were contained in an inaccessible folder on the pulse generating PCs, allocation to the treatment was adequately concealed. In order to administrate treatment, members of the research team had to enter the unique subject code. The participant, researchers and health-care personnel were all blind for the treatment condition. To assess adequate blinding,

participants were asked to guess which treatment they received. The code was broken after the last participant had completed the last measurement.

2.5. Study Outcome and psychometrics

The primary outcome was change in depression severity measured by the HAMD-17 (Hamilton 1960) immediately post-treatment. Secondary outcome measures consisted of changes in depression severity as assessed at five and fifteen week follow-up with the HAMD-17. Furthermore, we calculated response (50% improvement of HAMD-17) and remission (HAMD-17 < 8) rates, assessed weekly changes in depression severity during treatment and at five and fifteen week follow-up with the self-rated Inventory of Depressive Symptomatology Self-Report (IDS-SR) (Rush et al. 1986; Rush et al. 1996), and assessed changes of anxiety symptoms as measured with the Beck Anxiety Index (BAI) (Beck et al. 1988). At baseline, an expectancy scale with regard to the effect of treatment was administered, ranging from one to ten, one meaning low expectations and ten meaning high expectations. The degree of treatment resistance was quantified using the Maudsley Staging Method (MSM), a sum score based on duration and severity of illness and treatment history of the current episode (Fekadu, Wooderson et al. 2009b). A higher score is associated with a worse depression outcome (Fekadu, Wooderson et al. 2009a; Fekadu, Wooderson et al. 2009b; van Belkum et al. 2018).

2.6. Statistical analyses

Sample size calculation was based on change in HAMD-17 scores between baseline (week -1) and directly post-treatment (week 5) as reported in a previous publication (Martiny, Lunde, Bech 2010). Assuming a two-sided alpha level 0.05, and a beta of 0.8, we calculated n=25 per group. Participants that dropped out were replaced so that the total sample at follow-up consisted of 50 participants. Therefore, the total number of participants that started the study proportionally increased. Data were analyzed using the intent-to-treat (ITT) principle so that all randomized participants were included in the analyses.

Analyses were performed with IBM SPSS version 24.0 software (IBM, Chicago IL, USA). Baseline characteristics in each group were compared with Chi-square test, independent t-test or the Mann-Whitney U test where appropriate.

To test for the effect of treatment on the main outcome measure, a linear mixed model with a random intercept was applied with post-treatment HAMD-17 score as dependent variable and treatment group, time (baseline, week 5), the interaction between time and treatment group, and baseline HAMD-17-score as covariates.

Secondarily, we added the remaining time points (week 10 and week 20) as covariates. Subsequently we corrected for duration, type of episode (single or recurrent), number of episodes, treatment expectancy, and treatment guess by adding these variable as covariate to the model. Post-hoc we applied a linear mixed model in the four different strata, equal to the analysis of the main outcome. To test for the effect of treatment on the secondary outcome measures, we applied two different linear mixed models with a random intercept with 1) post-treatment IDS or 2) post-treatment BAI scores as dependent variable, and treatment group, time, and the interaction between time and treatment group as covariates for both dependent variables. The level of statistical significance was set at $\alpha < 0.05$.

3. Results

3.1. Sample description

Between May 2013 and October 2016 we included 55 participants, 27 female (49%). The trial ended after the aimed sample-size was reached. Participants were randomized to either the active treatment group (n = 29) or to sham treatment (n = 26). After randomization and before starting the first treatment session, two participants (both randomized for the active treatment) refused further participation. Three participants dropped out during the study: one (in the active treatment group) dropped out less than a week after starting the treatment sessions due to admittance to a closed ward because of severe suicidal ideations. Retrospectively, it became clear that these suicidal ideations were already present on baseline but had not been reported by this participant. Two other participants (one active, one sham) discontinued intervention due to absence of subjective treatment effect. Two participants did not attend the appointment at the 20-week follow-up measurement. In total, 50 participants completed all treatment sessions, 25 in each group. Data from all 55 participants were analyzed. See supplemental for the CONSORT flow diagram.

