Durability of antidepressant response to vagus nerve stimulation (VNS™)

Harold A. Sackeim, Stephen K. Brannan, A. John Rush, Mark S. George, Lauren B. Marangell, and John Allen

1 Department of Biological Psychiatry, New York State Psychiatric Institute, New York, NY, USA
2 Departments of Psychiatry and Radiology, Columbia University, New York, NY, USA
3 Cyberonics, Inc., Houston, TX
4 Department of Psychiatry, University of Texas, Southwestern Medical Center, Dallas, TX, USA
5 Departments of Psychiatry and Radiology, Medical University of South Carolina, Charleston, SC, USA
6 Department of Psychiatry, Baylor College of Medicine, Houston, TX
7 South Central Mental Illness Research Educational and Clinical Center, Houston, TX

Abstract

This study characterized the durability of improvement in patients who responded early or late while receiving vagus nerve stimulation (VNS). In both a pilot and pivotal study, patients were identified who had at least a 50% reduction in symptom scores 3 months (early responders) or 12 months (late responders) after starting VNS. Probabilities were determined for maintenance of response at 12-month and 24-month time-points. Consistency of improvement throughout the 24-month study period was evaluated, testing for change in serial depression ratings. In the pilot study, 30.5%, 23.7% and 45.8% were early responders, later responders, and non-responders, respectively. These rates were 14.6%, 19.5%, and 65.9% in the pivotal trial. The potential confound of alterations in antidepressant treatment was examined in the pivotal trial. The potential confound of alterations in antidepressant treatment was examined in the pilot trial. In the pilot study, 72.2% and 61.1% of early responders (n = 18) were responders at 12 and 24 months, respectively; 78.8% of late responders (n = 14) were responders at 24 months. In the pivotal trial, early responders (n = 30), 63.3% and 76.7% maintained response at 12 and 24 months, respectively; late responders (n = 40), 65.0% maintained response at 24 months. Early and late responders had fewer changes in medication than non-responders across the pivotal study period. In both studies, analyses of serial depression ratings showed stable improvement in early and late responders. These samples had exceptional levels of chronicity and treatment resistance. Yet patients who showed substantial clinical benefit maintained the improvement at remarkably high rates. This durability of benefit was not attributable to alterations in other treatments.

Introduction

Vagus nerve stimulation (VNS) has shown efficacy in treatment-resistant major depression (TRD). In a pilot study, 30% of 59 participants met response criteria after 3 months of stimulation (Sackeim et al., 2001a). A subsequent, randomized, sham-controlled trial failed to distinguish active and sham VNS conditions in the 3-month primary outcome (Rush et al., 2005a). However, including sham participants who received active VNS after the randomized phase, 30% of 205 evaluable participants were responders at 12 months (Rush et al., 2005b). This rate exceeded that of a treatment-as-usual comparison group (George et al., 2005). Since VNS involves the risks resulting from surgical implantation (Rush et al., 2005b; Sackeim et al., 2001a), it is only of value if its clinical benefits persist. This issue is of special importance in TRD. Relapse after remission with ECT appears heightened in TRD (Sackeim et al., 1990, 2000, 2001b). Similarly, some TRD patients may improve with a change in antidepressant...
pharmacotherapy, but rapidly lose benefit (Flint and Rifat, 2001).

As with pharmaco-resistant epilepsy (Kawai et al., 2002; Morris and Mueller, 1999), a substantial number of TRD patients show clinical benefit only after a considerable delay since starting VNS. There is limited information on the durability of benefit. Furthermore, persistent benefit could be due to prophylactic action of VNS or concomitant medications (Nahas et al., 2005). This study examined the long-term outcomes of patients who participated in the pilot (Sackeim et al., 2001a) and pivotal (Rush et al., 2005a, b) studies of VNS in TRD. Three subgroups were formed in each study: early responders, late responders, and non-responders.

This study had three aims. First, descriptive information was derived on the durability of benefit. We calculated conditional probabilities for maintenance of response over 1- and 2-yr periods, and also contrasted symptom scores over time of early and late responders and non-responders. Second, we compared the subgroups in demographic and clinical characteristics to identify features that distinguished the subgroups. Third, to examine causes of persistent clinical benefit, we contrasted the subgroups in the pivotal study in rates of medication change prior to and following onset of response.

