Clinical Significance of Transcranial Magnetic Stimulation (TMS) in the Treatment of Pharmacoresistant Depression: Synthesis of Recent Data

By Mark A. Demitrack, Michael E. Thase

ABSTRACT - Transcranial magnetic stimulation (TMS) is a novel treatment for patients with major depressive disorder. Although clearly safer and better tolerated than many other pharmacotherapeutic options or electroconvulsive therapy, questions have persisted about the magnitude of the efficacy of TMS in patients with pharmacoresistant depression, and the clinical significance of these outcomes. Previous studies have explored whether specific patient characteristics are associated with a greater likelihood of clinical benefit. In the largest such analysis conducted to date, the authors confirmed previous observations that the lower the number of prior failed antidepressant treatments, the better the clinical outcome of treatment with TMS. This relationship between prior treatment resistance and subsequent treatment outcome is consistent with previous evidence from antidepressant studies. The authors examined the clinical significance of the treatment effects seen with TMS in pharmacoresistant major depression in their recently completed studies by comparing these outcomes with the results reported in several large, comprehensive published reference datasets of antidepressant medications studied in both treatment-responsive and treatment-resistant patient populations. The efficacy of TMS demonstrated in randomized controlled trials was comparable to that of pharmaceutical antidepressants studied in similarly designed registration trials and to the adjunctive use of atypical antipsychotic medications in controlled trials of antidepressant non-responders. These data may be helpful in treatment-planning decisions when using TMS in clinical practice. Psychopharmacology Bulletin. 2009;42(2):5–38.

INTRODUCTION

Major depression is a severe, disabling, and potentially lethal clinical disorder (Kessler et al., 2003; Murray and Lopez, 1996; American Foundation for Suicide Prevention, 2009). While there are a wide variety of pharmaceutical agents available as treatments for major depression, almost all, have been studied in and approved for use only as initial treatments. Of these agents, only about half of patients respond to an initial course of antidepressant pharmacotherapy (Sackeim,
2001; Rush, Thase and Dubé, 2003), and for these patients, current standard of care involves an empirical series of treatment attempts, typically using medication switches, antidepressant combinations, or adjunctive therapy with mood stabilizers, benzodiazepines, atypical antipsychotics, or other agents (Greden, 2002). The adverse event burden and tolerability of some of these more complex interventions is not trivial and is a major factor that hinders patient adherence to treatment (Papakostas, 2008). Moreover, in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) project, the therapeutic potential of several of the more advanced treatment options, including lithium augmentation (Nierenberg et al., 2006) and switching to some of the alternate antidepressants studied (nortriptyline, mirtazapine and tranylcypromine; Fava et al., 2006; McGrath et al., 2006), was limited by difficulties with implementing trials of adequate dose and duration. Similarly, although there is increasing evidence that at least some of the atypical antipsychotics are effective as adjuncts to antidepressants, the potential for later emerging side effects, including weight gain and dyslipidemia, warrants both caution and careful clinical management.

It has been conservatively estimated that at least 15% of all patients with major depressive disorder remain refractory to any treatment intervention (Greden, 2002; Rush et al., 2003). While a complex relationship exists between disease chronicity and ineffective treatment (Leon et al., 2003), clinical evidence argues that the greater the number of treatment failures, the less likelihood of a good treatment response to subsequent interventions (Prudic et al., 1996; Sackeim, 2001). The recently reported results of the STAR*D study are the most vivid example of this clinical phenomenon (Trivedi et al., 2006a, 2006b; Rush et al., 2006a, 2006b; Fava et al., 2006; Nierenberg et al., 2006; McGrath et al., 2006). In that work, there was a progressive likelihood of poorer response with each successive treatment failure. For example, after the first treatment attempt, about 30% of patients remitted. By the time that a patient had experienced 3 definitive treatment failures, the likelihood of achieving remission with the fourth treatment option offered fell below 10%. Good clinical outcomes are also dependent upon the safety and side effect profile of the medication and hence the patient’s ability to tolerate the treatment at a sufficient dose for an adequate length of time. When studied with rigorous assessment tools, reaching this minimum level of treatment adequacy occurs usually only in about one in four treatment attempts (Prudic et al. 2004). Taken together, these facts underline the clinical urgency for physicians to identify and respond to early evidence of treatment resistance using treatments that have proven efficacy in these more difficult-to-treat patients.
Transcranial magnetic stimulation (TMS) is a non-invasive method used to electrically stimulate superficial cortical neurons, using rapidly alternating MRI-strength magnetic fields generated by a specially-designed magnetic coil held in contact with the surface of the head (Davey and Epstein, 2000). Although a substantial body of work has established that TMS is safe and effective in the treatment of patients with major depression, questions have persisted about the relative efficacy of TMS (i.e., as compared to other options used in patients with difficult-to-treat depression) and whether or not particular subgroups of patients are more or less likely to benefit from this intervention.

To further refine an understanding of the efficacy profile of TMS, two separate analyses of large clinical trial datasets were explored in order to systematically identify potential predictors of antidepressant treatment response. In the first analysis, using data from 195 patients across six different clinical trials, Fregni and colleagues (2006) reported that younger age and less prior treatment resistance were associated with better response to TMS. In a separate study, Brakemeier and colleagues (2007) performed a similar analysis in an open-label treatment study of 70 patients. Consistent with the results of Fregni and colleagues, they also found that those patients with the least treatment resistance showed the greatest clinical benefit.

The authors have recently reported the overall results of the first large, multisite, double-blind randomized controlled trial of TMS in the treatment of pharmacoresistant major depression (O’Reardon et al., 2007). The results from that work demonstrate that TMS is safe and effective in patients with unipolar major depression. In subjects enrolled, there was a wide range of magnitude of treatment resistance prior to study entry. Specifically, inclusion criteria stipulated that patients had to have failed to receive benefit from at least one but no more than four previous antidepressant treatments at minimal effective dose and duration in the current or most recent episode of their illness. We subsequently reported an analysis of pretreatment predictors of outcome in this large, multisite randomized controlled trial (Lisanby et al. 2008). Consistent with the predictor analyses reported in the works of Fregni (2006) and Brakemeier (2007) described above, we replicated the observation that the best predictor of treatment response was the degree of prior treatment resistance. In particular, treatment benefit with TMS was most striking in those patients in our randomized controlled trial dataset who were early in the course of their treatment resistance, namely those who had failed to benefit from one previous treatment at an effective dose and duration in the current illness episode.

The primary goals of the current review are to provide a complete summary of the efficacy outcomes observed within the treatment
cohort of the original randomized controlled trial, and because there are few large randomized controlled trials available of treatment options for patients who have clearly demonstrated resistance to pharmacotherapy, to provide a comprehensive comparison of outcomes to those with similar clinical endpoints as reported in other reference datasets from the published literature of antidepressant medications as studied in both treatment-responsive and pharmacoresistant forms of major depression.