Table 5.1: Sociodemographic and clinical parameters.

Variable	Active (n = 29)	Sham (n = 26)	p-value
Age (years) (mean (SD))	49 (13)	45 (12)	.309[a]
Female gender	15 (52%)	12 (46%)	.680[b]
Marital status			.422[b]
Single	10 (34%)	9 (35%)	
Married	15 (52%)	16 (62%)	
Divorced	4 (14%)	1 (4%)	
Educational background			.460[b]
Primary	2 (7%)	0 (0%)	
Lower secondary	9 (31%)	11 (42%)	
Upper secondary	13 (45%)	12 (46%)	
University	5 (17%)	3 (12%)	
Presence of somatic complaints	18 (62%)	21 (81%)	.127[b]
MDD-type			.324[b]
MDD first episode	14 (48%)	16 (62%)	
MDD recurrent episode	15 (52%)	10 (38%)	

Variable	Active (n = 29)	Sham (n = 26)	p-value
Number of episodes (median (IQR))	2 (1 - 2.5)	1 (1 - 3)	.746[c]
Duration of current episode (months) (median (IQR))	23 (16 - 66)	33 (12 - 107)	.468[c]
Presence of comorbidity	14 (48%)	11 (42%)	.657[b]
Anxiety Disorders[d]	9 (31%)	6 (23%)	.508[b]
Personality Disorders[e]	5 (17%)	5 (19%)	.849[b]
Miscellaneous[f]	2 (7%)	0 (0%)	.173[b]
MSM-score (mean (SD))	7.8 (1.60)	8.3 (2.29)	.402[a]
Treatment expectancy (scoring 1-10) (mean (SD))	5.8 (2.3)	6.4 (2.1)	.373[a]
Correct guess to treatment allocation	12 (48%)	9 (36%)	.254[b]
Stratum			.986[b]
S1	5 (17%)	5 (19%)	
S2	20 (69%)	18 (69%)	
S3	1 (3%)	1 (4%)	
S4	3 (10%)	2 (8%)	

[a] 2-tailed t-test.

[b] Chi-square.

[c] Mann-Whitney U.

[d] General Anxiety Disorder, Panic Disorder, Social Phobia, Obsessive Compulsive Disorder, Post-traumatic stress disorder, and Not Otherwise Specified.

[e] Avoidant Personality Disorder, Obsessive–compulsive personality disorder, and Not Otherwise Specified.

[f] Asperger's disorder, Attention deficit hyperactivity disorder.

Abbreviations: MDD = major depressive disorder, IQR = interquartile range, MSM = Maudsley Staging Method.

Table 5.1 shows that the treatment groups were similar in socio-demographic data and clinical measurements (recurrence and number of episodes, duration of present episode, presence of comorbidity and the MSM-score). On the expectancy scale, mean treatment expectancy was 5.8 (SD 2.3) for participants in the active group and 6.3 (SD 2.1) for participants in the sham group. In the active group, twelve participants (48%) guessed their condition as 'active'; in the sham group sixteen participants (64%) guessed their treatment condition as 'active', indicating adequate concealment of the treatment condition. These differences between the two treatment conditions were not statistically significant.

eTable 5.1 shows the treatment history of the current episode. Most participants had used an SSRI and an SNRI, with no significant differences between both treatment groups. More participants in the sham condition had received ECT, but numbers

Table 5.2: Change in HAMD-17 scores over time, overall and per stratum.

