Appendix A: General Methodological Principles of Study Design

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member. The critical appraisal of the evidence enables us to determine whether: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve health outcomes for patients. An improved health outcome is one of several considerations in determining whether an item or service is reasonable and necessary.

CMS divides the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the relevance of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention’s risks and benefits.

The issues presented here represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has unique methodological aspects.

1. Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes; and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematical assessment of factors related to outcomes.
- Larger sample sizes in studies to help ensure adequate numbers of patients are enrolled to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can
be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias)
- Co-interventions or provision of care apart from the intervention under evaluation (confounding)
- Differential assessment of outcome (detection bias)
- Occurrence and reporting of patients who do not complete the study (attrition bias)

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

- Randomized controlled trials
- Non-randomized controlled trials
- Prospective cohort studies
- Retrospective case control studies
- Cross-sectional studies
- Surveillance studies (e.g., using registries or surveys)
- Consecutive case series
- Single case reports

When there are merely associations but not causal relationships between a study’s variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study’s selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess the evidence.

2. Generalizability of Clinical Evidence to the Medicare Population
The applicability of the results of a study to other populations, settings, treatment regimens, and outcomes assessed is known as external validity. Even well-designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease, and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing, and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study’s external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator’s lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.

Given the evidence available in the research literature, some degree of generalization about an intervention’s potential benefits and harms is invariably required in making coverage decisions for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation), and similarities of the intervention studied to those that would be routinely available in community practice.

A study’s selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations because one of the goals of our determination process is to assess health outcomes. We are interested in the results of changed patient management not just altered management. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention’s benefits are clinically significant and durable, rather than marginal or short-lived.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

3. Assessing the Relative Magnitude of Risks and Benefits
Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits. Health outcomes are one of several considerations in determining whether an item or service is reasonable and necessary. For most determinations, CMS evaluates whether reported benefits translate into improved health outcomes. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology’s benefits and risk of harm to Medicare beneficiaries.
## Appendix B: Evidence Table

<table>
<thead>
<tr>
<th>Study Group</th>
<th>Author/Year</th>
<th>Study Description</th>
<th>Demographics</th>
<th>Outcomes Analyzed</th>
<th>Results Intervention group</th>
<th>Results Control group</th>
<th>Concomitant Therapies</th>
</tr>
</thead>
<tbody>
<tr>
<td>D01</td>
<td>Rush et al. 2000</td>
<td>VNS Pilot study</td>
<td>30, mean age 47.5, 67% female, 100% Caucasian</td>
<td>10-week outcome of 30 patients</td>
<td>40% had 50% improvement HRSD-28</td>
<td>None</td>
<td>No medication changes allowed except for lorazepam</td>
</tr>
<tr>
<td>D01</td>
<td>Sackeim et al. 2001</td>
<td>VNS Pilot study</td>
<td>60, mean age 46.8, 65% female, 98% Caucasian</td>
<td>10-week outcome of 60 patients</td>
<td>30.5% had 50% improvement HRSD-28</td>
<td>None</td>
<td>No medication changes allowed except for lorazepam</td>
</tr>
<tr>
<td>D01</td>
<td>Marangell et al. 2002</td>
<td>VNS Pilot study</td>
<td></td>
<td>1-year outcome of 30 patients</td>
<td>40% to 46% had 50% improvement HRSD-28</td>
<td>None</td>
<td>Any change allowed</td>
</tr>
<tr>
<td>D01</td>
<td>Nahas et al. 2005</td>
<td>VNS Pilot study</td>
<td></td>
<td>1- and 2-year outcomes of available 60 patient cohort</td>
<td>44% at one year, 42% at 2 years had 50% improvement in HRSD-28</td>
<td>None</td>
<td>Any change allowed</td>
</tr>
<tr>
<td>D02</td>
<td>Rush et al. 2005</td>
<td>Randomized sham controlled trial</td>
<td>222, mean age 47, 63% female, 96% Caucasian</td>
<td>3-month outcomes of randomized sham</td>
<td>VNS:15% had 50% improvement HRSD-24.</td>
<td>Sham</td>
<td>No medication changes allowed</td>
</tr>
<tr>
<td>Study</td>
<td>Authors</td>
<td>Study Design</td>
<td>Patient Characteristics</td>
<td>Outcome Measures</td>
<td>Sham Improvement</td>
<td>Model Comparison</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>D02</td>
<td>Rush <em>et al.</em> 2005</td>
<td>Observational trial (extension of 3-month RCT)</td>
<td>205, mean age 46, 63.9% female, 97% Caucasian</td>
<td>1-year outcomes of extension study</td>
<td>0.45 point improvement in HRSD-24 per month</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>D04</td>
<td>Dunner <em>et al.</em> 2006</td>
<td>Observational trial</td>
<td>124, mean age 46, 68.5% female, 90% Caucasian</td>
<td>1-year outcome of “usual treatment” patients</td>
<td>11.6% and 18.4% had 50% improvement in IDS-SR-30</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>D02/D04</td>
<td>George <em>et al.</em> 2005</td>
<td>Comparison of two groups</td>
<td>Comparison of D-02 and D-04 results</td>
<td>Model estimated average reduction in IDS-SR30 for D02 was 0.4 points per month greater than D04</td>
<td>Any change allowed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix C: Inclusion/Exclusion Criteria

