Blinding integrity in randomized sham-controlled trials of repetitive transcranial magnetic stimulation for major depression: a systematic review and meta-analysis

Marcelo T. Berlim1,2, Hannah J. Broadbent3 and Frederique Van den Eynde1

1 Neuromodulation Research Clinic, Douglas Mental Health University Institute and McGill University, Montréal, Québec, Canada
2 Depressive Disorders Program, Douglas Mental Health University Institute and McGill University, Montréal, Québec, Canada
3 Psychology and Human Development, Institute of Education, University of London, London, UK

Abstract

Repetitive transcranial magnetic stimulation (rTMS) is a safe and effective treatment for major depression (MD). However, the perceived lack of a suitable sham rTMS condition might have compromised the success of blinding procedures in clinical trials. Thus, we conducted a systematic review and meta-analysis of randomized, double-blind and sham-controlled trials (RCTs) on high frequency (HF-), low frequency (LF-) and bilateral rTMS for MD. We searched the literature from January 1995 to July 2012 using Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials and Scopus. The main outcome measure was participants’ ability to correctly guess their treatment allocation at study end. We used a random-effects model and risk difference (RD).

Overall, data were obtained from seven and two RCTs on HF- and bilateral rTMS, respectively. No RCT on LF-rTMS reporting on blinding success was found. HF- and bilateral rTMS trials enrolled 396 and 93 depressed subjects and offered an average of approximately 13 sessions. At study end, 52 and 59% of subjects receiving HF-rTMS and sham rTMS were able to correctly guess their treatment allocation, a non-significant difference (RD = 0.04; z = 0.51; p = 0.61). Furthermore, 63.3 and 57.5% of subjects receiving bilateral and sham rTMS were able to correctly guess their treatment allocation, also a non-significant difference (RD = 0.05; z = 0.49; p = 0.62). In addition, the use of angulation and sham coil in HF-rTMS trials produced similar results. In summary, existing sham rTMS interventions appear to result in acceptable levels of blinding regarding treatment allocation.

Introduction

Repetitive transcranial magnetic stimulation (rTMS) is a non-invasive technique for modulating cortical and subcortical function that is being increasingly investigated as a novel treatment for several neuropsychiatric disorders (George et al., 2009) and particularly for major depression (MD; Daskalakis et al., 2008). rTMS treatment for MD usually involves one of three protocols: high frequency-rTMS (HF-rTMS) applied to the left dorsolateral prefrontal cortex (DLPFC); low frequency-rTMS (LF-rTMS) applied over the right DLPFC; bilateral HF-/LF-rTMS (George and Aston-Jones, 2010; Wassermann and Zimmermann, 2012). However, questions still remain as to whether the sham strategies used in these trials were effective in maintaining blinding integrity (Loo and Mitchell, 2005; Daskalakis et al., 2008; Rossi et al., 2009), particularly because inadequate allocation concealment may lead to exaggerated estimates of treatment effects (Schulz et al., 1995; Gluud, 2006; Wood et al., 2008).

Real and sham rTMS ideally have to be different in terms of their brain effects and as similar as possible concerning the subjective experience associated with brain stimulation. Randomized, double-blind and sham-controlled trials (RCTs) on rTMS to date have employed one of two strategies as a control/sham intervention: the angulation of a real magnetic coil at 45° or 90° over the scalp (which produces the associated rTMS sound and part of the subsidiary muscular twitching); the use of a ‘sham coil’ that is visually and auditorily identical to the real rTMS coil but that does not generate a magnetic field or an underlying muscular activity. For additional information on sham rTMS methods please see Rossi et al. (2009) and Sandrini et al. (2011).
Several authors have reported that these two sham rTMS strategies often fail to fully replicate the local cutaneous sensation associated with the use of real rTMS (Brunoni and Fregni, 2011; Rosa and Lisanby, 2012). Furthermore, the use of coil angulation has been shown to produce some cortical stimulation (Loo et al., 2000), particularly in 45° arrangements (Lisanby et al., 2001). Thus, the potential lack of a reliable sham/placebo method might have compromised the success of blinding in RCTs of rTMS for MD, and thus limit the estimation of its antidepressant properties (Daskalakis et al., 2008).