	Overall[a]	
	Active	Sham
	Mean (SD)	Mean (SD)
HAMD week 0	22 (3.2)	22 (2.5)
HAMD week 5	16 (5.4)	17 (5.4)
HAMD week 10	15 (6.1)	14 (6.6)
HAMD week 20	15 (6.6)	13 (6.2)

	Stratum 1[b]		Stratum 2[c]	
	Active	Sham	Active	Sham
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
HAMD week 0	21 (3.0)	20 (2.1)	21 (2.3)	22 (2.1)
HAMD week 5	13 (3.7)	13 (5.0)	16 (4.4)	17 (4.7)
HAMD week 10	9 (8.2)	13 (7.6)	15 (4.6)	14 (7.0)
HAMD week 20	14 (10.0)	9 (4.0)	14 (6.0)	13 (5.0)

	Stratum 3		Stratum 4[e]	
	Active	Sham	Active	Sham
	Score[d]	Score[d]	Mean (SD)	Mean (SD)
HAMD week 0	28	26	27 (2.6)	26 (0.0)
HAMD week 5	24	30	17 (10.4)	17 (0.0)
HAMD week 10	25	18	20 (3.6)	13 (2.8)
HAMD week 20	21	23	20 (2.5)	19 (1.4)

[a] time*group interaction: F(3;66) 0.933; p = .338.

[b] time*group interaction: F(3;22) 1.069; p = .383.

[c] time*group interaction: F(3;99) 0.612; p = .609.

[d] n = 1.

[e] time*group interaction: F(3;13) 0.535; p = .667.

Abbreviations: HAMD = Hamilton-17.

were low and this difference was statistically not significant. Almost all had received a psychotherapeutic intervention, with no differences between groups. With regard to comorbidity it is clear that the most prevalent comorbid disorders were anxiety disorders. Five participants in both groups had comorbid personality disorders (DSM IV cluster C or not otherwise specified).

3.2. Effects of treatment

Mean severity at baseline was a HAMD-17-score of 22 for both groups (table 5.2). In general, participants did improve significantly over time (F 14.768; $p < .001$) (figure 5.2), but showed no difference between intervention and control group; the interaction time*group was not significant (F 0.933; $p = .338$). Correction for duration, type of episode (single or recurrent), number of episodes, treatment expectancy, and treatment guess did not affect these outcomes (eTable 5.2). The number of participants who responded (50% improvement of HAMD-17) was similar for those receiving active (3 (10%)) and sham (2 (8%)) treatment. Remission numbers were similar as well (active: 1 (3%) and sham: 1 (4%)).

In addition, there was a difference in stratum 3 between active and sham treatment of 8 points where the tPEMF participants improved and the sham participants worsened

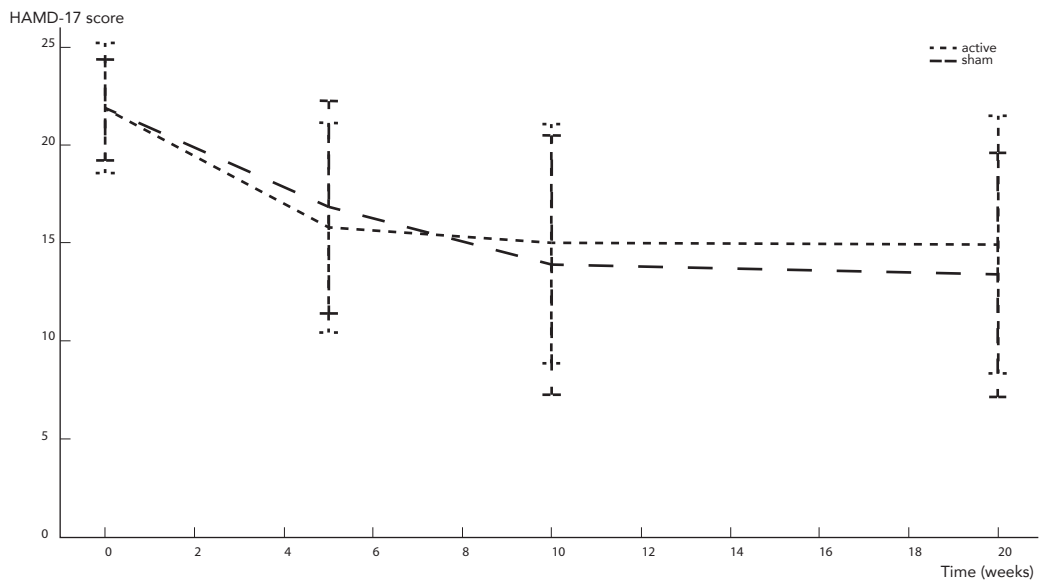


Figure 5.2: Decrease of Hamilton-17 scores over time for both groups.