METHODS

Overview of the TMS Clinical Development Program

A double-masked, randomized controlled trial of the antidepressant efficacy of TMS in patients with pharmacoresistant unipolar depression was conducted at 23 clinical sites in the United States (N = 20), Australia (N = 2) and Canada (N = 1). Active enrollment occurred between January 2004 and August 2005. The study was designed to examine the efficacy of the Neuronetics NeuroStar TMS Therapy System compared to a sham TMS treatment condition (O’Reardon et al., 2007). A second, open-label trial (Avery et al., 2008) that followed the same treatment sequence as the randomized controlled trial was available for all patients who: a) had participated in the first study for at least 4 weeks and, b) had not received significant clinical benefit from their randomized assignment. The criterion defining failure of clinical benefit in the randomized controlled trial was applied in a blinded manner (i.e., both investigators and patients were unaware of the eligibility criterion for enrollment in the extension study). The specific criterion to determine eligibility for entry into the open-label extension study was failure to achieve at least a 25% improvement in total score on the 17-item Hamilton Depression Rating Scale (HAM-D17) compared to baseline assessment.

In both the randomized and open-label studies, TMS was administered as a monotherapy. Treatment parameters included stimulation at 120% of observed motor threshold, applied at a pulse frequency of 10 pulses per second (in repeated 30-second duration cycles of 4 seconds on and 26 seconds off), for a total of 3000 pulses per session. Treatment location was targeted to the left dorsolateral prefrontal cortex (DLPFC), defined as the site 5 cm anterior to the optimal area for stimulating the contralateral thumb, with the motor threshold location and treatment location oriented in the same left superior oblique plane. The treatment location was recorded with a mechanical positioning system to permit repeatable placement of the magnetic coil at each subsequent treatment session. All TMS sessions were delivered using the NeuroStar TMS Therapy System, under an approved investigational device exemption.
A more detailed discussion of the study design and the efficacy and safety outcomes of both the randomized controlled trial and the open-label extension study have been reported elsewhere (O’Reardon et al. 2007; Janicak et al., 2008; Avery et al., 2008; Demitrack and Lisanby, 2008).

**Subject Description and Assessment of Prior Treatment Resistance**

Briefly, all patients met Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnostic criteria for unipolar, nonpsychotic major depressive disorder, confirmed by the Structured Clinical Interview for the DSM-IV (Ventura et al., 1998). A recurrent course of illness was reported in nearly all patients. In addition, patients were moderately to severely ill by symptom measures at baseline and moderately to severely resistant to pharmaceutical antidepressant treatment in the current illness episode. In the current episode alone, at the time of entry into the randomized controlled trial, the overall patient population had received an average of 5.5 antidepressant treatments (SD 3.5, range 1 to 23). There was no limit placed on the number of lifetime treatment failures. Adequacy of each antidepressant treatment attempt was determined with the Antidepressant Treatment History Form (ATHF; Sackeim, 2001). The ATHF is a reliable and validated method of rating the adequacy of antidepressant treatment trials and has been shown to have prospective validity in predicting future outcomes based on the number of prior failures of adequate treatment (Prudic et al. 1996). Use of the ATHF in other studies has also shown that for each adequate antidepressant treatment trial, a patient has usually experienced an average of 4 treatment attempts (Prudic et al. 2004).

Patients who had failed one adequate antidepressant trial in the current illness episode, as confirmed by the ATHF, comprised the majority 54.5% (N = 164 patients) of the overall study population. Among the remainder, 31.6% (N = 95) were classified as having not responded to two adequate treatment trials, with 10.0% (N = 30) and 4.0% (N = 12) not responding to three and four treatment trials, respectively. A summary of clinical and demographic characteristics for the subgroup of patients who had failed one previous antidepressant treatment in the current episode is shown in Table 1. With the exception of their level of prior antidepressant treatment resistance, there were no statistically significant or clinically meaningful differences between characteristics of the ATHF = 1 patient subgroup and those of the remainder of the study population (Lisanby et al. 2008). A complete description of the inclusion and exclusion criteria for patient recruitment is described elsewhere (O’Reardon et al. 2007).
**TABLE 1**

**DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE PATIENT STUDY POPULATION WHO FAILED TO RECEIVE BENEFIT FROM ONE PRIOR ADEQUATE ANTIDEPRESSANT MEDICATION TREATMENT FAILURE IN CURRENT ILLNESS EPISODE**

<table>
<thead>
<tr>
<th>VARIABLE NAME</th>
<th>NEUROSTAR TMS (N = 88)</th>
<th>SHAM TMS (N = 76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in Years Mean (SD)</td>
<td>48.6 (10.8)</td>
<td>51.9 (9.6)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Female N (%)</td>
<td>44 (50.0)</td>
<td>42 (55.3)</td>
</tr>
<tr>
<td>Ethnic Origin N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Caucasian</td>
<td>81 (92.0)</td>
<td>67 (88.2)</td>
</tr>
<tr>
<td>• All Other</td>
<td>7 (8.0)</td>
<td>9 (11.8)</td>
</tr>
<tr>
<td>Depression History N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Single Episode</td>
<td>3 (3.4)</td>
<td>2 (2.6)</td>
</tr>
<tr>
<td>• Recurrent</td>
<td>85 (96.6)</td>
<td>74 (97.4)</td>
</tr>
<tr>
<td>Duration of Current Episode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Length in Months [Mean (SD)]</td>
<td>12.8 (9.9)</td>
<td>11.9 (9.1)</td>
</tr>
<tr>
<td>• ≥2 Years N (%)</td>
<td>17 (19.3)</td>
<td>10 (13.2)</td>
</tr>
<tr>
<td>Secondary Diagnoses N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• None</td>
<td>57 (64.8)</td>
<td>55 (72.4)</td>
</tr>
<tr>
<td>• Any Anxiety Disorder</td>
<td>31 (35.2)</td>
<td>21 (27.6)</td>
</tr>
<tr>
<td>Antidepressant Treatment in Current Illness Episode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Number of Overall Antidepressant Treatment Attempts</td>
<td>4.6 (3.0)</td>
<td>4.7 (3.8)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4.6 (3.0)</td>
<td>4.7 (3.8)</td>
</tr>
<tr>
<td>Median (Range)</td>
<td>4 (1,18)</td>
<td>4 (1,23)</td>
</tr>
<tr>
<td>• Number Antidepressant Treatments of Adequate Dose/Duration</td>
<td>1.0 (0.0)</td>
<td>1.0 (0.0)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.0 (0.0)</td>
<td>1.0 (0.0)</td>
</tr>
<tr>
<td>Median (Range)</td>
<td>1 (1,1)</td>
<td>1 (1,1)</td>
</tr>
<tr>
<td>Employment Status N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Unemployed Due to Current Illness</td>
<td>39 (44.3)</td>
<td>41 (53.9)</td>
</tr>
<tr>
<td>Baseline Symptom Severity [Mean (SD)]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• MADRS Total Score</td>
<td>32.1 (5.8)</td>
<td>32.9 (6.0)</td>
</tr>
<tr>
<td>• HAM-D24 Total Score</td>
<td>30.0 (5.0)</td>
<td>30.3 (5.0)</td>
</tr>
<tr>
<td>• HAM-D17 Total Score</td>
<td>22.3 (3.3)</td>
<td>23.0 (3.8)</td>
</tr>
<tr>
<td>• IDS-SR Total Score</td>
<td>41.3 (8.7)</td>
<td>43.0 (10.3)</td>
</tr>
<tr>
<td>• CGI Severity Total Score</td>
<td>4.7 (0.6)</td>
<td>4.7 (0.7)</td>
</tr>
</tbody>
</table>