A recent systematic review on the blinding success of rTMS applied to the DLPFC has shown that participants in real and sham rTMS groups were not significantly different in their ability to correctly guess their intervention allocation (Broadbent et al., 2011). However, the quantitative analyses employed (using separate logistic regressions) did not weigh RCTs based on their overall precision, i.e. studies that yielded more precise estimates of the effect size (e.g. those with larger samples) should be assigned more weight in the statistical analyses (Borenstein et al., 2009; Huf et al., 2011).

Therefore, to summarize the best available evidence on the success of blinding strategies in RCTs of rTMS for treating MD, we carried out the present systematic review and meta-analysis. The main outcome measure was the success of blinding as indexed by participants’ ability to correctly guess their treatment allocation (i.e. active vs. sham rTMS) at study end. Also, where possible, we conducted sensitivity analyses to assess the impact of alternative stimulation parameters on blinding integrity. Finally, we investigated whether response to treatment following rTMS was associated with overall rates of blinding integrity.

Methodology of the literature review

Search strategy

We identified articles for inclusion in this meta-analysis by two complementary strategies: (1) screening the bibliography of the previous meta-analyses on rTMS for MD published to date (McNamara et al., 2001; Burt et al., 2002; Kozel and George, 2002; Martin et al., 2002, 2003; Couturier, 2005; Herrmann and Ebmeier, 2006; Gross et al., 2007; Lam et al., 2008; Schutter, 2010; Slotema et al., 2010; Allan et al., 2011) and the systematic review by Broadbent and colleagues (Broadbent et al., 2011) and also of all included RCTs; (2) searching Medline, EMBASE, PsycINFO, the Cochrane Central Register of Controlled Trials (CENTRAL), Scopus and ProQuest Dissertations & Theses (PQDT) from 1 January 1995 to 20 July 2012.

The search procedures (including syntaxes, parameters and results) are described in detail in the Supplementary material.

Study selection

Candidate studies (judged on the basis of their title and abstract) had to satisfy the following criteria (Higgins and Green, 2008).

Study validity: random allocation; double-blind (i.e. patients and clinical raters blinded to treatment conditions); sham-controlled (i.e. coil angled on the scalp or use of a specific sham coil); parallel or crossover design (with only data from the initial randomization being used for the latter to avoid carry-over effects); ≥5 subjects with MD randomized per study arm.

Sample characteristics: subjects aged 18–75 yr with a diagnosis of primary major depressive episode (unipolar or bipolar) according to DSM-IV (APA, 1994) or the International Classification of Diseases criteria (WHO, 1992).

Treatment characteristics: HF-rTMS (i.e. ≥5 Hz), LF-rTMS (i.e. ≤1 Hz) or bilateral HF-/LF-rTMS (either simultaneously or sequentially) given for ≥5 sessions.

Publication-related: articles written in English.

Studies were excluded if they: enrolled subjects with ‘narrow’ diagnoses (e.g. post-partum depression) or secondary MD (e.g. vascular depression); did not report on blinding integrity at blinded study end.

Data extraction

Data were recorded in a structured fashion as follows. Sample characteristics: mean age; gender; primary diagnosis; presence of treatment-resistant depression. rTMS-related: stimulation frequency and intensity (including the total number of stimuli delivered); number of treatment sessions; type of sham. Integrity of blinding: number of subjects in the active and sham rTMS groups who were able to correctly guess their treatment allocation at study end.

Data synthesis and analyses

Analyses were performed using Comprehensive Meta-Analyses version 2.0 (Biostat, USA) and IBM SPSS version 20 (IBM Corporation, USA).