Table 5.3: Change in IDS-scores over time.

	Active	Sham
	Mean (SD)	Mean (SD)
IDS week 0	44 (9.8)	45 (8.9)
IDS week 1	41 (9.7)	41 (11.2)
IDS week 2	41 (10.4)	38 (11.1)
IDS week 3	40 (10.5)	38 (12.8)
IDS week 4	37 (11.4)	36 (13.4)
IDS week 5	38 (13.3)	38 (12.4)
IDS week 10	34 (13.3)	35 (13.2)
IDS week 20	38 (14.9)	33 (15.1)

time*group interaction: F(7;340) 0.683; p = .687.

Abbreviations: IDS = Inventory of Depressive Symptomatology Self-Report.

on the HAMD-17. However, there were only two participants in this stratum (table 5.2). In the other strata the difference between active and sham treatment was minimal and non-significant. Also, no clear differences were observed between participants who used antidepressants at baseline versus participants who did not use antidepressant medication at baseline (eTable 5.3).

3.3. Secondary outcome measures

With regard to our secondary outcome measures, we found improvement of IDS-SR scores over time (F 10.002; p < .001), but no difference between the two treatment conditions directly post-treatment, and no interaction of time by group (F 0.683; p = .687) (table 5.3). The interaction time*group for the difference in BAI-score of participants was not statistical significant (F: 2.363; p = .055) (table 5.4).

3.4. Reported adverse effects

eTable 5.4 shows reported adverse effects of treatment. A total of 22 participants (40%) reported adverse events with no differences between both groups. Experience of headaches was mostly mentioned. Of those who dropped out of the study, two participants experienced headaches. They were equally distributed over both treatment conditions. None of the adverse events were cause of concern for the participant or reasons to seek medical attention.

Table 5.4: Decrease of BAI-scores over time.

	Active	Sham
	Mean (SD)	Mean (SD)
BAI week 0	22 (10.3)	28 (12.4)
BAI week 2	20 (10.6)	21 (11.4)
BAI week 5	16 (9.7)	23 (11.8)
BAI week 10	14 (9.7)	19 (11.7)
BAI week 20	18 (9.5)	18 (10.9)

time*group interaction: F(4;192) 2.362; p = .055

Abbreviations: BAI = Beck Anxiety Index.

4. Discussion

In this study we aimed to replicate the study of Martiny and colleagues (Martiny, Lunde, Bech 2010) that has shown promising results of tPEMF for patients with TRD. Using a similar design and power calculation we observed an improvement of depression severity over time that continued for fifteen weeks after the last stimulation. However, we found no differences in improvement between the active treatment group and the sham group and were not able to replicate the earlier findings (Martiny, Lunde, Bech 2010).

Several clinical variables related to treatment resistance might have influenced our results. The degree of treatment resistance in our sample was measured with the Maudsley Staging Method (MSM) (Fekadu, Wooderson et al. 2009b). Participants in our sample suffered from moderate treatment resistance (MSM-score: 8), with higher scores indicating a worse depression outcome (Fekadu, Wooderson et al. 2009a; Fekadu, Wooderson et al. 2009b; van Belkum et al. 2018). To explore whether different levels of treatment resistance may have played a role, we stratified our treatment group into four different strata, based on severity and duration of illness, which are known TRD determinants. Sub-analyses of the strata again revealed no differences in improvement of HAMD-17 between treatment conditions. Adding duration and number of antidepressants used as covariates to our analyses also did not substantially affect outcome. Furthermore, on these clinical parameters no clear difference exists between our sample and the previous sample of Martiny (Martiny, Lunde, Bech 2010). Thus, our negative findings are not likely to be due to the distribution of clinical factors that contribute to treatment resistance.

