Efficacy and Safety of Transcranial Magnetic Stimulation in the Acute Treatment of Major Depression: A Multisite Randomized Controlled Trial

John P. O’Reardon, H. Brent Svolason, Philip G. Janicak, Shirlene Sampson, Keith E. Isenber, Ziad Nahas, William M. McDonald, David Avery, Paul B. Fitzgerald, Colleen Loo, Mark A. Demitrack, Mark S. George, and Harold A. Sackeim

Background: We tested whether transcranial magnetic stimulation (TMS) over the left dorsolateral prefrontal cortex (DLPFC) is effective and safe in the acute treatment of major depression.

Methods: In a double-blind, multisite study, 301 medication-free patients with major depression who had not benefited from prior treatment were randomized to active (n = 155) or sham TMS (n = 146) conditions. Sessions were conducted five times per week with TMS at 10 pulses/sec, 120% of motor threshold, 3000 pulses/session, for 4 – 6 weeks. Primary outcome was the symptom score change as assessed at week 4 with the Montgomery–Asberg Depression Rating Scale (MADRS). Secondary outcomes included changes on the 17- and 24-item Hamilton Depression Rating Scale (HAM-D) and response and remission rates with the MADRS and HAMD.

Results: Active TMS was significantly superior to sham TMS on the MADRS at week 4 (with a post hoc correction for inequality in symptom severity between groups at baseline), as well as on the HAMD17 and HAMD24 scales at weeks 4 and 6. Response rates were significantly higher with active TMS on all three scales at weeks 4 and 6. Remission rates were approximately twofold higher with active TMS at week 6 and significant on the MADRS and HAMD24 scales (but not the HAMD17 scale). Active TMS was well tolerated with a low dropout rate for adverse events (4.5%) that were generally mild and limited to transient scalp discomfort or pain.

Conclusions: Transcranial magnetic stimulation was effective in treating major depression with minimal side effects reported. It offers clinicians a novel alternative for the treatment of this disorder.

Key Words: Clinical trial, efficacy, major depression, safety, TMS

Maj or depressive disorder (MDD) is a common, recurrent, and frequently chronic disorder that is a leading contributor to functional impairment and disability (Murray and Lopez 1996). Treatment is often challenging; an estimated 20%–40% of patients do not benefit sufficiently from existing antidepressant interventions including trials of medication and psychotherapy (Greden 2001). A substantial proportion of patients manifest a chronic, treatment-resistant course of illness, resulting in a need for additional treatment options (Rush et al. 2006; Trivedi et al. 2006). Transcranial magnetic stimulation (TMS) has been proposed as one such alternative (George et al. 1997, 1995b; George and Wassermann 1994).

During TMS, a time-varying current is discharged in an insulated coil placed on the scalp surface. This generates a brief dynamic magnetic field that is orthogonal in orientation to current flow in the coil (Amassian et al. 1992; Roth et al. 1991a, 1991b). The scalp and skull are transparent to the magnetic field, which induces current flow when it reaches a conductive medium such as neural tissue and with it the potential to modulate neural circuitry in a therapeutic fashion.

Several single-center, controlled studies of TMS have been conducted that, in most cases, have supported the hypothesis that TMS manifests antidepressant properties when delivered to the left or right dorsolateral prefrontal cortex (DLPFC; Avery et al. 2005; Fitzgerald et al. 2003; George et al. 1997, 2000; Klein et al. 1999; Loo et al. 1999, 2001; Pascual-Leone et al. 1996). This target was initially selected (George et al. 1995b) based on imaging findings implicating this region in the pathophysiology of depression and in antidepressant effects, as well as studies linking specific lesion locations to dysregulation of mood (Bench et al. 1992, 1995; Brody et al. 2001; George et al. 1995a, 1993a, 1993b; Kimbrell et al. 2002; Mayberg 2002, 2003; Nobler et al. 2001; Sackeim 2001a; Sackeim 2001b; Robinson et al. 1984, 1988).

Meta-analyses conducted of this literature have largely concurred that either a slow rate of stimulation (≤ 1 pulse/sec) over the right DLPFC or fast stimulation (at 5–20 pulses/sec) over the left DLPFC have greater antidepressant effects than matched sham stimulation conditions (Burt et al. 2002; Couterier 2005; Kozel and George 2002; Martin et al. 2002; McNamara et al. 2001). Such effects, however, have been questioned in terms of clinical significance, with at least one meta-analysis, using con-

From the Department of Psychiatry (JPO), University of Pennsylvania, Philadelphia; Department of Psychiatry (HBS), Stanford University, Palo Alto, California; Department of Psychiatry (PGJ), Rush University Medical Center, Chicago, Illinois; Mayo Clinic College of Medicine (SS), Rochester, Minnesota; Department of Psychiatry (KEI), Washington University School of Medicine, St. Louis, Missouri; Department of Psychiatry and Behavioral Sciences (WMM), Emory University, Atlanta Georgia; Department of Psychiatry and Behavioral Sciences (DA), Seattle, Washington; University of Washington School of Medicine; Alfred Psychiatry Research Centre (PBF), Melbourne, Australia; the Alfred and Monash University School of Psychology, Psychiatry and Psychological Medicine; School of Psychiatry (CL), University of NSW, Sydney, Australia; Department of Psychiatry (ZN, MSG), Medical University of South Carolina, Charleston, South Carolina; Neuronetics, Inc. (MAD), Malvern, Pennsylvania; Department of Biological Psychiatry (HAS), New York State Psychiatric Institute and College of Physicians and Surgeons, Columbia University, New York. Address reprint requests to John P. O’Reardon, M.D., Laboratory for Transcranial Magnetic Stimulation, University of Pennsylvania, 3535 Market Street, Suite 4005, Philadelphia, PA 19104.