We used a random-effects model because we assumed that the true treatment effects had likely varied between the included RCTs (Riley et al., 2011). Blinding integrity was investigated with risk difference (RD). Also, to rule out the presence of baseline differences in depression severity between active and sham rTMS groups, we computed the pooled standardized mean differences (s.m.d.) for subjects’ baseline depression scores. Additionally, we performed cumulative analyses to retrospectively identify the time-point when HF-rTMS (compared to sham rTMS) first reached conventional levels of statistical non-significance in terms of blinding integrity (Egger et al., 2001). Moreover, we conducted sensitivity analyses (Egger et al., 2001) to assess the potential impact of the number of sessions (10 vs. ≥10)
and the total number of magnetic pulses delivered (i.e. ≤25,000 vs. >25,000) on the effect-size estimates for blinding integrity. Finally, to investigate the relationship between clinical improvement following rTMS and blinding integrity, we conducted a linear regression with the RD for participants correctly guessing their treatment allocation at study end as the dependent variable and the RD for response to treatment (i.e. >50% reduction in post-treatment depressive symptoms (Rush et al., 2006); data derived from a previous meta-analysis (M. T. Berlim, F. Van den Eynde, S. T. Perdomo and Z. J. Daskalakis, unpublished observations) as the independent variable.

Heterogeneity was assessed using the Q statistics and the I² index (Cooper et al., 2009). Values of $p < 0.1$ for the former and $>25\%$ for the latter were deemed as indicative of study heterogeneity (Borenstein et al., 2009). Finally, we used funnel plots, Egger’s regression intercept (Egger et al., 1997) and Duval and Tweedie’s trim and fill procedure (Duval and Tweedie, 2000) to test for the presence of publication bias (Borenstein et al., 2009; Cooper et al., 2009).

Results

Literature search

From the RCTs on rTMS for MD included in the previous meta-analyses (McNamara et al., 2001; Burt et al., 2002; Kozel and George, 2002; Martin et al., 2002, 2003; Couturier, 2005; Herrmann and Ebmeier, 2006; Gross et al., 2007; Lam et al., 2008; Schutter, 2010; Slotema et al., 2010; Allan et al., 2011) and in the systematic review by Broadbent et al. (2011), six were selected for the present investigation (Nahas et al., 2003; Avery et al., 2006; Fitzgerald et al., 2006; Loo et al., 2007; Mogg et al., 2008; Herwig et al., 2010). Also, we included two RCTs retrieved from Medline, PsycINFO, EMBASE, CENTRAL, Scopus and PQDT (George et al., 2010; Blumberger et al., 2012). Figure 1 shows a PRISMA flowchart (Moher et al., 2009), and the Supplementary material provides a description of the study selection procedures.

Included RCTs: main characteristics

HF-rTMS

Overall, seven RCTs on HF-rTMS for MD were included in our meta-analysis (Nahas et al., 2003; Avery et al., 2006; Loo et al., 2007; Mogg et al., 2008; George et al., 2010; Herwig et al., 2010; Blumberger et al., 2012), totaling 396 subjects with MD, of whom 196 were randomized to active rTMS (mean age $= 48.5 \pm 4.2$ yr; 58.5% females) and 200 were randomized to sham rTMS (mean age $= 47.2 \pm 3.6$ yr; 54.3% females; Table 1). The mean number of HF-rTMS sessions and total magnetic pulses delivered were $13 \pm 2.7$ and $25,250 \pm 11,308$, respectively. Also, most RCTs used coil angulation (four out of seven).
We found no RCT reporting on the integrity of blinding after LF- vs. sham rTMS for MD.

Bilateral rTMS

Two RCTs on bilateral rTMS for MD were included in our meta-analysis (Fitzgerald et al., 2006; Blumberger et al., 2012), totalling 93 subjects with MD, of whom 49 were randomized to sequential bilateral rTMS (mean age = 52.4 ± 7.9 yr; 56.9% females) and 40 were randomized to sham rTMS (mean age = 44.75 ± 1.48 yr; 66.7% females; Table 2). The mean number of bilateral rTMS sessions and the total number of magnetic pulses delivered in the HF- and LF-rTMS protocols were 12.5 ± 3.5, 9375 ± 2651 and 5587 ± 1952, respectively. Also, both RCTs employed coil angulation.