There is one aspect of this study that differs from the original study by Martiny et al. (Martiny, Lunde, Bech 2010) and that might explain the differences in findings, which is difference between the stimulation caps used in both studies. In the current study, we applied 7 V over a cap with 19 small iron core coils positioned according to the international 10/20 system (Myslobodsky et al. 1990). The induced magnetic flux density of this cap is inhomogeneous with a relative stronger magnetic field directly under the coils and a weaker field between coils. In contrast, Martiny et al. used a set of seven air coils, placed on the anterior and posterior temporal regions, the upper parietal regions, and the center of the lower occipital region (Martiny, Lunde, Bech 2010). The relevance of the difference between these treatment-caps is unclear, as the stimulation parameters are the same and inhomogeneous electromagnetic fields have a biological effect (Grossman et al. 2017). However, the difference between the treatment-caps with regard to the precise localization of the coils and the supposed aim could still be of importance. In our set-up, one of the areas covered by the

electromagnets was the frontal lobe, an area often targeted with neurostimulation in treating depression (Brunoni et al. 2016; Lepping et al. 2014). This could be considered to be an advantage of our treatment-cap over the cap of Martiny et al. (Martiny, Lunde, Bech 2010) but the results clearly did not show this. Based on our rough estimation, it could also have been possible that Martiny et al. did in fact influence the local field potentials in the anterior cingulate cortex (ACC), instead of a more global stimulation (Martiny, Lunde, Bech 2010). The ACC also plays an important role in affect in general and depression in particular (Groenewold et al. 2013; Warren, Pringle, Harmer 2015), but non-invasive neurostimulation of the ACC is often difficult due to the depth of this area, even more so when low strength magnetic fields are used. The significance of this difference between caps is thus still unclear and it is questionable if this difference could explain the dissimilarity in findings between our two studies.

In finding an explanation for the difference in findings between these two studies, there is a possibility that the effect-size of tPEMF treatment is much lower than initially thought and therefore our study may have lacked power. However, we calculated our sample size based on the effect-sizes of the previous study (Martiny, Lunde, Bech 2010) and in line with pilot studies of rTMS and tDCS (Lefaucheur et al. 2014; Lefaucheur et al. 2017). Furthermore, the effect-sizes we found were negligible, which would limit the clinical relevance of this stimulation method if replicated in larger samples.

A clear effect in our study was that the average person improved over the course of the study and that this effect lasted for at least fifteen weeks after the last treatment. Placebo effects can be of considerable magnitude in the treatment of depression. For example, effect-sizes for treatment with citalopram are higher when MDD-patients are being treated in an open label study compared to an RCT, even if the same treatment regimen is used (Rutherford et al. 2017). Martiny et al. have reported that their active stimulator did emit a faint humming sound (Martiny, Lunde, Bech 2010). Participants could thus have been partially aware of the treatment condition they were in. However, no clear evidence was found in the study of Martiny et al. that participants did actually hear the faint humming noise, as reflected by the amount of people that correctly guessed their treatment condition which was no better than chance in both groups (Martiny, Lunde, Bech 2010). This also was similar to our study, in which participants were not able to guess in what condition they were.