Received September 29, 2006; revised November 25, 2006; accepted January 19, 2007.

0006-3223/07/$32.00 doi:10.1016/j.biopsych.2007.01.018

sorptive criteria, finding no clear benefit from active TMS (Martin et al. 2002). All meta-analyses have acknowledged the limitations of prior work including insufficient intensity and duration of treatment, inferior sham condition, single-center study design, and inadequate sample size.

We conducted a large, multicenter, randomized controlled trial of TMS to address key prior shortcomings. We administered TMS at fixed treatment parameters over the left DLPFC and compared it with a sham TMS intervention that mimicked the active condition to a greater extent than has been the case in previous research. The TMS parameters in the study were selected to achieve the highest feasible dose consistent with current safety guidelines (Wassermann 1998) and patient tolerability. The duration of the TMS course (4–6 weeks) was also longer than in previous studies, providing greater opportunity for cumulative therapeutic and adverse effects to be expressed.

Methods and Materials

Subjects

Eligible subjects were antidepressant medication-free outpatients, aged 18–70, with a DSM-IV diagnosis of MDD, single episode or recurrent, with a current episode duration of 3 years or less. The episode had a Clinical Global Impressions Severity of Illness (CGI-S) score of at least 4 and a total score of at least 20 on the 17-item Hamilton Depression Rating Scale (HAMD17). Symptom stability was required during a 1-week no-treatment lead-in period, with a HAMD17 total score of at least 18 and a decrease in score of 25% or less from that observed at the screening assessment. Prior antidepressant treatment during the current episode was assessed using the Antidepressant Treatment History Form (ATHF; Sackeim 2001b). Patients were required to have failed at least one but no more than four adequate antidepressant treatments in this or the most recent episode. Alternatively, patients were eligible if they had marked intolerance to antidepressants as demonstrated by four failed attempts to tolerate an adequate medication trial (lifetime).

Exclusionary criteria for study participation included a lifetime history of psychosis, bipolar disorder, or obsessive–compulsive disorder; posttraumatic stress disorder and eating disorders (if present in the past year); lack of response to an adequate trial of electroconvulsive therapy (ECT); prior treatment with TMS or a vagus nerve stimulator implant; pregnancy; a personal or close family history of a seizure disorder; presence of neurologic disorder or medication therapy known to alter seizure threshold; or medication therapy known to alter seizure threshold; or pregnancy; a personal or close family history of a seizure disorder; presence of neurologic disorder or medication therapy known to alter seizure threshold; or presence of ferromagnetic material in or in close proximity to the head. Routine laboratory studies (complete blood count, chemistry, thyroid stimulating hormone), urine toxicology screen, and electrocardiogram were performed at study screening, and subjects were required to be medically stable before entry.

Study Overview

The study was conducted at 23 study sites in the United States (20 sites), Australia (2 sites), and Canada (1 site), with active enrollment extending from January 2004 through August 2005. Institutional review board approval was obtained at all sites. The study was conducted under an Investigational Device Exemption from the U.S. Food and Drug Administration (FDA). All subjects signed an informed consent document before undergoing any study procedures. The study had three phases: a lead-in phase (1 week, no treatment), a 6-week acute treatment period (daily treatment with TMS or sham), and a taper phase (3 weeks reduced frequency of TMS or sham, start of antidepressant). Patients were randomized 1:1 to either active TMS or sham TMS. During the acute treatment phase, TMS sessions were scheduled daily in a 5-day sequence, for a maximum of 30 sessions (6 weeks), and typically administered on a Monday through Friday schedule. Then TMS was tapered in a blinded manner in six sessions across 3 weeks during which all patients were titrated onto monotherapy with an antidepressant medication. After 4 weeks of participation in the acute phase, if patients failed to show meaningful clinical benefit (i.e., < 25% reduction in baseline symptoms on the HAMD17), they could crossover to an open-label, acute treatment extension study. The specific criterion for early study exit was concealed from the investigators’ knowledge.

Study Device Description, TMS Session Procedures, and Ratings

The TMS sessions were delivered using the Neuronetics Model 2100 Therapy System investigational device (Neuronetics Inc., Malvern, Pennsylvania). Three separate magnetic coils, similar in weight, external appearance, and acoustic properties when actively pulsed, were used at each site, with one coil unblinded and used as the known active coil to determine motor thresholds. The other two coils differed in that the sham coil had an embedded magnetic shield. The latter limited the magnetic energy reaching the cortex to 10% or less than the active coil but nevertheless allowed the active and sham coils to have similar appearance, placement, and acoustic properties. All treatment personnel were blind as to coil assignment. All efficacy outcome measures were assessed by blinded study personnel (raters) who were not permitted access to the treatment sessions. Raters underwent certification in which their study participation was contingent on demonstrating adequate reliability in the conduct and scoring of interviews to derive HAMD and MADRS scores. Quality of ongoing ratings was assessed through the use of video monitoring reviewed by an independent expert. Patients were instructed not to disclose any details of the treatment session with the study raters during rating sessions. Furthermore, neither raters nor other study personnel at the specific centers were aware of the primary efficacy measure during the trial. Ratings were administered at baseline and at weeks 2, 4, and 6–9.