\(^1P = 0.039;\) all other comparisons not significant \((P > 0.05)\).

**Efficacy Outcomes**

Symptomatic improvement in both studies was measured by the Montgomery-Åsberg Depression Rating Scale (MADRS) and the
HAM-D, with both 24-item and 17-item versions reported separately for the HAM-D. The primary efficacy outcome in the randomized trial was the difference between response and remission to active and sham TMS using the last available total score on the MADRS through week 4 of the acute treatment phase. Secondary outcome measures included the MADRS total score at 6 weeks, total scores on the HAM-D17 and HAM-D24 at 4 and 6 weeks, and categorical endpoints for response and remission using the MADRS, HAM-D17, and HAM-D24 at 4 and 6 weeks. Response was defined as at least 50% reduction from baseline score. Remission was defined by a total score at endpoint of MADRS < 10, HAM-D24 ≤ 10, and HAM-D17 ≤ 7.

**Statistical Methods**

Efficacy analyses in the original randomized controlled trial were performed on the modified intent-to-treat sample of all evaluable patients, defined in the protocol as those with a baseline assessment and at least one postbaseline observation available for analysis. All analyses were conducted in a last observation carried forward (LOCF) manner. A complete description of the prespecified statistical analysis methods used in the original randomized controlled trial are discussed in detail elsewhere (O’Reardon et al., 2007). In the prespecified analysis of covariance (ANCOVA) model used for the overall study population, a treatment (active vs sham TMS) by ATHF subgroup (ATHF = 1 vs ATHF = 2–4) interaction was observed (P = 0.028), therefore a follow up analysis of the relationship between prior treatment resistance and clinical outcome was performed. This analysis demonstrated that the ATHF = 1 subgroup showed a strong, statistically significant, superior benefit for active TMS compared to sham treatment, whereas the remainder of the patients did not. A complete description of this prespecified analysis plan has been described and presented elsewhere (O’Reardon et al. 2007; Lisanby et al., 2008).

Since the open-label study group was not representative of the randomized study population, results for this study were reported as descriptive outcomes (i.e., no a priori hypothesis was tested). There were two potential routes of entry into the open-label study; failure to benefit from either previous active TMS treatment, or previous sham TMS treatment. In the data discussed here, we focus on the results observed in the latter group of patients that transitioned from sham treatment to open-label TMS, since this patient group represents the open-label analog to the 6 week acute treatment endpoint observed in the randomized controlled trial. A complete discussion of the methods of analysis for the open-label extension study and the outcomes in the entire study population are discussed in more detail elsewhere (Avery et al. 2008).
Methods for Comparison of NeuroStar Outcomes to Other Antidepressants

In order to appraise the clinical significance of the outcomes observed in both the randomized controlled trial and the open-label study employing the NeuroStar TMS system, results were compared with the findings reported in several large, peer-reviewed published reference datasets of antidepressant outcomes of pharmaceutical interventions for major depression (Table 2). These data included several large summaries of FDA registration submission databases obtained by those authors under the Freedom of Information Act (Khan et al., 2000, 2001 and 2007; Turner et al., 2008). These datasets provided a comprehensive overview of continuous outcome measures, i.e., mean change from baseline in total depression symptom scores, which are routinely used as the primary outcome measures in regulatory evaluations of new antidepressants. The recent large and comprehensive dataset utilized in the report by Turner and colleagues (2008) also summarized standardized effect sizes for these outcomes, which enables a reasonable statistical comparison of treatment effects across various studies. These publications are authoritative because they include the most exhaustive review of antidepressant clinical trial data available, they contain both positive and negative studies, and include results of all otherwise unpublished studies conducted by the relevant pharmaceutical companies. Also, these summaries are contemporary to the TMS results reported here and therefore similar study design approaches and identical depression rating scale outcomes were used. These data sources typically do not contain information on categorical clinical outcomes (e.g., response or remission rates), since these are traditionally used as secondary outcome measures in regulatory review of new antidepressants. Therefore, in order to accomplish a comparison of categorical outcomes, we used another published, large reference dataset of a comprehensive series of studies of bupropion where such categorical outcomes were described (Thase et al. 2005). Similar to the publications noted above for continuous outcomes, this latter report is informative since the author had full access to the entire registration dossier for this pharmaceutical agent and so was able to summarize information on both positive and negative clinical trial results. For two of the more recently approved antidepressants, duloxetine and selegiline (transdermal patch), the sponsors’ peer-reviewed primary data publications were utilized as these appeared to represent nearly exhaustive summaries of their respective development programs (Detke, 2002a, 2002b; Goldstein, 2002, 2004; Perahia, 2006; Feiger, 2006; Bodkin, 2002).