In line with the finding of Martiny et al., other studies with different stimulation parameters have also reported on global neuromodulation techniques (Leuchter et al. 2015; Rohan et al. 2013; Straaso et al. 2014). For example, a dose-remission study without a sham condition has found that augmentation with tPEMF stimulation in patients with TRD during eight weeks reduced HAMD-17 scores with 74% and 68%

(13 and 14 points) if treated with one vs. two daily tPEMF doses (Straaso et al. 2014). Another study, a double blind sham controlled RCT, applying Low Field Magnetic Stimulation (LFMS) to stimulate the whole cortex with oscillating electromagnetic fields in 63 depressed participants, has found that this had an immediate positive effect on depression severity (Rohan et al. 2013) . Another device with three rotating magnets has shown antidepressive effects in a double blind sham controlled RCT in 202 depressed participants (Leuchter et al. 2015). These studies point in the direction of an antidepressive effect of global neuromodulation devices, in support of the study by Martiny et al. (Martiny, Lunde, Bech 2010). However, the later studies all had methodological caveats: they either lacked a sham condition (Straaso et al. 2014) used a rating-scale not validated for measuring short-term change (Rohan et al. 2013), or found no difference on the primary outcome despite reporting some antidepressive effect (Leuchter et al. 2015).

To summarize, although there were minor differences in sample and set-up between the studies, we were not able to replicate the promising findings of an earlier tPEMF study using a similar design. For the time being the conclusion must therefore be that transcranial pulsed electromagnetic fields do not have consistent antidepressive effects. Alternatively, moderator variables that either enhance or preclude such effects may be at work and need to be identified. More studies will be needed for a more definite answer to the question whether tPEMF can be of clinical value in the treatment of TRD. In addition, studies into putative neurobiological mechanisms are needed to clarify biological plausibility for clinical effects to occur.

5. Acknowledgements

This study was funded by: UMCG Innovation Fund, project U-11-221, and Fonds NutsOhra, project 1103-068 (main applicant on both applications was RS).

The authors thank Charlotte Kohne, Chris Geraets, Christien van Buuren, Ella Bekhuis, Elroy Doornbos, Esther van Veen, Joyce van Meel, Lotte Staas, Lydia Datema, Maartje Bastiaans, Magda Tasma, Milou van Eldik, Nadine de Jong, Nina Schimmel, Philip Nan, Roelien Anna Vaals, Sjoukje Vroom, Stella Druiven, and Thom Lysen for their assistance in conducting this study.

6. Funding and Disclosure

RK is owner of Magnolia Therapeutics, a company that develops and sells magnetic stimulators and that offers magnetic stimulation and counseling directly to the public.