Treatment Parameters

Treatment was fixed at 120% magnetic field intensity relative to the patient’s observed resting motor threshold (MT), at a repetition rate of 10 magnetic pulses/sec, with a stimulus train duration (on time) of 4 sec and an intertrain interval (off time) of 26 sec. The left DLPFC was the treatment location and was determined by movement of the TMS coil 5 cm anterior to the motor threshold location along a left superior oblique plane with a rotation point about the tip of the patient’s nose (George et al. 1995b). Spatial coordinates were recorded with a mechanical coil positioning system to ensure placement reproducibility. The MT estimation was repeated weekly by visual observation of thumb or other finger movement (Pridmore et al. 1998) using the MT Assist (Neuronetics Inc.). The latter is a standardized, software-based mathematical algorithm that provides an iterated estimate of the MT.

During the first week of the acute phase only, treatment intensity could be adjusted to 110% of MT for tolerability but was then required to return to 120% MT from week 2 onward. A treatment session lasted for 37.5 min for a total of 3000 magnetic pulses delivered per session.

www.sobp.org/journal
Concomitant Treatments

All patients were free of antidepressants or other psychotropic medications directed at treating depression. Patients were allowed only limited use of either hypnotics or anxiolytics for treatment-emergent insomnia or anxiety, respectively. Up to 14 daily doses (lorazepam 2 mg/day equivalent) only were permitted (of either or both types of medications) during the acute phase.

Efficacy Assessments

The primary efficacy outcome was the difference between active and sham TMS using the last visit MADRS score through week 4 of the acute phase. Secondary outcome measures were the MADRS score at 6 weeks, 24-item and 17-item HAMD scores at 4 and 6 weeks, and categorical endpoints using MADRS, HAMD17, and HAMD24 at 4 and 6 weeks. Response was defined as at least 50% reduction from baseline score. Remission was defined by an absolute scale-specific score, as indicated in Figure 2. Several standardized HAMD factor scores were also derived: including the Depression Core factor (Items 1, 2, 3, 7, 8), the Maier subscale (Items 1, 2, 7, 8, 9, 10), the Gibbons subscale (Items 1, 2, 3, 7, 9, 10, 11, 14), the Anxiety/Somatization factor (Items 10, 11, 12, 13, 15, 17), the Retardation factor (Items 1, 7, 8, 14), and the Sleep factor (Items 4–6). Global clinical status was assessed using the observer rated Clinical Global Impressions Severity of Illness Scale (CGI-S).

Patient-reported outcomes were obtained using the Inventory of Depressive Symptoms—Self Report version (IDS-SR), and the Patient Global Impressions Improvement Scale (PGI-I). Various functional status and quality-of-life measures were collected and will be reported elsewhere.

Safety Assessments

Safety was assessed at every treatment visit by recording spontaneous adverse event reports that were coded using the current version of the Medical Dictionary for Regulatory Activities. Additional safety evaluations included targeted assessment of air-conduction auditory threshold at baseline, week 4, and week 6. Cognitive function was assessed periodically and will be the subject of a separate report.

Safety Assessments

Statistical Methods

Sample size was determined by requiring 90% power and a two-tailed 5% level in detecting a difference between the active and sham conditions with a putative MADRS effect size of approximately .40, based on the standard t test method. Efficacy analyses were performed on the strict intent-to-treat sample of all evaluable patients, defined in the protocol as those with a baseline and at least one postbaseline observation available for analysis. The null hypothesis for the primary outcome was tested with an analysis of covariance, using baseline score and ATHF medication resistance level as fixed-effect covariates, adjusting for site differences using a random effect (SAS Institute, Cary, North Carolina). Secondary outcome analyses for continuous measures were conducted in a similar fashion. All analyses were conducted in a last-observation carried forward (LOCF) manner through the indicated time points.

Results

Patient Disposition

Of the 325 patients randomized to TMS or sham, 301 (92.6%) had at least one postbaseline assessment, and their data comprised the a priori–specified analysis set. There were no baseline clinical or demographic differences between the 24 nonevaluable patients and the 301 evaluable subjects. Nonevaluable patients were evenly distributed between the active (n = 10) and the sham (n = 14) conditions, and there was no systematic difference between these groups in the reasons for discontinuation. Reasons given for study discontinuation through week 4, the primary efficacy time point, are provided in Figure 1.

Through the primary efficacy time point of 4 weeks, the overall discontinuation rate was low and similar in the active TMS (7.7%) and sham TMS (8.2%) treatment groups. Discontinuation because of adverse events of TMS was rare, similar across treatment conditions (4.5% of active TMS vs. 3.4% of sham TMS patients), and most commonly due to scalp discomfort. Beyond the primary efficacy time point, 74 (47.7%) patients in the active

Figure 1. Reasons for study discontinuation through the primary efficacy time point (week 4).

www.sobp.org/journal
TMS group and 92 (63.0%) patients in the sham TMS treatment group elected to enter the open-label extension study.