The majority of the reference comparison datasets that we have used consist of studies performed in patients preferentially recruited based on the prior likelihood of response to treatment, i.e., patients...
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>TYPE OF SOURCE</th>
<th>ANTIDEPRESSANT MEDICATIONS</th>
<th>NUMBER OF STUDIES REPORTED</th>
<th>OVERALL SAMPLE SIZE FOR EACH MEDICATION REPORTED</th>
<th>CLINICAL OUTCOME REPORTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan et al. (2000)</td>
<td>Aggregated FDA Registration Dossiers</td>
<td>Fluoxetine</td>
<td>4</td>
<td>958</td>
<td>Change from Baseline in HAM-D17 Total Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sertraline</td>
<td>3</td>
<td>993</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paroxetine</td>
<td>11</td>
<td>1075</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venlafaxine (IR)</td>
<td>6</td>
<td>1365</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nefazodone</td>
<td>8</td>
<td>1748</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mirtazapine</td>
<td>10</td>
<td>1240</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bupropion</td>
<td>3</td>
<td>1352</td>
<td></td>
</tr>
<tr>
<td>Khan et al. (2001)</td>
<td>Aggregated FDA Registration Dossiers</td>
<td>Venlafaxine (XR)</td>
<td>3</td>
<td>610</td>
<td>Change from Baseline in HAM-D17 Total Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Citalopram</td>
<td>4</td>
<td>1050</td>
<td></td>
</tr>
<tr>
<td>Khan et al. (2007)</td>
<td>Aggregated FDA Registration Dossiers</td>
<td>Escitalopram</td>
<td>4</td>
<td>1312</td>
<td>Change from Baseline in HAM-D17 Total Score</td>
</tr>
<tr>
<td>Turner et al. (2008)</td>
<td>Aggregated FDA Registration Dossiers</td>
<td>Bupropion (SR)</td>
<td>3</td>
<td>857</td>
<td>Standardized Effect Size (Hedge’s g)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Citalopram</td>
<td>5</td>
<td>971</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Duloxetine</td>
<td>8</td>
<td>1880</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Escitalopram</td>
<td>4</td>
<td>1295</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fluoxetine</td>
<td>5</td>
<td>1142</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mirtazapine</td>
<td>10</td>
<td>977</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nefazodone</td>
<td>6</td>
<td>818</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paroxetine</td>
<td>16</td>
<td>1317</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>TYPE OF SOURCE</th>
<th>ANTIDEPRESSANT MEDICATIONS REPORTED</th>
<th>NUMBER OF STUDIES REPORTED FOR EACH MEDICATION</th>
<th>OVERALL SAMPLE SIZE FOR STUDIES</th>
<th>CLINICAL OUTCOME REPORTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thase et al. (2005)</td>
<td>Aggregated Registration Dossier from Manufacturer Database</td>
<td>Paroxetine (CR)</td>
<td>3</td>
<td>235</td>
<td>Categorical Outcome (Response Status) HAM-D17 Total Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sertraline</td>
<td>5</td>
<td>1017</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venlafaxine (IR)</td>
<td>6</td>
<td>1070</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venlafaxine (XR)</td>
<td>3</td>
<td>605</td>
<td></td>
</tr>
<tr>
<td>Detke et al. (2002a); Detke et al. (2002b); Goldstein et al. (2002); Goldstein et al. (2004); Perahia et al. (2006)</td>
<td>Published Registration Clinical Trials</td>
<td>Bupropion (IR,SR)</td>
<td>7</td>
<td>1244</td>
<td>Change from Baseline in HAM-D17 Total Score</td>
</tr>
<tr>
<td>Bodkin et al. (2002); Feiger et al. (2006)</td>
<td>Published Registration Clinical Trials</td>
<td>Selegiline (Transdermal Patch)</td>
<td>2</td>
<td>442</td>
<td>Change from Baseline in HAM-D17 Total Score</td>
</tr>
<tr>
<td>Shelton et al. (2005); Corya et al. (2006); Thase et al. (2007)</td>
<td>Published Registration Clinical Trials</td>
<td>Olanzapine/Fluoxetine (Combination Antidepressant Treatment)</td>
<td>4</td>
<td>1588</td>
<td>Change from Baseline in MADRS Total Score</td>
</tr>
<tr>
<td>Berman et al. (2007); Marcus et al. (2007)</td>
<td>Published Registration Clinical Trials</td>
<td>Aripiprazole (Adjunctive Antidepressant Treatment)</td>
<td>2</td>
<td>743</td>
<td>Change from Baseline in MADRS Total Score</td>
</tr>
</tbody>
</table>
with historical evidence of resistance to antidepressant treatment were typically excluded from such studies. There are, unfortunately, fewer comparative datasets of large, multisite clinical trials conducted in patient populations where the inclusion criteria specifically and preferentially selected subjects based on demonstrated evidence of their failure to benefit from prior treatment, as was done for the NeuroStar TMS study results reported here. The two largest reference sources of such studies in treatment-resistant patient populations are the recently completed reports for the development of the combination of olanzapine and fluoxetine (Shelton et al. 2005; Corya et al. 2006; Thase et al. 2007) and studies of the recently FDA-approved adjunctive treatment with aripiprazole (Berman, 2007; Marcus, 2007).

In a similar manner, in order to assess the clinical significance of the open-label study results, we compared those outcomes to the results of the recently reported, large multisite STAR*D study (Trivedi et al. 2006a, 2006b; Rush et al. 2006a, 2006b; Fava et al. 2006; Nierenberg et al. 2006; McGrath et al. 2006). STAR*D sequentially exposed patients to progressively more complex pharmacotherapy regimens following repeated failures to benefit from the preceding antidepressant treatment assignment. While there are differences in the design of the STAR*D and the NeuroStar TMS clinical studies, we believe the comparison is reasonable for several reasons. In the STAR*D study, patients entering at Level 1 could not have previously been treated with any of the agents used in the first two treatment levels of STAR*D; therefore, these patients represented a relatively treatment-naïve patient population at study entry. Patients were subsequently permitted to enter Level 2 in the STAR*D study if they had failed to achieve remission after 12 to 14 weeks of treatment in Level 1. In other words, at entry to Level 2 they had shown clear evidence of failure to benefit from one prior treatment at an effective dose and duration in their current illness episode. Therefore, the outcomes in the NeuroStar TMS ATHF = 1 population were compared only with those of the STAR*D study Level 2, since these two different groups provide the closest possible parallel for prior treatment history across the two studies.

Results from the NeuroStar TMS Therapy clinical trials were compared with those of antidepressant medication trials (both randomized controlled and open label studies) that variably employed several different clinical outcomes and statistical measures. For the randomized controlled trials, the primary outcome measure in virtually all the antidepressant medication registration studies reported was the difference in the mean change score from baseline for the 17-item HAM-D, compared between the two treatment groups at study endpoint. Interestingly, for all of the randomized controlled trials of
antidepressant medications that enrolled treatment-resistant patients, the MADRS was used as the primary outcome measure. Since the NeuroStar studies obtained both measures, we report the results for the relevant rating scale to provide the most accurate comparison across studies. A commonly used secondary outcome measure in most studies is the categorical outcome measure of response rate, again compared between the two treatment groups at study endpoint. Calculation of effect size offers a more general statistical method of comparing results across different studies (Kraemer and Kupfer, 2005). Two commonly used methods to calculate effect size are the standardized effect size (e.g., Hedge’s g, which compensates for bias in circumstances of unequal sample sizes) for continuous outcome measures and the success rate difference (SRD) and the number needed to treat (NNT, or 1/SRD) for categorical outcomes. To compute values for Hedge’s g we followed the method previously described by Turner and colleagues in their recent report (Turner et al. 2008).

Because the STAR*D trials used remission (as measured by the HAM-D) at study endpoint, this measure alone is used to compare outcomes between STAR*D Level 2 and Neurostar TMS therapy trial during open-label treatment.