Method

Samples

Inclusion and exclusion criteria in the pilot and pivotal studies were similar and the methods have previously been reported (Rush et al., 2005a, b; Sackeim et al., 2001a). In the pilot study, 59 out of 60 patients contributed to efficacy analyses. Patients presented with a chronic and/or recurrent, major depressive episode (unipolar or bipolar) and had a minimum entry score of 20 on the Hamilton Rating Scale for Depression (HRSD_{28}, 28-item; Hamilton, 1967). They received at least two adequate antidepressant trials in the current episode from at least two different classes, based on Antidepressant Treatment History Form (ATHF) criteria (Sackeim, 2001). The pivotal study enrolled 235 patients, of whom 205 patients contributed to efficacy analyses. The inclusion and exclusion criteria were identical to the pilot, except patients could not have had more than six adequate treatment trials in the current episode and had a minimum score of 18 on the HRSD_{24} at entry. The Institutional Review Board at each site approved the study. All patients provided written informed consent.

Study design

The pilot study was an open-label investigation. There was a 2-wk, single-blind, lead-in following implantation. This was followed by a 2-wk parameter adjustment period, during which the amplitude of the VNS pulse was titrated upwards. Stimulation parameters were then kept fixed for 10 wk, and medication regimens were also unchanged. Twelve weeks after implantation (2 wk recovery, 2 wk adjustment, and 8 wk fixed dose) both medication regimens and VNS parameters could be altered as clinically indicated. The pivotal study used a similar design except patients were randomized to active or sham VNS conditions. In the sham group, VNS was not activated until 12 wk post-implantation. However, 21 of 110 sham-treated patients were excluded for HRSD_{24} scores below 18 at randomized phase termination.

In this report, HRSD_{24} scores were the outcome measures analysed for both studies. The Beck Depression Inventory (BDI-II; Beck et al., 1996) and the Inventory for Depressive Symptoms (IDS-SR; Rush et al., 1985) provided a self-report measure in the pilot and pivotal studies, respectively. Each measure was obtained at baseline and quarterly intervals for 24 months.

Statistical analyses

The pilot and pivotal samples were compared in demographic and clinical features using t-tests on continuous measures and χ² analyses for dichotomous variables. Early responders had a reduction in HRSD_{24} scores of at least 50% from baseline at the assessment 3 months (12 wk) after implantation (or crossover for sham patients). Late responders met this criterion at the 12-month, but not the 3-month, assessment. All other patients were termed non-responders.

The probabilities of maintaining response were computed at 12 months and 24 months for early responders. The criterion for maintenance of response at the follow-up time-points was reduced to an improvement of at least 40% relative to baseline. This avoided characterizing a minor decrease (e.g. from 51% to 49%) as loss of benefit. The probability of maintaining response at the 24-month time-point was also determined for late responders.

Although high probabilities of consistent benefit might suggest that the improvement was durable, examination of arbitrary time-points left open the possibility of substantial fluctuation in symptomatic improvement over time. Repeated-measures analyses of variance (ANOVAs) were conducted on raw
HRSD$_{24}$ scores at the nine quarterly assessments, with response grouping (three levels) as a between-subject factor and assessment interval as the repeated-measures factor. The Hundt–Feldt correction for sphericity was applied. The findings were comparable when only patients with complete data were analysed or when the complete samples were examined using the last observation carried forward (LOCF) method for missing data. Only the LOCF findings are reported. In the pivotal trial, 1705 out of 1845 (92.4%) possible HRSD assessments (nine per patient) were completed during the 24-month period.

BDI scores in the pilot study and IDS-SR scores in the pivotal study were substituted for HRSD$_{24}$ scores and the response subgroups were redefined. When the durability analyses were repeated the results were fully consonant with findings using the HRSD$_{24}$. The findings with the two self-report measures are not reported.

To determine whether the three response subgroups differed in demographic and clinical features, one-way ANOVAs were conducted on continuous variables and $\chi^2$ analyses on dichotomous variables. An a-priori focus was on the difference among the subgroups in treatment resistance. To corroborate findings, onset of response was treated as a continuous variable in parametric survival analyses, and event time was defined as the quarter when patients first met the response criterion ($\geq$50% change in HRSD$_{24}$ scores) and the improvement ($\geq$40% change in HRSD$_{24}$ scores) was sustained over at least 6 months (2 quarters). The clinical features that distinguished the response subgroups in the preceding analyses were used as covariates in the survival analyses, which applied a lognormal distribution to the event times.