Sample Characteristics

Demographic and clinical features were similar in both groups. The CGI-S severity score of 4.7 for the study cohort corresponded to a depression severity between moderately and markedly ill. A minority of the sample reported a current episode duration greater than 2 years. The degree of treatment resistance in the current episode did not differ between groups, with an average of 1.6 failed adequate antidepressant treatments in the current episode and about half the study population failing to benefit from at least two treatments (by ATHF criteria). The degree of functional impairment in the study cohort was substantial, with about half of the sample unemployed and one third on disability because of their mood disorder.

As indicated in Table 1, baseline symptom severity at randomization differed between groups on the primary outcome measure (MADRS, \(p = .036\)) but not on any of the other symptom scales. This difference arose in the context of the protocol not specifying a minimal symptom severity score on the MADRS for study inclusion, differing in this respect from the HAMD17 (for which a minimum score of 20 was required at the screening visit). Six participants had baseline MADRS scores in the mild range (< 20, range 14–19) and were randomized unevenly across the two groups (i.e., 4 to active, and 2 to sham). This resulted in a small but significant difference in baseline symptom scores (i.e., 1.1 point).

Continuous Efficacy Outcome Measures

Efficacy results for continuous outcomes with the MADRS, HAMD24, and HAMD17 are shown in Figure 2.

MADRS

At the primary efficacy time point, week 4, the baseline to endpoint change on the MADRS showed a statistical trend favoring active TMS in the a priori–specified evaluable study population (\(p = .057\)). Given the observed baseline imbalance on the MADRS, a supplementary analysis was conducted on a study sample that included only those patients (\(n = 295\)) with a minimum baseline score of 20 on the MADRS, excluding the six patients noted earlier. In this analysis, the baseline to endpoint change was statistically significant at week 4 (\(p = .038\). At the secondary efficacy time point, week 6, the baseline to endpoint change on the MADRS continued to show a statistical trend favoring active TMS in the a priori–specified evaluable study population (\(p = .057\)) and in the subset of patients with a baseline MADRS > 20 (\(p = .052\)).

Table 1. Demographic Features, Clinical Features, and Baseline and End Point (Weeks 4 and 6) Symptom Scores of the Study Sample

<table>
<thead>
<tr>
<th>Demographic Variables</th>
<th>Active TMS (n = 155)</th>
<th>Sham TMS (n = 146)</th>
<th>(p) Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (% female)</td>
<td>86 (55.5)</td>
<td>74 (50.7)</td>
<td>.421</td>
</tr>
<tr>
<td>Age (years ± SD)</td>
<td>47.9 ± 11.0</td>
<td>48.7 ± 10.6</td>
<td>.509</td>
</tr>
<tr>
<td>Ethnic origin, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>146 (94.2)</td>
<td>131 (89.7)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>9 (5.8)</td>
<td>15 (10.3)</td>
<td>.201</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full time</td>
<td>58 (35.6)</td>
<td>45 (28.3)</td>
<td></td>
</tr>
<tr>
<td>Part time</td>
<td>27 (16.6)</td>
<td>31 (19.5)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>78 (47.9)</td>
<td>83 (52.2)</td>
<td></td>
</tr>
<tr>
<td>Receiving disability compensation</td>
<td>28 (32.9)</td>
<td>31 (34.1)</td>
<td></td>
</tr>
<tr>
<td>Clinical Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrent illness course (%)</td>
<td>149 (95.5)</td>
<td>136 (93.8)</td>
<td>.611</td>
</tr>
<tr>
<td>Duration of current episode in months, Mean (SD)</td>
<td>13.6 (9.9)</td>
<td>13.2 (9.5)</td>
<td>.728</td>
</tr>
<tr>
<td>n (%) of population with current episode &gt; 2 years</td>
<td>36 (23.2)</td>
<td>23 (15.8)</td>
<td>.112</td>
</tr>
<tr>
<td>Number of fully adequate antidepressant treatments in current episode</td>
<td>1.6</td>
<td>1.6</td>
<td>.816</td>
</tr>
<tr>
<td>Baseline Symptom Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS total score (SD)</td>
<td>32.8 (6.0)</td>
<td>33.9 (5.7)</td>
<td>.036</td>
</tr>
<tr>
<td>HAMD17 total score (SD)</td>
<td>22.6 (3.3)</td>
<td>22.9 (3.5)</td>
<td>.508</td>
</tr>
<tr>
<td>HAMD24 total score (SD)</td>
<td>30.1 (5.0)</td>
<td>30.5 (4.9)</td>
<td>.568</td>
</tr>
<tr>
<td>CGI-Severity (SD)</td>
<td>4.7 (6.0)</td>
<td>4.7 (7.7)</td>
<td>.594</td>
</tr>
<tr>
<td>IDS-SR total score (SD)</td>
<td>42.0 (9.4)</td>
<td>43.4 (9.9)</td>
<td>.197</td>
</tr>
<tr>
<td>Week 4 Symptom Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS total score (SD)</td>
<td>27 (11.1)</td>
<td>29.8 (10.1)</td>
<td>.057</td>
</tr>
<tr>
<td>HAMD17 total score (SD)</td>
<td>17.4 (6.5)</td>
<td>19.4 (6.5)</td>
<td>.006</td>
</tr>
<tr>
<td>HAMD24 total score (SD)</td>
<td>23.4 (8.9)</td>
<td>25.9 (8.8)</td>
<td>.012</td>
</tr>
<tr>
<td>Week 6 Symptom Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS total score (SD)</td>
<td>26.8 (12.8)</td>
<td>30.10 (8.0)</td>
<td>.058</td>
</tr>
<tr>
<td>HAMD 17 total score (SD)</td>
<td>17.1 (7.7)</td>
<td>19.6 (7.0)</td>
<td>.005</td>
</tr>
<tr>
<td>HAMD 24 total score (SD)</td>
<td>23.2 (10.6)</td>
<td>26.9 (4.4)</td>
<td>.015</td>
</tr>
</tbody>
</table>