RESULTS

Efficacy Outcomes

TMS Therapy Randomized Controlled Study

In patients who had failed to benefit from one prior antidepressant treatment at minimal effective dose and duration in their current illness episode, there was a statistically significantly superior outcome on the primary outcome measure, MADRS total score change from baseline after 4 weeks of treatment with NeuroStar TMS (Active TMS LS mean = -7.1 points vs Sham TMS LS mean = -2.1 points, P = 0.0006) (Fig. 1a). A similar pattern of efficacy for active TMS compared to sham TMS was also observed on the secondary outcome measures for both the HAM-D24 (P = 0.0006) and HAM-D17 (P = 0.0011) (Fig. 1b, Fig. 1c). As is shown in the Figures, statistical significance persisted on all 3 continuous outcome measures through the secondary time point of week 6.

Pre-specified secondary outcomes also included several factor scores for the HAM-D, including the Core Depression (Sum of HAM-D Items 1,2,3,7,8), Maier (Sum of HAM-D Items 1,2,7,8,9,10), Gibbons (Sum of HAM-D Items 1,2,3,7,9,10,11,14), Anxiety/Somatization (Sum of HAM-D Items 10,11,12,13,15,17), Psychomotor Retardation
**FIGURE 1A**

**Randomized Controlled Trial: Baseline to Endpoint Change in MADRS Total Score**

<table>
<thead>
<tr>
<th>Week</th>
<th>NeuroStar TMS (N = 88)</th>
<th>Sham TMS (N = 76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Week 2</td>
<td>-2</td>
<td>-3.5</td>
</tr>
<tr>
<td>Week 4</td>
<td>-5.8</td>
<td>-7.3</td>
</tr>
<tr>
<td>Week 6</td>
<td>-7.2</td>
<td>-8</td>
</tr>
</tbody>
</table>

Note: P-value represents the change from baseline contrast between treatment group comparison using ANCOVA model, change from baseline = baseline MADRS score, center and treatment.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

**FIGURE 1B**

**Randomized Controlled Trial: Baseline to Endpoint Change in HAM-D24 Total Score**

<table>
<thead>
<tr>
<th>Week</th>
<th>NeuroStar TMS (N = 88)</th>
<th>Sham TMS (N = 76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Week 2</td>
<td>-2</td>
<td>-3.5</td>
</tr>
<tr>
<td>Week 4</td>
<td>-5.8</td>
<td>-7.3</td>
</tr>
<tr>
<td>Week 6</td>
<td>-7.2</td>
<td>-8</td>
</tr>
</tbody>
</table>

Note: P-value represents the change from baseline contrast between treatment group comparison using ANCOVA model, change from baseline = baseline HAM-D24 score, center and treatment.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
(Sum of HAM-D Items 1,7,8,14), and Sleep (Sum of HAM-D Items 4,5,6) subscale factors. As shown in Table 3, there was statistically significant superiority of active TMS on all subscale outcomes with the exception of the Sleep factor score.

Categorical outcomes showed statistically significant superiority of active TMS compared to sham for all 3 depression rating scales for response rate after 4 weeks of acute treatment, with persistence and growth in this rate through week 6 (Figs. 2a, 2b, 2c). The more stringent outcome of remission showed statistically significant superiority for active TMS compared to sham at 6 weeks for 2 of the 3 depression symptom scales (Figs. 3a, 3b, 3c).

TMS Therapy Open-Label Extension Study

The primary effectiveness outcome was the change from baseline (obtained as the last observed value at exit from the randomized study) in MADRS total score. These data are summarized in Figure 4a, which displays the mean change from baseline for the Sham to TMS treatment group through the end of the acute treatment phase (week 6).

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

Note: P-value represents the change from baseline contrast between treatment group comparison using ANCOVA model, change from baseline = baseline HAM-D17 score, center and treatment.

* $P<0.05$, ** $P<0.01$, *** $P<0.001$.

Protocol-specified secondary efficacy outcomes included change from baseline on the HAM-D24 and HAM-D17 total scores and showed
### TABLE 3

**Randomized Controlled Trial Secondary Outcome Measures: Hamilton Depression Rating Scale Factor Scores**

<table>
<thead>
<tr>
<th>HAM-D Factor Score</th>
<th>Neurostar TMS (N = 88)</th>
<th>Sham TMS (N = 76)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BASELINE (SD)</td>
<td>CHANGE FROM BASELINE LS MEAN (SD)</td>
</tr>
<tr>
<td>Core Depression Factor</td>
<td>9.2 (2.1)</td>
<td>-2.1 (3.0)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 1,2,3,7,8)</td>
<td>0.0043</td>
<td>0.0103</td>
</tr>
<tr>
<td>Maier Depression Factor</td>
<td>11.2 (1.9)</td>
<td>-2.8 (3.3)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 1,2,7,8,9,10)</td>
<td>0.0004</td>
<td>0.0008</td>
</tr>
<tr>
<td>Gibbons Depression Factor</td>
<td>14.4 (2.6)</td>
<td>-3.4 (4.2)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 1,2,3,7,9,10,11,14)</td>
<td>0.0007</td>
<td>0.0028</td>
</tr>
<tr>
<td>Anxiety/Somatization Factor</td>
<td>7.3 (2.0)</td>
<td>-2.0 (2.3)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 10,11,12,13,15,17)</td>
<td>0.0075</td>
<td>0.0082</td>
</tr>
<tr>
<td>Psychomotor Retardation Factor</td>
<td>7.7 (1.7)</td>
<td>-1.9 (2.4)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 1,7,8,14)</td>
<td>0.0003</td>
<td>0.0030</td>
</tr>
<tr>
<td>Sleep Factor</td>
<td>3.6 (1.8)</td>
<td>-1.0 (2.1)</td>
</tr>
<tr>
<td>(Sum of HAM-D Items 4,5,6)</td>
<td>0.1262</td>
<td>0.1648</td>
</tr>
</tbody>
</table>

Notes: P-value represents the change from baseline contrast between treatment group comparison using ANCOVA model, change from baseline = baseline HAM-D factor score, center and treatment. All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
CLINICAL SIGNIFICANCE OF TRANSCRANIAL MAGNETIC STIMULATION

FIGURE 2A

RANDOMIZED CONTROLLED TRIAL: MADRS RESPONSE RATES

<table>
<thead>
<tr>
<th>Week</th>
<th>NeuroStar TMS (N = 88)</th>
<th>Sham TMS (N = 76)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2</td>
<td>10.2</td>
<td>4.0</td>
<td>0.0471</td>
</tr>
<tr>
<td>Week 4</td>
<td>20.5</td>
<td>9.2</td>
<td>0.0083</td>
</tr>
<tr>
<td>Week 6</td>
<td>25.0</td>
<td>9.2</td>
<td>0.0026</td>
</tr>
</tbody>
</table>