In the pivotal study, type and dosage of psycho- tropic medications were recorded at each visit. Dosage was scored using ATHF criteria for adequacy (1–5 scale) (Sackeim, 2001). A medication change was recorded when a patient started a new medication or a change in dosage corresponded to an increase of at least 1 point by ATHF criteria. Each patient was scored as having such a change or not during each of the sixteen 45-d intervals that comprised the 24-month study period. A random regression model (RRM) contrasted early responders and non-responders in rates of medication change during the two 45-d intervals before early response and the subsequent fourteen 45-d intervals. Another RRM contrasted late responders and non-responders in rates of medication change during the eight 45-d intervals before and after the designation of late response.

**Results**

**Demographic and clinical characteristics**

The pilot and pivotal samples did not differ in age or the distribution of gender (Table 1). The pilot study had more patients with bipolar disorder and chronic depression. This sample also had a longer duration of current episode, higher baseline HRSD$_{24}$ scores, and had received both more antidepressant treatment trials and more adequate trials (ATHF defined).

**Pilot study**

Figure 1 presents the clinical change from baseline on the HRSD$_{24}$ at the 12- and 24-month assessments for early responders, with the long-term outcomes tracked separately for patients with 50–59%, 60–79%, and ≥80% improvement at the 3-month time-point.

Using the 40% improvement criterion as the threshold for maintaining response, 13 of 18 (72.2%) early responders maintained benefit at 12 months, and 11 of 18 (61.1%) still benefited at 24 months. Of 14 late responders, 11 (78.8%) maintained benefit at 24 months (Figure 2).

Serial HRSD$_{24}$ scores for the early and late responders and non-responders are presented in Figure 3. The repeated-measures ANOVA yielded significant main effects for patient subgroup ($F=17.01$, d.f. = 2, $p<0.0001$) and time-point ($F=32.66$, d.f. = 5.1, 284.6, $p<0.0001$) and a significant interaction ($F=7.06$, d.f. = 10.2, 284.6, $p<0.0001$). Probably reflecting regression to the mean, at the assessment immediately following response designation some worsening was evident in both responder subgroups. Nonetheless, both responder subgroups showed strong maintenance of benefit at all intervals, with little variation in the extent of improvement. From onset of response until the end of follow-up, early responders averaged 61.6±20.6% improvement, late responders averaged 60.8±21.4% improvement, and non-responders averaged 24.5±18.8% improvement.

A one-way ANOVA indicated that the three groups differed in average percent change in HRSD$_{24}$ scores ($F=24.87$, d.f. = 2, 56, $p<0.0001$). Post-hoc comparisons indicated that both responder subgroups differed from non-responders, but not from each other.

The three subgroups did not differ in any of the variables in Table 1 other than history of ECT (early responders 44.4%, late responders 64.3%, non-responders 81.5%; $\chi^2=6.64$, d.f. = 2, $p<0.04$), total number of treatment trials in the current episode (early responders 11.50±4.93, late responders 13.33±6.06, non-responders 19.81±8.36; $F=8.16$,
d.f. = 2, 56, \( p = 0.0008 \)), and number of adequate trials in the current episode (early responders 3.44 ± 1.50, late responders 4.64 ± 2.31, non-responders 5.93 ± 3.09; \( F = 5.31, \) d.f. = 2, 56, \( p < 0.008 \)). These bivariate analyses suggested that treatment resistance was greater among late responders relative to early responders and was even greater among non-responders. When history of ECT and number of adequate trials in the current episode were predictors in a simultaneous logistic regression on response status, ECT history had no effect (\( \chi^2 = 1.99, \) d.f. = 2, \( p = 0.37 \)), but the effect of medication resistance remained significant (\( \chi^2 = 6.01, \) d.f. = 2, \( p < 0.05 \)). In contrast, when both of these variables were used as covariates in the survival analysis on time to sustained benefit, the effect of ECT history was significant (\( \chi^2 = 4.81, \) d.f. = 1, \( p < 0.03 \)), but the number of adequate treatment trials in the current episode was not (\( \chi^2 = 2.48, \) d.f. = 1, \( p = 0.12 \)).

**Pivotal study**

Figure 4 depicts the clinical change on the HRSD_{24}, observed at 12 months and 24 months for early responders. Overall, 19 of 30 (63.3%) early responders maintained benefit at 12 months, and 23 of 30 (76.7%) early responders still benefited at 24 months. Of the 40 late responders, 26 (65.0%) maintained benefit at 24 months (Figure 5).