CGI-S, Clinical Global Impressions Severity of Illness; HAMD, Hamilton Depression Rating Scale (17 and 24 item); IDS-SR, Depressive Symptoms—Self Report version; MADRS, Montgomery–Asberg Depression Rating Scale; TMS, transcranial magnetic stimulation.

*p values represent between group contrasts at baseline except for the weeks 4 and six symptom scores where the \(p\) values indicate the contrast of the within group changes from baseline.
Results for the HAMD24 and HAMD17 are presented for the a priori–specified evaluable data set only (n/H11005 301). No baseline imbalance in symptom severity occurred with these scales, and the findings were unaltered when the six patients with low scores on the MADRS were excluded (data available from the authors on request). At the primary efficacy time point, week 4, the baseline to endpoint change on both the HAMD17 and the HAMD24 yielded a significant main effect of treatment group favoring active TMS (p = H11005.006 and p = H11005.012, respectively). This outcome was sustained at the secondary efficacy time point, week 6, with a significant advantage in favor of active TMS (p = H11005.005 for HAMD17, p = H11005.015 for HAMD24). As seen in Figure 2A–C, it is notable that the time course of clinical effect showed a sustained improvement across the acute treatment phase for the active TMS group, whereas the sham TMS group showed a pattern of early change that dissipates at the later time points.

**CGI-S**

Clinician-rated global illness severity showed greater improvement with active TMS compared with sham TMS as early as week 2 of the acute treatment phase (p = H11005.047) and continued through the primary efficacy time point of week 4 (p = H11005.009) and the secondary efficacy time point of week 6 (p = H11005.012).

**Categoric Outcomes**

**Response Rates.** As seen in Figure 3A–C, at the week 4 primary efficacy time point, the response rate as defined by the three symptom-rating scales (MADRS, HAMD17, and HAMD24) was higher with active compared with sham TMS. This effect was sustained through the secondary efficacy time point at week 6.

**Remission Rates.** As seen in Figure 3A–C, at the week 4 primary efficacy time point, a significant difference in remission rates was not detected. At the secondary efficacy time point of week 6, however, remission rate was higher with active compared with sham TMS on the MADRS and HAMD24.

**HAMD Factor Scores.** The pattern of change in the HAMD17 scale was examined further by the assessment of previously established factor scores for the HAMD. These results are shown in Table 2 and demonstrate significantly better outcomes for active TMS compared with sham TMS on core depression symptoms (Core Depression Factor, Maier Factor, and Gibbons Factor), anxiety symptoms (Anxiety/Somatization Factor), and vegetative symptoms (Retardation Factor) at weeks 4 and 6.

**Patient-Reported Mood Symptoms and Global Improvement.** Patient-reported mood change and global improvement were assessed using the IDS-SR and the PGI—Improvement scales. The pattern of change followed the clinician-reported measures but showed a less robust effect. The IDS-SR exhibited a trend toward improved outcome in active compared with sham TMS at the primary efficacy time point, week 4 (p = H11005.058), which was maintained at the secondary efficacy time point, week 6 (p = H11005.053). There was no statistically significant separation between the groups on the PGI-I at either time point.

**Antidepressant Effects During TMS Taper.** During this phase, response and remission rates of the active TMS group improved incrementally. With the MADRS, the response rate increased from 23.9% to 27.7% and the remission rate from 14.2% to 20.6% for active TMS. In contrast, addition of medication to the sham TMS group produced little meaningful clinical change during this phase: a change from 12.3% to 13.7% for the response rate and from 5.5% to 8.9% for remission on the MADRS.

**Safety Outcomes**

**Spontaneous Adverse Events.** There was a higher incidence of scalp discomfort and pain with active than sham TMS (Table 3). These events were generally reported as mild or moderate in severity and diminished rapidly in incidence after the first treatment week. Scalp discomfort had the potential to compromise the study blind, and a separate analysis was conducted to examine the relationship between clinical outcome and the experience of cutaneous discomfort. The findings were negative regarding an association between any of these adverse event terms and the primary outcome measure (data not shown).
Serious Adverse Events. There were no deaths in this study, and no seizures were reported. During the acute treatment phase, 16 serious adverse events were reported, 9 in the active TMS group and 7 in the sham TMS group. Events reflecting disease-related exacerbation were the most common serious adverse events. These included suicidality (1.9% with sham vs. 6% with active TMS), exacerbation of depression (1.9% with sham vs. 6% with active), and a single suspected suicide gesture (in the sham group). Overall risk of exacerbation of suicidality was evaluated by determining the proportion of patients in either group who increased in score on the suicide item of the HAMD (item 3) from a value of 0 or 1 at baseline to a value of 3 or 4 at any time point during the acute treatment phase. Cumulatively, 10 events meeting this criterion were observed in the sham TMS group compared with 1 event in the active TMS group.