Note: Responder defined as ≥50% reduction in total score compared to baseline. P-value calculated using a logistic regression model: Responder = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

FIGURE 2B

RANDOMIZED CONTROLLED TRIAL: HAM-D24 RESPONSE RATES

<table>
<thead>
<tr>
<th>Week</th>
<th>NeuroStar TMS (N = 88)</th>
<th>Sham TMS (N = 76)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2</td>
<td>11.4</td>
<td>5.3</td>
<td>0.0670</td>
</tr>
<tr>
<td>Week 4</td>
<td>21.6</td>
<td>9.2</td>
<td>0.0047</td>
</tr>
<tr>
<td>Week 6</td>
<td>25.0</td>
<td>13.2</td>
<td>0.0139</td>
</tr>
</tbody>
</table>

Note: Responder defined as ≥50% reduction in total score compared to baseline. P-value calculated using a logistic regression model: Responder = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
CLINICAL SIGNIFICANCE OF TRANSCRANIAL MAGNETIC STIMULATION

FIGURE 2C

RANDOMIZED CONTROLLED TRIAL: HAM-D17 RESPONSE RATES

![Bar chart showing response rates for NeuroStar TMS and Sham TMS across weeks 2, 4, and 6.]

- **P = 0.0042**
- **P = 0.0025**
- **P = 0.0012**

**Note:** Responder defined as ≥50% reduction in total score compared to baseline. P-value calculated using a logistic regression model: Responder = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

FIGURE 3A

RANDOMIZED CONTROLLED TRIAL: MADRS REMISSION RATES

![Bar chart showing remission rates for NeuroStar TMS and Sham TMS across weeks 2, 4, and 6.]

- **P = 0.0042**
- **P = 0.0025**
- **P = 0.0120**

**Note:** Remission defined as MADRS total score <10 at time point of observation. P-value calculated using a logistic regression model: Remission = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
CLINICAL SIGNIFICANCE OF TRANSCRANIAL MAGNETIC STIMULATION

RANDOMIZED CONTROLLED TRIAL: HAM-D24 REMISSION RATES

Note: Remission defined as HAM-D24 total score <11 at time point of observation. P-value calculated using a logistic regression model: Remission = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.
All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

RANDOMIZED CONTROLLED TRIAL: HAM-D17 REMISSION RATES

Note: Remission defined as HAM-D17 total score <8 at time point of observation. P-value calculated using a logistic regression model: Remission = Center, Treatment Group.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001.
All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
results consistent with those observed for the primary outcome on the MADRS (Fig. 4b, Fig. 4c).

Clinical improvement was also evident on the categorical outcome measure of response (Figs. 5a, 5b, 5c), for the MADRS, HAM-D24 and HAM-D17, respectively. Finally, clinical improvement was evident on the measure of remission, as defined by a MADRS total score of \( \leq 10 \), HAM-D24 total score of <10, or HAM-D17 total score of \( \leq 7 \) at endpoint, respectively (Figs. 6a, 6b, 6c).

**Summary of Treatment Effects and Effect Sizes**

**Randomized Controlled Studies**

A summary of the mean change score from baseline on the HAM-D17 is shown in Figure 7 for both randomized controlled studies conducted for eleven currently marketed pharmaceutical antidepressants and for the NeuroStar TMS study results. In this summary, the HAM-D17 is used since that scale is the most ubiquitous primary outcome measure in the cited pharmaceutical reference studies.

In a recent comprehensive meta-analysis, Turner and colleagues calculated the standardized effect size for trials of 12 marketed antidepressant agents (12,564 subjects) registered in the FDA clinical trial database.

---

**FIGURE 4A**

**Open-Label Trial: Baseline to Endpoint Change in MADRS Total Score**

Note: P-value represents the change from baseline contrast within treatment group comparison using ANCOVA model, change from baseline = baseline MADRS score and treatment.

\* = \( P < 0.05 \), \( ** = P < 0.01 \), \( *** = P < 0.001 \), \( **** = P < 0.0001 \).

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
**Clinical Significance of Transcranial Magnetic Stimulation**

**Open-Label Trial: Baseline to Endpoint Change in HAM-D24 Total Score**

Note: P-value represents the change from baseline contrast within treatment group comparison using ANCOVA model, change from baseline = baseline HAM-D24 score and treatment.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001, **** = P < 0.0001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.

**Open-Label Trial: Baseline to Endpoint Change in HAM-D17 Total Score**

Note: P-value represents the change from baseline contrast within treatment group comparison using ANCOVA model, change from baseline = baseline HAM-D17 score and treatment.

* = P < 0.05, ** = P < 0.01, *** = P < 0.001, **** = P < 0.0001.

All computations performed on modified intent-to-treat, evaluable population in last-observation-carried forward (LOCF) manner.
CLINICAL SIGNIFICANCE OF TRANSCRANIAL MAGNETIC STIMULATION

**FIGURE 5A**

OPEN-LABEL TRIAL: MADRS RESPONSE RATES

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.3</td>
<td>32.6</td>
<td>53.5</td>
</tr>
</tbody>
</table>

(95%CI: 6.8 – 30.7) (95%CI: 19.1 – 48.5) (95%CI: 37.7 – 68.8)

- **NeuroStar Open-Label TMS (N = 43)**

Note: Responder defined as ≥50% reduction in total score compared to baseline. Response rate is calculated as the ratio of the actual number of responders at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.

**FIGURE 5B**

OPEN-LABEL TRIAL: HAM-D24 RESPONSE RATES

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.3</td>
<td>37.2</td>
<td>53.5</td>
</tr>
</tbody>
</table>

(95%CI: 11.8 – 38.6) (95%CI: 23.0 – 53.3) (95%CI: 37.7 – 68.8)

- **NeuroStar Open-Label TMS (N = 43)**

Note: Responder defined as ≥50% reduction in total score compared to baseline. Response rate is calculated as the ratio of the actual number of responders at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.

Psychopharmacology Bulletin: Vol. 42 - No. 2
**OPEN-LABEL TRIAL: HAM-D17 RESPONSE RATES**

Week 2  | Week 4  | Week 6  
---|---|---
14.0 (95%CI: 5.3 – 27.9)  | 34.9 (95%CI: 21.0 – 50.9)  | 48.8 (95%CI: 33.3 – 64.5)

**OPEN-LABEL TRIAL: MADRS REMISSION RATES**

Week 2  | Week 4  | Week 6  
---|---|---
7.0 (95%CI: 1.5 – 19.1)  | 9.3 (95%CI: 2.6 – 22.1)  | 23.3 (95%CI: 11.8 – 38.6)

Note: Remission defined as MADRS total score <10 at time point of observation. Remission rate is calculated as the ratio of the actual number of remitters at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.