The repeated-measures ANOVA on serial HRSD_{24} scores yielded significant main effects of patient...
subgroup ($F = 68.97$, d.f. = 2, 202, $p < 0.0001$) and time-point ($F = 60.70$, d.f. = 6.7, 1357.2, $p < 0.0001$) and a significant interaction ($F = 16.29$, d.f. = 13.4, 1357.2, $p < 0.0001$) (Figure 6). Again both responder subgroups showed strong maintenance of benefit at all assessments after the designation of response. Early responders averaged 66.4% improvement at 3 months, and this value ranged from 51.8% to 59.8% at the assessments between 9 and 24 months. Late responders averaged a 65.1% improvement at 12 months, with a range from 45.4% to 49.8% at the assessments between 18 and 24 months. Improvement in non-responders across all assessments ranged from 3.4% (3 months) to 19.0% (21 months). From the designation of response
until the end of follow-up, early responders averaged 54.7 ± 16.1% improvement, late responders averaged 51.3 ± 20.5% improvement, and non-responders averaged 12.9 ± 19.0% improvement over the entire follow-up period. A one-way ANOVA indicated that the three subgroups differed in these changes (F = 102.9, d.f. = 2, 202, p < 0.0001). Post-hoc comparisons indicated that each responder group differed from non-responders.

Unlike the results for the pilot study, bivariate analyses indicated that the three subgroups did not differ on any variable in Table 1, including history of ECT and degree of treatment resistance.

Figure 7 presents the rate of medication change for the subgroups. The RRM for the comparison of early responders and non-responders yielded main effects of response status (F = 7.66, d.f. = 1, 166, p = 0.006) and assessment interval (F = 2.40, d.f. = 15, 2293, p = 0.002), but no interaction (F = 1.25, d.f. = 15, 2293, p = 0.23). Across the 24 months, early responders (12.3 ± 9.6%)...
averaged fewer medication changes than non-responders (19.9 ± 10.4%). The RRM for the late responder vs. non-responder comparison yielded similar effects: a trend for a main effect of response classification \((F = 3.65, \text{d.f.} = 1, 173, \ p = 0.06)\), a main effect of assessment interval \((F = 5.24, \text{d.f.} = 15, 2440, \ p < 0.0001)\), and no interaction \((F = 0.89, \text{d.f.} = 15, 2440, \ p = 0.58)\). Participants with medication changes tended to be represented more among non-responders (19.9 ± 10.4%) than late responders (16.4 ± 9.8%). These findings indicating that medication changes were less frequent among early or late responders than non-responders probably underestimated the magnitude of the effect. Principally due to dropout and loss to follow-up, there were missing values for medication changes at one or more assessment intervals for 3 of 30 (10.0%) early responders, 5 of 40 (12.5%) late responders, and 40 of 135 (29.6%) non-responders \(( \chi^2 = 4.74, \text{d.f.} = 2, p = 0.03)\). For the sixteen 45-d intervals, on average early responders had missing data on 0.63 ± 2.14 occasions, late responders on 0.80 ± 2.22 occasions, and non-responders on 2.19 ± 3.97 occasions \((F = 4.11, \text{d.f.} = 2, 202, p = 0.02)\). Differential dropout due to lack of efficacy probably accounted for this pattern.

As noted, the representation of patients with bipolar disorder was greater in the pilot study than the pivotal study (27.1 vs. 9.8%) (Table 1, \(p = 0.0006)\). However, in both studies the representation of early and late responders and non-responders did not differ for the bipolar and unipolar subgroups. Bipolar patients also did not appear exceptional in durability of benefit. Across the two studies, six of the nine (66.7%) bipolar early responders maintained benefit at both the 12-month and 24-month assessments, and four out of five (80.0%) bipolar late responders showed substantial improvement at the 24-month assessment.

**Discussion**

Most participants who responded during the pilot and pivotal studies showed durable clinical benefit. For early responders, 63% and 72% (pivotal/pilot studies) showed substantial clinical benefit at 12 months, and these rates were 77% and 61% at 24 months. Likewise, 79% and 65% of late responders showed substantial clinical benefit at 24 months. Similarly, analyses of serial symptom scores, whether clinician-rated or self-reported, demonstrated that symptomatic improvement, once achieved, was generally stable throughout follow-up.

The degree of treatment resistance and the delayed onset of response are two exceptional aspects of the findings. The extent of medication resistance, history of previous ECT, and duration of current episode substantially exceeded that reported in studies of ECT (Husain et al., 2004; McCall et al., 2000; Sackeim et al., 1993, 2000), whose primary indication is treatment resistance (APA, 2001).