Discussion

This is the first large, multisite, randomized controlled trial of daily left prefrontal TMS in medication-free patients with major depression who had failed to receive adequate benefit from prior antidepressant treatment. The findings indicate that TMS, administered at these parameters for a period of 4–6 weeks, is safe and effective in the treatment of major depression.

The pattern of symptom improvement was consistent across the various study outcome measures. Active treatment with TMS was significantly superior to sham TMS treatment for the change in mean symptom score using the HAMD17 and HAMD24 at weeks 4 ($p = .006$, $p = .012$ respectively) and 6 ($p = .005$, $p = .015$). After correction for the baseline score imbalance, the MADRS also showed this pattern ($p = .038$). Clinically important change, as reflected in terms of the categoric outcomes of response and remission, was also achieved in a substantial portion of patients. At 6 weeks, the active TMS group was about twice as likely to have achieved remission compared with the sham TMS group (MADRS:14.2% vs. 5.2%, HAM-D17: 15.5% vs. 7.1%, HAMD24: 17.4% vs. 8.2%). Improvement on the self-report IDS scale was at the level of a trend ($p = .058$ week 4; $p = .053$ week 6). Clinical outcomes were observed in the setting of a favorable tolerability profile, with less than 5% of patients on active TMS discontinuing treatment because of adverse effects by the primary efficacy endpoint of 4 weeks.

The trajectory of improvement, as indicated in Figure 2, implies that more than 2 weeks of TMS, compared with sham, is required in this population before a significant improvement is detected. Similarly, it appears that an additional 2 weeks of TMS beyond the initial 4 weeks, as indicated in Figure 3, can have an important clinical impact. The remission rates doubled during that period of time.

Table 2. HAMD Factor Scores—Contrasts Between Active TMS and Sham TMS during the Acute Treatment Phase

<table>
<thead>
<tr>
<th>HAMD Rating Scale Factor Score</th>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressive Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core depression factor</td>
<td>&gt;.100</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>Maier factor</td>
<td>&gt;.100</td>
<td>.003</td>
<td>.003</td>
</tr>
<tr>
<td>Gibbons factor</td>
<td>&gt;.100</td>
<td>.007</td>
<td>.006</td>
</tr>
<tr>
<td>Anxiety Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety/somatization factor</td>
<td>&gt;.100</td>
<td>.025</td>
<td>.023</td>
</tr>
<tr>
<td>Vegetative symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retardation factor</td>
<td>.057</td>
<td>.007</td>
<td>.003</td>
</tr>
<tr>
<td>Sleep factor</td>
<td>&gt;.100</td>
<td>&gt;.100</td>
<td>&gt;.100</td>
</tr>
</tbody>
</table>

HAMD, Hamilton Depression Rating Scale; TMS, transcranial magnetic stimulation. All contrasts, except for the sleep factor, reflect a superior outcome at weeks 4 and 6 for the active TMS compared with the sham TMS treatment condition. $p$ values shown for analysis of covariance model; post hoc contrast between active and sham TMS at the time points indicated.
Table 3. Adverse Events Occurring in the Active Treatment Group at a Rate of 5% or More and at Least Twice the Rate for Sham (with ME-Coded Preferred Terms Shown)

<table>
<thead>
<tr>
<th>Body System</th>
<th>Active TMS (n = 165)</th>
<th>Sham TMS (n = 158)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred term</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td><strong>Eye disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye pain</td>
<td>10 (6.1)</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Gastrointestinal Disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toothache</td>
<td>12 (7.3)</td>
<td>1 (.6)</td>
</tr>
<tr>
<td><strong>General Disorders and Site Administration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application site discomfort</td>
<td>18 (10.9)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Application site pain</td>
<td>59 (35.8)</td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>Facial pain</td>
<td>11 (6.7)</td>
<td>5 (3.2)</td>
</tr>
<tr>
<td><strong>Musculoskeletal and connective tissue disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle twitching</td>
<td>34 (20.6)</td>
<td>5 (3.2)</td>
</tr>
<tr>
<td><strong>Skin and subcutaneous tissue disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain of skin</td>
<td>14 (8.5)</td>
<td>1 (.6)</td>
</tr>
</tbody>
</table>

MedDRA, Medical Dictionary for Regulatory Activities.

In this study, TMS was well tolerated and safe. The dropout rate for any reason from active TMS was 7.7% at 4 weeks, and discontinuation specifically because of side effects was 4.5%. This is lower than the discontinuation rates generally reported from clinical trials of marketed antidepressants. Adverse events reported were principally limited to scalp discomfort or pain within the confines of the TMS session itself and were mostly transient phenomena in the first weeks of the TMS course. In fact, the incidence of headache did not differ between active and sham TMS conditions.

This study also suggests that the upper limit on the safe administration of TMS may be somewhat greater than we currently suspect. Despite TMS being administered here at 120% of motor threshold and 3000 pulses/session, an elevated rate of serious adverse events relative to sham was not detected. The most serious side effect that has been reported with TMS is a seizure (Anderson et al. 2006), and none were observed in this study. No adverse effects on mood in terms of either treatment-emergent suicidality or treatment-emergent mania or hypomania were observed. In fact, the frequency of treatment-emergent suicidal ideation was numerically lower in the active TMS group (1 event observed with active vs. 10 events observed with sham during the acute treatment phase).