Note: Responder defined as ≥50% reduction in total score compared to baseline. Response rate is calculated as the ratio of the actual number of responders at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.
### OPEN-LABEL TRIAL: HAM-D24 REMISSION RATES

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>18.6</td>
<td>32.6</td>
</tr>
</tbody>
</table>

(95%CI: 1.5 – 19.1)  
(95%CI: 8.4 – 33.4)  
(95%CI: 19.1 – 48.5)

Note: Remission defined as HAM-D24 total score <11 at time point of observation. Remission rate is calculated as the ratio of the actual number of remitters at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.

### OPEN-LABEL TRIAL: HAM-D17 REMISSION RATES

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>14.0</td>
<td>25.6</td>
</tr>
</tbody>
</table>

(95%CI: 1.5 – 19.1)  
(95%CI: 5.3 – 27.9)  
(95%CI: 13.5 – 41.2)

Note: Remission defined as HAM-D17 total score <8 at time point of observation. Remission rate is calculated as the ratio of the actual number of remitters at the specified time point divided by the total number of subjects (N = 43). 95% confidence intervals are shown in parentheses above each bar.
Using Hedge’s g, they reported that the overall mean weighted effect size across all antidepressants was 0.31 for the HAM-D17. In the randomized controlled trial for the NeuroStar TMS system, the value of Hedge’s g for the HAM-D17 is 0.52. These data are shown in Figure 8a.

There have been two pharmaceutical antidepressant treatments (i.e., olanzapine/fluoxetine combination and aripiprazole augmentation) that have been studied in randomized controlled studies in patient populations which, like the NeuroStar TMS study, were specifically selected for their resistance to prior antidepressant treatment. Figure 7b shows the summary of mean change scores from baseline for these treatments. In this instance, the data are presented for the MADRS, since that scale was used as the primary outcome measure in both treatments and results for the HAM-D were not reported in those studies.

Using the same statistical method as reported above for antidepressants studied in non-treatment-resistant patients, the mean effect size by Hedge’s g for the olanzapine/fluoxetine combination is 0.20 and
The mean effect size for the aripiprazole augmentation studies is 0.35 in these treatment-resistant patient populations. These data are shown in Figure 8b.

The categorical outcome of response rate is shown in Figure 9, studied for both treatment-responsive and treatment-resistant patient populations using the data from Thase and colleagues (2005) and also from the two published aripiprazole augmentation studies (Berman et al. 2007; Marcus et al. 2008). Data from Thase et al. 2005 are shown on the left side of this figure, summarizing the response outcomes for a large meta-analysis of randomized controlled studies from a pharmaceutical database for the antidepressant bupropion. These studies were conducted in patient populations preferentially selected for their responsiveness to prior antidepressant treatment. In the center of this figure the response outcomes for the two aripiprazole augmentation studies are shown, which are, as noted above, results from randomized controlled trials in treatment-resistant patients. For completeness of display, categorical response
FIGURE 8A

**SUMMARY OF EFFECT SIZES (HEDGE’S G) IN RANDOMIZED CONTROLLED TRIALS OF ANTIDEPRESSANT MEDICATIONS AND FOR NEUROSTAR TMS THERAPY**

Pooled Effect Size for Antidepressant Medication RCTs

0.31 (95% CI: 0.26 – 0.36)

Effect Size for NeuroStar TMS Randomized Controlled Trial

0.52 (95% CI: 0.21 – 0.83)

Note: Please refer to text for details. Antidepressant medication reference source from Turner et al. (2008). Method for standardized effect size computation (Hedge’s g) as reported in Turner et al. (2008).

FIGURE 8B

**SUMMARY OF EFFECT SIZES (HEDGE’S G) IN RANDOMIZED CONTROLLED TRIALS OF ANTIDEPRESSANT MEDICATIONS AND FOR NEUROSTAR TMS THERAPY (TREATMENT-RESISTANT STUDY POPULATIONS)**

Olanz/Fluox (Thase, 2007)

0.33 (95% CI: 0.13 – 0.52)

Olanz/Fluox (Corya, 2006)

0.27 (95% CI: – 0.01 – 0.55)

Olanz/Fluox (Shelton, 2005)

0 (95% CI: – 0.21 – 0.25)

Aripiprazole (Marcus, 2008)

0.34 (95% CI: 0.14 – 0.54)

Aripiprazole (Berman, 2007)

0.35 (95% CI: 0.14 – 0.56)

NeuroStar TMS Therapy

0.52 (95% CI: 0.21 – 0.83)

Note: Please refer to text for details. Antidepressant medication reference sources from Shelton et al. (2005); Corya et al. (2006); Thase et al. (2007); Berman et al. (2007); Marcus et al. (2008). Method for standardized effect size computation (Hedge’s g) as reported in Turner et al. (2008).
outcomes for NeuroStar TMS are shown for both the HAM-D17 outcome and for the MADRS outcome.

As noted above, number needed to treat, or NNT, is a clinically meaningful representation of the effect size for categorical outcome measures. The treatment effects reported in Figure 9 and summarized above can be reported as NNT as shown in Table 4 for the cited studies.

Open-Label Studies

Results from the 6-week open-label study for the NeuroStar TMS system represent an open-label confirmation of the treatment results observed in the active TMS group in the 6-week randomized controlled study. Because the open-label group, by protocol definition, had failed to receive benefit from one adequate antidepressant treatment (verified by the ATHF) in the current episode at the time of entry into the randomized controlled trial, outcomes in these patients correspond most closely to those seen in the patients treated with a monotherapy antidepressant medication at Level 2 of the STAR*D trial. The HAM-D17 remission rate, defined using an endpoint total score of <8, was 26% in the NeuroStar TMS Therapy group, which contrasts with a mean remission rate of 21.2% reported among the Level 2 antidepressant medication monotherapy options offered in the STAR*D study (Rush et al. 2006a).
### TABLE 4

Effect Size Estimates [Number Needed to Treat (NNT)] Based on Responder Rate in Randomized Controlled Trials of Antidepressant Treatments: NeuroStar TMS Therapy vs Treatment-Responsive and Treatment-Resistant Reference Datasets

<table>
<thead>
<tr>
<th>Responder Rate</th>
<th>WEEK 4</th>
<th>WEEK 6</th>
<th>Treatment-Responsive Reference Data</th>
<th>Treatment-Resistant Reference Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Treatment Condition</td>
<td>20.5%</td>
<td>25.0%</td>
<td>25.0%</td>
<td>27.3%</td>
</tr>
<tr>
<td>Placebo/Sham Treatment Condition</td>
<td>9.2%</td>
<td>10.5%</td>
<td>9.2%</td>
<td>10.5%</td>
</tr>
<tr>
<td>Number Needed to Treat (NNT)</td>
<td>8.8</td>
<td>6.9</td>
<td>6.3</td>
<td>6.0</td>
</tr>
</tbody>
</table>
DISCUSSION

There are few evidence-based treatment options that have been specifically studied in patients with major depression who have been rigorously demonstrated to have failed to benefit from prior antidepressant treatment. For a majority of patients, when the initial treatment option fails, the next steps are largely guided by the treating clinician’s anecdotal experience with prior treatments in other patients. Increasing the number of treatment options available that are based on randomized controlled trial evidence of clinical benefit is critical for establishing a more informed treatment algorithm for difficult-to-treat depression.