Treatment-resistance and chronicity are thought to predict poorer outcome in treatment of depressive episodes and higher rates of relapse in responders or remitters (Flint and Rifat, 1997; Khan et al., 1991; Posternak et al., 2002; Prudic et al., 1996; Sackeim et al., 1990). For example, when remission is achieved with ECT the relapse rate in the following 12 months is estimated to be 60–70% among patients who have not responded to one or more adequate medication trials, approximately double that of patients who come to ECT without having received an adequate medication trial in the current episode (Sackeim et al., 1990, 2000). In contrast, while only one-third of patients responded to VNS within 1 yr, approximately two-thirds maintained clinical benefit through 21 months of follow-up.

The second exceptional finding is the delay in acute response. By HRSD\(_{34}\) criteria, the pilot study had 18 early and 14 late responders, and the pivotal study had 30 early and 40 late responders. In the pilot and pivotal studies, the mean time to onset of response (maintained for at least 6 months) was 10.6 ± 4.2 and 10.1 ± 2.2 months, respectively. These values
underestimated the frequency of late response, as they did not include participants who improved after the 12-month assessment. Regardless, if late response was due to VNS, either acting alone or synergistically with medications, it reflects a novel time-course. Given these delayed effects, outcome of a VNS trial for an individual patient may not be fully known until 12–24 months have elapsed. We were unsuccessful in identifying predictors of early or late benefit.

In the absence of a long-term, sham-controlled, randomized trial, one cannot conclude that the acute or sustained effects observed in the pilot and pivotal studies were attributable to VNS. The three major alternatives are that the therapeutic benefits reflected placebo effects, the action of altered medication regimens, or the natural history of the illness. The possibility of placebo effects is heightened by the fact that VNS involves a surgical intervention and frequent early visits for dose adjustment. However, the counter arguments appear determinative. In major depression, placebo effects are seen early and are typically transient (Quitkin et al., 1993, 1998). In contrast, in the pilot and pivotal studies a substantial proportion of participants showed delayed benefit, which was typically sustained. Both studies recruited participants with marked treatment resistance and long duration of current episode, features that would argue against a placebo response.

Whether the clinical benefit seen with VNS could be attributed to medication alterations was partially addressed in this study. The rate of medication changes in early and late responders was lower than in non-responders. Indeed, since the absolute rates of medication change were low in the responder subgroups, it is hard to imagine how a change in medication regimen could have been a significant factor leading to clinical response or sustaining the benefit once achieved. Finally, although the natural history of TRD is not well characterized, the 12-month data from the treatment-as-usual study suggest that spontaneous recovery rates are extremely low (Dunner et al., 2006; George et al., 2005).

The findings indicated that improvement was durable for most participants who showed acute clinical benefit with VNS. It did not appear that onset of benefit or its continuance was due to changes in medication regimens. Furthermore, a substantial number of participants showed delayed improvement. These findings trigger two key issues. First, the delay and durability of action may be related. Treatments that are remarkably quick in onset of antidepressant effects, such as sleep deprivation and ECT, are also characterized by high rates of relapse (Hernandez et al., 2000; Leibenluft and Wehr, 1992). If neuroplastic changes are a necessary component to sustained antidepressant effects, as suggested by the neurogenesis literature (Newton and Duman, 2004; Santarelli et al., 2003), treatments that require longer periods to take effect may be characterized by greater durability due to time-dependent changes in patterns of connectivity that sustain persistence of benefit. Investigating the time-course of neurobiological alterations of VNS relative to other interventions may prove useful.

Second, the findings did not address whether improvement is sustained only when VNS is responsible for acute response or whether VNS can sustain the improvement initially produced by other treatments. The low rate of medication change among the responders in the pivotal study suggested that the acute improvement was probably due to VNS. Many treatment-resistant patients have a history of transient response to ECT or various antidepressant medications. In the pilot and pivotal studies, participants were rarely re-challenged with these treatments precisely because their benefits had been fleeting in the past. However, it is not known whether sustained exposure to VNS results in prophylactic benefit regardless of the intervention that produces initial improvement. If VNS can continue the action not only of itself but also of other interventions that produce acute response, the number of patients who might benefit from VNS would increase substantially.

Acknowledgements

The research reported herein was supported in part by grants from Cyberonics, Inc.

Statement of Interest

This paper reports data from two studies that were sponsored by Cyberonics, Inc. Drs Sackeim, Rush, George, and Marangell served as principal investigators in the first study and were also investigators in the second. Each has served as a consultant and a member of the Depression Advisory Board to Cyberonics, Inc. The other two authors, Dr Brannan and Dr Allen, are employees of Cyberonics, Inc.

References