How can the clinical significance of these results be placed in context for the practitioner? In answering this question, it is important to consider the treatment resistance of the population of patients studied. On average, the patients included had an average of 1.6 failed adequate trials in the current episode of major depression, and nearly half had failed to benefit from 2 or more treatments. The importance of prior resistance to antidepressant treatment is to diminish significantly the likelihood of responding to subsequent interventions. This pattern has been demonstrated with respect to outcomes with electroconvulsive therapy (Prudic et al. 1990, 1996; Sackeim et al. 1990), pharmacotherapy (Trivedi et al. 2006), and vagus nerve stimulation (Sackeim et al. 2001).

The most comprehensive demonstration of this in the recent literature is from the large, open-label, seminaturalistic STAR*D trial (Trivedi et al. 2006). In that program, there was a progressive reduction in remission rates with each stage of resistance. In fact, patients who had failed two prior antidepressant treatments experienced a mean remission rate (defined as a HAMD17 score < 8) in level 3 treatment of only 16.2% after a 10- to 12-week course of treatment (Fava et al. 2006). In the controlled trial results reported here, at the end of 6 weeks of treatment with TMS alone, patients experienced a remission rate of 15.5%, and this increased to 22.6%, by the same HAMD17 criteria, after 9 weeks during the taper phase of the study. Given that controlled trials generally report somewhat lower clinical response and remission rates than are seen in open-label experience, the results reported here compare favorably to those seen in similarly treatment-resistant patients in the STAR*D reports.

Controlled trials, where an active antidepressant medication is measured against its relevant within-study control, provide the most appropriate comparison for the TMS results. In the large analysis of the FDA database of approved antidepressant medications, Khan and colleagues (2000) noted that the mean percentage reduction from baseline in total HAMD17 score across the entire data set of antidepressants was 40.7% for active treatment and 30.9% for placebo. On average, this represents an overall relative advantage of about 10% in the reduction of total score from baseline on the HAMD17 when comparing active antidepressant to placebo treatment. By comparison, in this study, at 4 weeks, active TMS treatment resulted in a reduction in HAMD17 score of approximately 25% compared with 15% in the sham group. This represents an overall relative advantage of about 8% for active TMS compared with sham in reduction of the total HAMD17 score. Similarly, application of the metric of number needed to treat (NNT; Kraemer and Kupfer 2006) as an indicator of effect size, yielded an NNT for TMS of 11 at week 4 and 9 at week 6. Thase et al. (2005), recently reported a pooled estimate of response rates in a large sample (N = 1795) of patients treated with bupropion, various selective serotonin reuptake inhibitors, or placebo and noted an overall response rate in active-treated patients of 62.8% compared with 50.8% in placebo treated patients, yielding an NNT in that sample of eight. Thus, by these metrics, the antidepressant efficacy of TMS is comparable to that of standard pharmacotherapy.

This study has several limitations that should be considered when interpreting the results. Although a rigorous method was used to assess prior treatment resistance, namely, the ATHF, this method relies on retrospective report. Therefore, greater certainty regarding the level of treatment resistance would have been obtained from a prospective, open-label antidepressant lead-in phase and subsequently enrolling only nonresponders to that intervention. This prospective method, however, requires considerably more time and cost to implement and is therefore not commonly used. The data documented in this trial support a moderate level of treatment resistance for the study population. Moreover, the observed response and remission rates with the sham TMS intervention were very low, providing additional evidence of the treatment resistance of this patient population.

The study employed an innovative approach to sham methodology that represents a clear advance over prior work. Special efforts were made to match the active and sham conditions in procedure, sound, and sensation as closely as possible, while substantially limiting the exposure of the cortex to the actual magnetic field, as discussed in the study methods. Formal query of patients and treaters to assess the adequacy of the blind, however, was not conducted. Results from prior studies indicate that previously TMS-naïve patients primarily base judgment of whether they received an active or sham TMS procedure on clinical outcome (Fitzgerald et al. 2003). A secondary analysis (results to be presented in a subsequent report) also indicated that scalp discomfort with active TMS did not correlate with

www.sobp.org/journal

J.P. O'Reardon et al.
treatment outcome. Thus, unblinding of the active condition is an unlikely explanation for the therapeutic advantage of active TMS.

The primary efficacy endpoint in this study was at week 4 of the 6-week acute treatment phase. At or after that time point, if clinically justified, patients were eligible to enter the open-label extension study. Complete randomization was thus truly present through the primary efficacy time point. Therefore, outcomes at time points after week 4 should be interpreted in this context. On the other hand, because the treatment assignment blind was maintained throughout the study, it is worth noting that active TMS subjects continuing through the end of the acute phase and into the taper phase of the study showed persistent and perhaps accumulating benefit in comparison to their continuing counterparts in the sham TMS group. Finally, the method of coil positioning used to identify the left DLPFC was not based on guidance by means of an magnetic resonance imaging–assisted neuronavigational method. Rather, we used a probabilistic surface anatomy approach targeting 5 cm anterior to the motor threshold location, like most clinical trials of TMS.