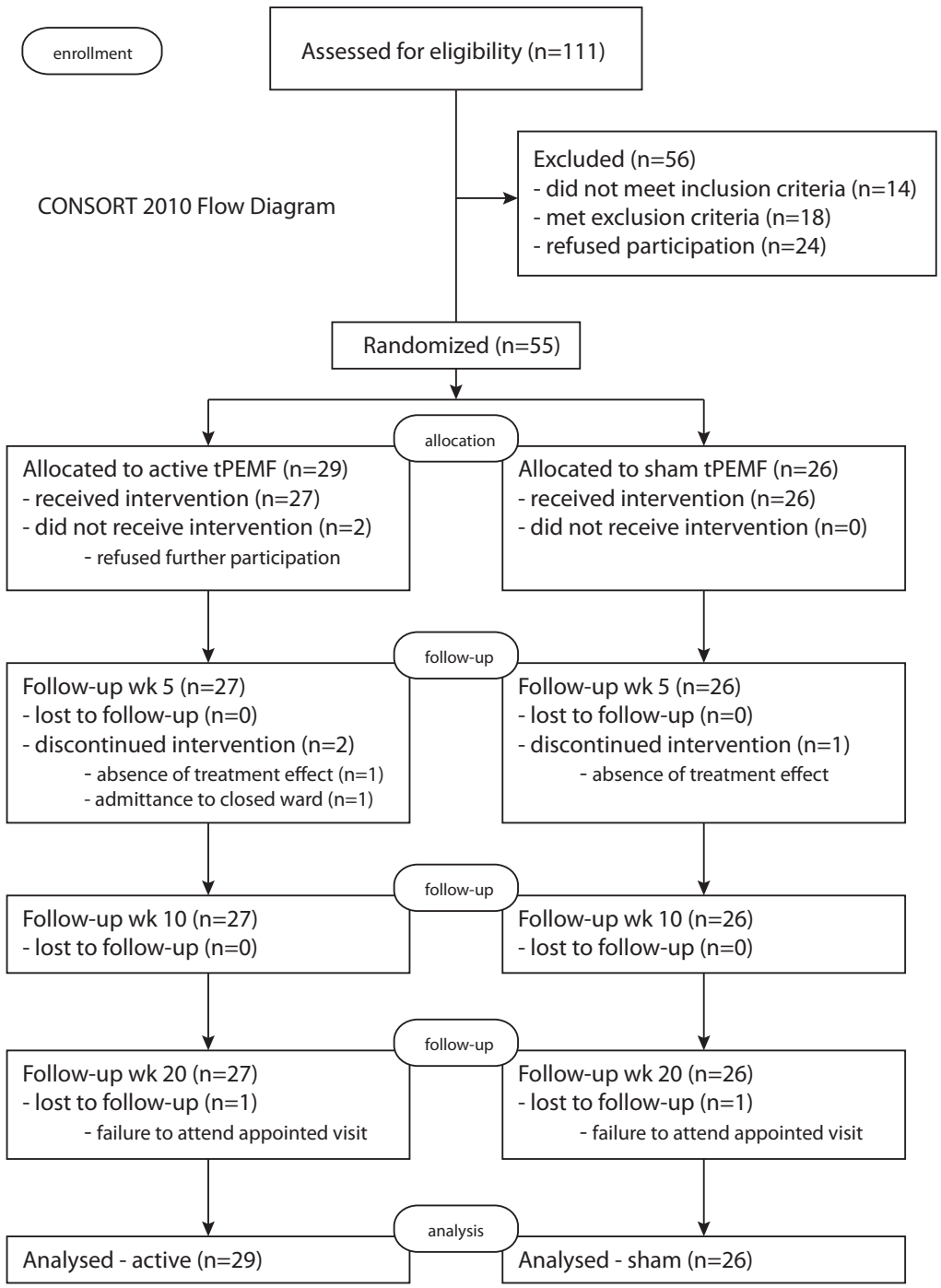
No conflicts of interests for SvB; MdB; EO; FW; RH; HK; AA; RS.

Supplementary Methods

Changes in protocol

The trial started out as a single center trial. Due to a low recruitment rate, we changed the study to a multicenter trial. Furthermore we changed our eligibility criteria. In the first protocol, inclusion criteria stated that participants should not have responded to two or more antidepressants. We changed this to not having responded to one or more antidepressant, as this was more in line with the method of Martiny et al. (2010). Furthermore, in the first protocol, mood stabilizers and antipsychotics were not allowed in the last four weeks before or during the course of the study. We changed this to the allowance of these medications during the whole period. Blinding was not affected by these changes and the Medical Ethical Committee of the UMCG approved all changes.

Supplementary Results



eFigure 5.1: Flow-chart of patient disposition.

eTable 5.1: Treatment history of current episode.

Treatment-history current episode	Active	Sham	p-value
Number of antidepressant used (median (IQR))	2 (1 - 3)	2 (1 - 4)	.390[a]
Antidepressant used in current episode			
TCA (%)	12 (41%)	10 (38%)	
SSRI (%)	20 (69%)	21 (81%)	
SNRI (%)	15 (52%)	15 (58%)	
MAOI (%)	1 (3%)	3 (12%)	
MAOA (%)	1 (3%)	1 (4%)	
Misc (%) (Agomelatine, Bupropion, Mirtazapine, Trazodon)	7 (24%)	12 (46%)	
Other psychopharmacology used in current episode			
Benzodiazepines (%)	5 (17%)	8 (31%)	
Antipsychotics (%)	6 (21%)	4 (15%)	
Lithium (%)	5 (17%)	5 (19%)	
Antiepileptics (%)	3 (10%)	3 (12%)	
Psychotherapeutic treatment in current episode			.926[b]
None (%)	1 (3%)	1 (4%)	
Supportive (%)	12 (41%)	13 (50%)	
1 protocolized (%)	13 (45%)	10 (38%)	
2+ protocolized (%)	3 (10%)	2 (8%)	
ECT (% yes)	0 (0%)	3 (12%)	.060[b]
Intensified treatment of current episode			.270[b]
None (%)	18 (62%)	18 (69%)	
Day-care <12 weeks or < 3days/week	0 (0%)	0 (0%)	
Day-care >12 weeks or > 3days/week	8 (28%)	3 (12%)	
Clinical admission	3 (10%)	5 (19%)	