Integrity of blinding

HF-rTMS

Overall, no difference was observed between active and sham HF-rTMS groups in terms of the number of participants correctly guessing their treatment allocation at study end [52% (102/196) vs. 59% (118/200), respectively; RD = -0.04; 95% confidence interval (CI) -0.2 to 0.12; z = -0.51; p = 0.61; Fig. 2]. Furthermore, no differences in terms of blinding integrity were found between the use of coil angulation and of sham coil (Q = 0.007; d.f. = 1; p = 0.93). More specifically, in RCTs using coil angulation, 52.6% (41/78) vs. 56.7% (42/74) of participants receiving active and sham rTMS correctly guessed their treatment allocation (RD = 0.05; 95% CI -0.29 to 0.2; z = -0.38; p = 0.71), respectively, whereas in those using a sham coil, 51.7% (61/118) vs. 60.3% (76/126) of participants receiving active and sham rTMS correctly guessed their treatment allocation (RD = -0.03; 95% CI -0.28 to 0.22; z = -0.24; p = 0.81), respectively (Fig. 3).

Heterogeneity between RCTs on HF-rTMS for MD exceeded that expected by chance (d.f. = 6; Q = 14.39; p = 0.026; I² = 58.3), implying that the variance among the effect sizes was greater than expected by sampling error. However, sensitivity analyses did not identify any specific study as the main source of heterogeneity. Additionally, the associated funnel plot was reasonably symmetrical (Fig. 4). Publication bias was assessed more conservatively with Egger’s regression intercept, which was 1.55 (d.f. = 5; t = 0.77; two-tailed p = 0.47) and no RCT was trimmed in the Duval and Tweedie’s trim and fill procedure, thus suggesting a low risk of publication bias.

Bilateral rTMS

Overall, no difference was observed between active and sham bilateral rTMS groups in terms of the number of participants correctly guessing their treatment allocation at study end [63.3% (31/49) vs. 57.5% (23/40),
respectively; RD=0.05; 95% CI –0.15 to 0.25; z=0.49; p=0.62; Fig. 5).

Heterogeneity between RCTs on bilateral rTMS for MD did not exceed that expected by chance (d.f. = 1; Q1 = 0.23; p = 0.63; P = 0). As a result of the small number of included RCTs, we could not assess the risk of publication bias.

**Cumulative analyses**

RCTs on MD showed HF-rTMS to be similar to sham rTMS in terms of blinding integrity by the year 2003 (Fig. 6). Further studies essentially narrowed the 95% CI around relatively similar RD estimates.

**Sensitivity analyses**

We were only able to conduct sensitivity analyses on the set of RCTs on HF-rTMS for MD. First, we re-analysed blinding integrity data after excluding the study by George et al. (2010), as they used an electrical sham system that simulates the scalp sensations associated with real rTMS (Borckardt et al., 2008), thus clearly differing from traditional inactive sham coils. Our results show that the exclusion of this RCT did not significantly affect the initial effect-size estimates for the whole sample analysis (adjusted RD = –0.006; z = –0.06; p = 0.95) or for the angulation/sham coil sub-group analysis (Q = 0.28; d.f. = 1; p = 0.6). Also, the number of sessions (10 vs. > 10) and the total number of magnetic pulses delivered (i.e. < 25 000 vs. > 25 000) were not associated with differential estimates of blinding integrity (i.e. Q = 0.004; d.f. = 1; p = 0.95 and Q = 0.7; d.f. = 1; p = 0.4, respectively). The associated Forrest plots are shown in the Supplementary material.

**Response to treatment and blinding integrity**

The RD for response to treatment significantly predicted the RD for participants correctly guessing their treatment allocation at study end (B = 1.36; t = 4.2; p = 0.014), thus suggesting that perceived clinical improvement might influence the latter.

**Active vs. sham rTMS groups: depression severity at baseline**

No baseline difference on depression scores was found between the HF- and sham rTMS groups (S.M.D. = 0.01; p = 0.95) and the bilateral rTMS and sham rTMS groups (S.M.D. = –0.03; p = 0.87), thus ruling out illness severity at baseline as a confounding factor. The associated Forrest plots are shown in the Supplementary material.

**Discussion**

This is the first meta-analysis assessing blinding success in RCTs of rTMS for MD. Our results show that there is no significant difference between active and sham rTMS.
groups in terms of blinding integrity at study end and that real coil angulation and sham coil strategies appear to be equal in this respect (at least in the case of HF-rTMS). Overall, slightly more than half the subjects in each group correctly guessed their treatment allocation (i.e. chance level). Also, statistically non-significant results on blinding integrity between HF- and sham rTMS were first reached in 2003. Finally, our finding that participants presenting with a clinical response at study end were more likely to guess that they had received real rTMS is probably explained by the fact that symptomatic improvement is more often associated with active than sham/placebo interventions (Schulz et al., 1995; Day and Altman, 2000; Hrobjartsson and Boutron, 2011) and we believe that this has not affected the validity of the blinding methodology.