In conclusion, TMS administered over the left DLPFC using the parameters reported here for a period of up to 6 weeks was effective in treating major depression and with a good tolerability profile. These results indicate that TMS offers clinicians a novel alternative in the treatment of this disorder (Sackeim 2001a).

The authors thank the patients who participated in this clinical trial. In addition, we thank the Steering Committee for this trial (Mark George, MUSC, Harold Sackeim, Columbia University; Alan Schatzberg, Stanford), the principal investigators who were members of the Neurotics TMS Study Group, and their staff; Scott Aaronson, Sheppard Pratt Health System; David Avery, University of Washington Medical Center; Randolph Canterbury, University of Virginia; Z. Jeffrey Daskalakis, Centre for Addiction and Mental Health; James Ferguson, Radiant Research; Paul Fitzgerald, The Alfred Psychiatry Research Centre; William Gilmer, Northwestern University; Roshen Gutierrez, Psy-Care, Inc.; Mustafa Husain, University of Texas—Southwestern Medical Center; Keith Isenberg, Washington University School of Medicine; Philip G. Janicak, Rush University Medical Center; Andrew Krystal, Duke University Medical Center; Sarah H. Lisanby, New York State Psychiatric Institute; Colleen Loo, University of New South Wales; Daniel Maixner, University of Michigan Medical Center; Lauren Marrangell, Baylor College of Medicine; William McDonald, Emory University; Ziad Nahas, Medical University of South Carolina; John P. O’Reardon, University of Pennsylvania; Elliott Richelson, Peter Rosenquist, Wake Forest University Health Sciences; Shirlene Sampson, Mayo Clinic; Brent Sokolov, Stanford University.

Dr. Avery has received grant support from the National Institute of Mental Health (NIMH) and Neurontics; has acted as a consultant to Bristol Myers Squibb (BMS), Cyberonics, GlaxoSmithKline (GSK), Eli Lilly, Janssen, Pharmaceutica Products, Neurontics, Performance Plus, and Takeda; and has been a member of speakers bureaus for BMS, Cyberonics, GSK, Eli Lilly, Janssen, and Pfizer. Dr. Demitrack is an employee of Neurontics, the study sponsor, and owns equity in Lilly, Wyeth, and Neurontics. Dr. Fitzgerald has received research support from Neurontics. Dr. George has received grant support from Abbott, Cephor, Cortex, Cyberonics, Dantec, DarPharma, DARPA, GSK, Jazz, National Institutes of Health (NIH), NARSAD, Neutontus, NeuroPace, Neurontics, and the Stanley Foundation; has acted as consultant to Abbott, Aspect Biomedical, Aventis, Jazz Pharmaceuticals, Argyn Pharmaceuticals, Neuraliive, Neurontics, and NeuroPace; and has been a member of speakers bureaus for Janssen, Lilly, GSK, Parke Davis, Picker Int, Cyberonics, and MedPhysics/Amersham. Dr. Isenberg has received grant support from NIMH and Neurontics and has acted as a consultant to Wellpoint and Barnes Jewish-Christian Behavioral Health. Dr. Janicak has received grant support from Astra Zeneca, BMS, Janssen, Neurontics, and Solvay; has acted as consultant to Astra Zeneca, BMS, Janssen, Neurontics, and Shire; and is a member of speaker bureaus for Abbott, Astra Zeneca, BMS, Janssen, Pfizer, and Shire. Dr. Loo has received research support from National Health and Medical Research Council (Australia), Neurotics, NSW Schizophrenia Fellowship, Pfizer Neuroscience Research Grant scheme, and Rebecca Cooper Medical Research Foundation. Dr. McDonald has received grant support from the American Foundation for Suicide Prevention, Faqua Foundation, National Institute of Mental Health, National Institute of Neurological Disease and Stroke, National Alliance for Research in Schizophrenia and Depression, Neurontics, and Janssen; has acted as consultant to Neurontics, Janssen, Forest, BMS; has been a member of speakers bureaus for BMS, Forest, Janssen, and Solvay; and has equity in Amgen, Teva, Pfizer, and Abbott. Dr. Nahas has received grant support from Cyberonics, Eli Lilly, Integra, Medtronic, Neurontics, NeuroPace, and NIMH; has acted as a consultant to Avanir Pharmaceutical, Aventis Pharmaceuticals, Cyberonics, Neurontics, and NeuroPace; and has been a member of the speakers bureau for Cyberonics. Dr. O’Reardon has received grant support from BMS, Cyberonics, Lilly, Magstim, Neurontics, Pfizer, and Sanoﬁ; acted as consultant for Lilly and Neurontics; and is a member of speakers bureaus for BMS, Cyberonics, and Lilly. Dr. Sackeim has received research support from the Cyberonics, MECTA, National Alliance for Research in Schizophrenia and Depression, and NIMH and has acted as a consultant to Cyberonics, Eli Lilly, Magstim, MECTA, Neurontics, NeuroPace, Novartis, and Pfizer. Dr. Sampson has received research support from Neurontics and Pfizer and has acted as a consultant to NeuroPace. Dr. Solason has received grant support from Astra Zeneca, BMS, Forest, GSK, and Neurontics and has served as a consultant to Cephalon, Cyberonics, Forest, Lilly, NeuroPace, and Seprocar.

Clinical trial is posted on www.clinicaltrials.gov. Listing No. NCT 00104611. Supported by a grant from Neuronetics Inc.


www.sobp.org/journal