<table>
<thead>
<tr>
<th>D01</th>
<th>D02</th>
<th>D04</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject is diagnosed with major depressive disorder or bipolar disorder (I or II) according to DSM-IV diagnosis criteria;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject is in a chronic (&gt;2 years) current major depressive episode and/or has had a history of recurrent MDEs (at least four lifetime MDEs including a current MDE);</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>Subject for the current MDE, “…must have had an unsatisfactory response to at least two adequate trials of different classes of antidepressant medication, but not more than six, regardless of antidepressant category based on participant/family interviews, medical records, and, when available, pharmacy records.” An adequate trial was defined using a modified Antidepressant Treatment History Form and an Antidepressant Resistance Rating scale;”</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>&gt;20 on the HRSD-24 at baseline and &gt; 18 on the HRSD-24 14 days post implantation</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>IQ &gt;70 based on investigator's judgment;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>&gt;18 and &lt;80 years of age</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>Subject with BPD had demonstrated a resistance to lithium treatment or had a medical contraindication to treatment with lithium or was know to be intolerant to lithium;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject had a history of treatment with psychotherapy (current or previous MDE) or at least 6 weeks duration that did not result in a substantial clinical improvement;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject was stable on current antidepressant medication regimen of ≤5 antidepressant medication for ≥4 weeks prior to baseline or subject was not taking antidepressant medications for ≥4 weeks prior to baseline;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject was stable on current atypical antipsychotics and anticonvulsant medication for ≥4 weeks prior to first baseline visit or subject was not taking atypical antipsychotics and anticonvulsant medications for ≥4 weeks prior to first baseline visit.</td>
</tr>
<tr>
<td>*</td>
<td></td>
<td></td>
<td>Assessments after 8 and 10 weeks of sham VNS have an average score of ≥ 18</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Age 18 to 70</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>For the current MDE, not responsive to 2 or more medication classes</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>For the current MDE not responsive to 6 weeks of psychotherapy</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Baseline HDRS-28 of 20 or greater</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Score &lt; 50 on the Global Assessment of Function</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>&gt;20 on the HRSD-24 at baseline</td>
</tr>
<tr>
<td>D01</td>
<td>D02</td>
<td>D04</td>
<td>Exclusion Criteria</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject met DSM-IV criteria for atypical depression at the time of study entry or subject had ever had psychotic symptoms in any MDE;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject had a history of nonmood psychotic disorders;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject with BPD had a history of rapid cycling;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Subject currently had a secondary diagnosis of, or signs of, delirium, dementia, amnestic, or other cognitive disorders per DSM-IV;</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>“Subject did not have an acceptable clinical response due to failure (resistance based on antidepressant resistance rating [ARR] score $\geq 3$) with $\geq 7$ antidepressant treatments (regardless of category) during the current MDE”</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>“Suicide attempt within the previous 12 months that required medical treatment, or $\geq 2$ suicide attempts in the past 12 months, or established plan for suicide during study, or was likely to attempt suicide within 6 months;”</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject had a history of myocardial infarction or cardiac arrest;</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
<td>Subject had received general anesthesia within 30 days prior to enrollment;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject had taken an investigational drug within a clearance duration of five times the half-life of the investigational drug or within $\geq 4$ weeks prior to first baseline visit, whichever longer;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject had a significant cardiac or pulmonary condition currently under treatment resulting in an ASA score $\geq$ III;</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Subject had a demand cardiac pacemaker, implantable defibrillator, or other implantable stimulators.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Suicide intent</td>
</tr>
<tr>
<td>Study</td>
<td># of sites</td>
<td>Outcomes by site</td>
<td>Primary outcome</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>D01</td>
<td>4</td>
<td>Variation.</td>
<td>$\geq 50%$ decrease in HAM-D 24 at 10 weeks of therapy</td>
</tr>
<tr>
<td>D02 acute phase</td>
<td>21</td>
<td>Variation. seven sites had $&lt;10%$ response rate, four sites had $\geq 25%$ response rate</td>
<td>$\geq 50%$ decrease in HAM-D 24 at 10 weeks of therapy</td>
</tr>
<tr>
<td>D02 long-term phase</td>
<td>21</td>
<td>no</td>
<td>Repeated measures analysis of HAM-D 24</td>
</tr>
<tr>
<td>D04</td>
<td>13</td>
<td>no</td>
<td>$\geq 50%$ improvement in IDS-SR at</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the last two measured quarters</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>--------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Additional References from Public Comment on Proposed Decision


Burke W, Moreno F. Suicidality and treatment resistant depression: Results from a 24 month trial of vagus nerve stimulation (poster) presented at the 156th American Psychiatric Association annual meeting in May 20-25, 2006 Toronto Canada.


Carpenter LL, Moreno F, Kling MA. Effects of vagus nerve stimulation on cerebrospinal fluid in depressed patients. Biol Psychiatry. 2002;51(suppl 1):152S-153S.


Cassak D. Cyberonics' two-front war. In Vivo. 2006;24.


Lesser IM, Leuchter AF, Trivedi MH et al Characteristics of insured and noninsured outpatients with depression STAR*D. Psychiatric Services 56:995-1004, 2005


Martinez JM, Zboyan HA. Vagus nerve stimulation therapy in a patient with treatment-resistant depression: a case report of long-term follow-up and battery end-of-service. CNS Spectr. 2006;11:143-147.


Miller IW. Keitner GI. Schatzberg AF, et al.: The treatment of chronic depression, part 3:

Miller et al. 1999. Unable to identify.


Pardo JV, Sheikh SA, Abuzzahab FS, Lee JT, Rittberg B. Metabolic abnormalities in treatment resistant depression (TRD): comparing the metabolism in medicated/vagus nerve stimulator (VNS) implanted patients to the patients without medication/VNS implant. Biol Psychiatry. 2002;51(suppl 1):33S.


Results of VNS long-term depression study are positive. J Psychosoc Nurs Ment Health Serv. 2001;39:12.