[a] Mann-Whitney U

[b] Chi-square

eTable 5.2: Correction for duration of episode, DSM-code, number of episodes, treatment expectancy, and treatment guess did not substantially affect the primary outcome, both univariate and multivariate.

UNIVARIATE	Numerator df	Denominator df	F	p-value
Duration	1	48.356	0.795	.377
time*group[a]	1	48.356	0.795	.377
Number of episodes	1	36.371	0.052	.820
time*group[b]	3	14.852	0.937	.424
Single or recurrent episode	1	49.704	4.330	.070
time*group[c]	3	148.513	0.929	.428
Treatment guess	1	50.961	2.183	.146
time*group[d]	3	14.865	0.935	.425
Treatment expectancy	1	48.027	0.439	.511
time*group[e]	3	14.550	1.107	.348
MULTIVARIATE[f]	Numerator df	Denominator df	F	p-value
Time	3	143.975	26.213	< .001
Baseline HAMD-score	1	33.200	39.622	< .001
Duration	1	42.674	0.432	.515
Number of episodes	1	32.835	2.538	.121
Single or recurrent episode	1	46.554	4.782	.034
Treatment guess	1	42.250	1.151	.289
Treatment expectancy	1	43.099	0.182	.672
Group	1	41.735	0.798	.377
time*group	3	145.199	1.098	.352

[a] Adjusted for duration.

[b] Adjusted for number of episodes.

[c] Adjusted for single or recurrent episode.

[d] Adjusted for treatment guess.

[e] Adjusted for treatment expectancy.

[f] Dependent Variable: Hamilton-score.

Abbreviations: df = degrees of freedom.

eTable 5.3: Change in HAMD-17 scores divided based on use of antidepressant.

	No AD use on baseline[a]		AD use on baseline[b]	
	Active	Sham	Active	Sham
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
HAMD week 0	22 (2.9)	22 (2.9)	22 (3.5)	22 (2.4)
HAMD week 5	14 (4.9)	16 (4.5)	17 (5.5)	17 (5.9)
HAMD week 10	15 (5.7)	12 (5.5)	15 (6.6)	15 (7.1)
HAMD week 20	13 (6.1)	13 (8.4)	17 (6.5)	13 (4.8)

[a] time*group interaction: F(3;55) 1.895; p = .141.

[b] time*group interaction: F(3;88) 1.049; p = .375.

Abbreviations: AD = antidepressant, HAMD = Hamilton-17.

eTable 5.4: Adverse events present across both treatment groups.

Study related adverse events	Active	Sham	p-value
Adverse Events present	10 (34%)	12 (46%)	.378
Adverse Events present in dropouts	1 (3%)	1 (4%)	.171
Presence of headache	8 (28%)	9 (35%)	.573
Presence of sleep disturbances	5 (17%)	2 (8%)	.289
Presence of concentration disturbances	1 (3%)	0 (0%)	.339
Presence of tingling sensations	2 (7%)	1 (4%)	.619
Presence of tension	1 (3%)	1 (4%)	.937
Presence of fatigue	1 (3%)	0 (0%)	.339
Presence of nausea	1 (3%)	1 (4%)	.937