The main results of our meta-analysis are partially in agreement with those of Broadbent and colleagues (Broadbent et al., 2011), although we did not find that subjects who received real rTMS were more likely to correctly guess their treatment allocation. Nevertheless, direct comparisons with this previous systematic review are difficult because we used meta-analytical procedures and focused on a more homogeneous patient population (i.e. depressed subjects) receiving specific rTMS protocols (i.e. HF- and bilateral rTMS).
A striking issue is that data on blinding integrity in RCTs of rTMS for MD are very rarely reported and it remains unclear whether this is a result of reporting bias and/or a lack of standardized assessments in primary studies. Thus, future RCTs on rTMS should systematically assess and report information on blinding integrity by using, for example, a two-choice ‘real/sham rTMS’ question and a visual analogue scale to gather subjects’ level of certainty about allocation (with ‘sham rTMS’ at one end and ‘real rTMS’ at the other; Broadbent et al., 2011). Moreover, in RCTs offering multiple rTMS sessions, we suggest that participants should be asked to guess their treatment allocation not only at study end, but also after the first few rTMS sessions in order to investigate whether their choice is dependent, for example, on eventual response to treatment.

Another important methodological issue is the difficulty in blinding researchers who deliver rTMS, although recent evidence suggests that this may be accomplished with the use of a novel sham system (Borckardt et al., 2008; George et al., 2010). Interestingly, however, our sensitivity analysis indicated a lack of difference in blinding success between real coil angulation and sham coil strategies and this may suggest that only participants and raters should be blinded to treatment allocation in rTMS trials. Nevertheless, this is a preliminary finding that should be explored in future studies.

In summary, the identification of a valid sham condition for rTMS research is of key importance for producing valid results (George and Aston-Jones, 2010). Considering that participants may have detected that they had received an atypical stimulation procedure because of previous knowledge on rTMS or because of a lack of sensory stimuli during the sham intervention, the probability of correctly guessing treatment allocation should have been high. However, our findings suggest this was not case and that current sham strategies can be used reliably in RCTs of rTMS for MD.

**Limitations**

First, the included RCTs enrolled a relatively small number of depressed subjects. Second, although blinding assessment can be useful to identify whether the blinding was broken (Boutron et al., 2007), this method is probably not ideal as, for example, correct guessing might be related to side-effects and/or amelioration of symptoms rather than to the blinding process itself (Fergusson et al., 2004). Third, we found significant heterogeneity among the included RCTs on HF-rTMS for MD and this might have influenced our effect-size estimates. However, it is difficult to determine the precise nature of this heterogeneity as it may be due, for example, to genuine differences in participants, interventions, co-interventions, outcomes, measurements, settings and to numerous other factors varying across data sets, studies and participants (Thompson, 1994; Petitti, 2001; Ioannidis, 2008). Nevertheless, careful inspection of the included RCTs did not reveal any major differences in their overall methodology. Finally, meta-analyses have been often criticized for the potential of publication bias and for the inclusion of poor-quality trials (Borenstein et al., 2009). In the present study, however, these concerns were addressed by the comprehensive systematic review of the literature and the use of stringent inclusion criteria and by the objective examination of study heterogeneity and publication bias.
Conclusion
Our systematic review and meta-analysis shows that information on blinding integrity in typical RCTs of rTMS for MD is rarely reported and reasons for this phenomenon remain unclear. Furthermore, our findings suggest that commonly used sham rTMS methods appear to adequately conceal treatment allocation, leading to acceptable levels of blinding integrity at study end. Nevertheless, the development of novel and optimized sham rTMS strategies such as electromagnetic sham (Rossi et al., 2007), focal electrical stimulation (Arana et al., 2008) or shielded magnetic coils (Sommer et al., 2006) is still a worthwhile endeavour.

Declaration of Interest
None.

Supplementary material
For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1461145712001691.

References