In a patient population carefully characterized by their failure to benefit from prior antidepressant medication treatment, we have shown that NeuroStar TMS Therapy provides statistically significant clinical improvement in the symptoms of depression. We have also specifically established that this clinical improvement depends upon the patient’s prior antidepressant treatment history, with the greatest benefit evident in patients who, in the current illness episode, have failed to receive satisfactory improvement from one antidepressant medication given at an effective dose for adequate duration. These observations are consistent with previous literature that confirms an inverse relationship between prior treatment failure and future treatment response.

Any new treatment must be considered both in terms of the adequacy of the evidence that the null hypothesis of no effect is rejected with a sufficient level of statistical confidence and also in terms of the clinical significance of the observed effect. In a recent review, Kraemer and Kupfer (2005) highlighted the importance of this concept and have argued that an examination of the size of treatment effects, through the use of mathematically appropriate measures of effect size, is critical to an understanding of the clinical significance or medical consequence of any statistically valid observation in a randomized controlled trial.

Over the past few decades, the development of new antidepressants often lacked sufficient prior evidence to establish the minimum criterion standard to decide whether the threshold of a clinically significant treatment effect had been reached. However, there is now a large, well-established literature of randomized controlled clinical trials of contemporary antidepressants that can be used to retrospectively address this question. The clinical effects of contemporary antidepressants, all of which are now in routine clinical use, are accepted by the clinical community at large as having substantial and clinically meaningful utility in the real world. Therefore we can have increased confidence that the effect sizes observed in the earlier randomized
clinical trial literature identify a minimum mathematical threshold for
determining the clinically interesting significance of any new antide-
pressant treatment.

It should be cautioned, though, that the clinical development literature
for antidepressant medications is still composed largely of outcomes
determined in patient populations specifically selected for their likelihood
of response to treatment. They were not selected because of their past his-
tory of demonstrated failure to benefit from prior treatment, as was done
for the NeuroStar TMS studies. In this report, we present the currently
available evidence from a from a database of clinical trials of treatment-
resistant depression. These studies still comprise, however, a minority of
the published randomized clinical trial literature for antidepressant med-
ications. Nevertheless, we believe meaningful interpretations can still be
drawn from the method of comparison that is presented here, because of
the presence of within-study control conditions that allow a calculation of
effect sizes. In the data we have presented here, NeuroStar TMS Therapy
shows a treatment effect that meets or exceeds clinically meaningful
thresholds of effect size when considered relative to evidence from stud-
ies of pharmaceutical antidepressants in either more treatment-responsive
patient populations or more treatment-resistant patient samples.

The data presented in this report clearly establish a role for NeuroStar
TMS Therapy as a treatment option for use in patients with demon-
strated resistance to initial antidepressant treatment. What constitutes
sufficient prior antidepressant treatment? In the randomized clinical
trial, a majority of patients within the “one adequate treatment failure”
population had experienced at least one prior episode of depression.
Within the current episode of illness alone, a majority of these patients
had also experienced multiple treatment attempts, but only one of these
attempts had reached an effective dose and duration of exposure as
determined by the ATHF. The median number of treatment attempts
in this episode was four and the range extended from one to as many as
twenty-three treatment attempts. This type of complexity in treatment
history is consistent with the scientific literature and with clinical expe-
rience in patients who fail to benefit from initial treatment. Studies
using the ATHF have previously shown that the majority of treatment
attempts fall short of their idealized dose and duration goal, usually
because of clinical issues with tolerability or adverse events. In general,
patients experience 3 to 5 treatment attempts for every one that reach-
es an appropriate level of treatment adequacy (Sackeim, 2001; Prudic
et al. 2004).

This complex interaction between treatment resistance and treatment
intolerance takes on additional relevance in treatment planning given the
fact that clinical studies support the view that adherence to treatments
declines substantially as treatment resistance increases (Papakostas, 2008; Fava et al., 2006; McGrath et al., 2006; Nierenberg et al., 2006; Rush et al., 2006b; Trivedi et al., 2006a). In the STAR*D study, which represents the largest reference dataset available that speaks to this issue, discontinuation rates due to treatment intolerance or adverse events rose steadily across the sequenced Levels [i.e., 8.6% at Level 1, 23.1% (range: 21.0% to 27.2%) at Level 2, 35.2% (range: 34.2% to 36.2%) at Level 3 and 32.1% (range: 21.6% to 41.4%) at Level 4]. From this perspective, TMS represents a clinically meaningful difference for patients who fail to benefit from initial pharmacotherapy, due to the absence of systemic adverse events during treatment with TMS and the low rate of discontinuation due to adverse events (less than 5% during acute treatment) (O’Reardon et al. 2007; Janicak et al. 2008).

There are limitations to the analyses performed here. While we have made an effort to organize the available clinical studies by the level of prior treatment resistance of the study populations, it is clear that different methodologies have been used to characterize treatment resistance across the various studies. Whether these differences in assessment had an impact on the assessment of clinical outcome is not clear. We believe, however, that the impact of any variation in assessment method is mitigated by the large sample sizes of the reference data sets and also by the consistency of the general observation in the clinical literature which supports the view that increasing gradations of treatment resistance predict increasingly reduced odds of responding to a subsequent treatment attempt.

The safety and efficacy of NeuroStar TMS Therapy as a treatment option for patients with pharmacoresistant depression is well-established by the studies reviewed here. These data are important since they represent a clear, prospectively-assembled, evidence-based data set describing statistically and clinically significant efficacy in a population of patients with a defined level of treatment resistance. This information will be helpful for treatment-planning decisions in patients who have failed to receive satisfactory benefit from initial antidepressant treatment.

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

DISCLOSURES
Dr. Demitrack is an employee of and owns stock options in Neuronetics, Inc and reports stock ownership in Eli Lilly and Company and Wyeth Pharmaceuticals, Inc.

CONTRIBUTORS

Drs. Demitrack and Thase have both contributed to the composition and editing of this manuscript. Both authors had access to the data and statistical analyses presented in this manuscript. Both authors have approved the final version of the manuscript.

REFERENCES


