

Final Statistical Summary Review for PMA P970003/S50 (Original and Various Amendments), Vagus Nerve Stimulator (VNS) Therapy System for Depression, Cyberonics, Inc.

I. **Introduction**

The VNS system is indicated for the adjunctive long-term treatment of chronic or recurrent depression in patients who are experiencing a major episode that has not had an adequate response to two or more antidepressant treatments. This review summarizes the important statistical issues and results for pivotal D-02 study (VNS plus Standard of Care), observational D-04 (Standard of Care alone), and D-02/D-04 comparison. The primary efficacy endpoint is the comparison of average rate of change per month (slope) and average change from baseline between D-02 and D-04 patients for the evaluable patient population. The secondary efficacy endpoint is the comparison of proportions of Response (defined as  $\geq$  50% decrease in scores from baseline) at 12 months.

The Hamilton Rating Score for Depression (**HRSD-24**) is the primary efficacy endpoint in the D-02 pivotal study. However, since HRSD-24 scores were collected only at baseline and at 12 months in the D-04 observational study, the Inventory of Depressive Symptomatology Self Report (**IDS-SR**) was used as the primary efficacy endpoint in the D-02/D-04 comparison via repeated-measures linear regression (**RMLR**) analyses. The **RMLR** requires both patient baseline and multiple post-baseline measurements to estimate average rate of change per month (slope) and the difference of two true slopes for D-02/D-04 comparison.

II. **Multi-Center Study Data**

There are 22 sites (centers) which participated in either D-02 or D-04 studies. Of these 22 sites, 12 sites participated in both D-02 and D-04 studies (called **overlapping** sites), 9 sites enrolled D-02 patients but no D-04 patients, and 1 site enrolled D-04 patients but no D-02 patients. The numbers of patients for “evaluable” and “12-month completer” patient population, separately by all participating and overlapping sites, are shown in Table 1.

Table 1. Number of Patients (N)<sup>a</sup> by “All Sites” and “**Overlapping**” Sites, D-02/D-04 Study

Site	D-02 Long-Term	D-02 Evaluable	D-02 12-Month Completers	D-04 Evaluable	D-04 12-Month Completers
All (22)	233	205 (185	177	124	112

		Unipolar, 20 Bipolar)			
<b>Overlapping</b> (12)	165	147	128	120	108

a. Sample size (N) was justified for the comparison of two response proportions, the *secondary* efficacy endpoints; not for the comparison of two slopes, the *primary* efficacy endpoint. The detailed distribution of patients by clinical site (Non-overlapping and **Overlapping**) is shown in Table 10, Appendix 1.

### III. **D-02 Pivotal Study**

The D-02 study included patients whose HRSD-24  $\geq$  18 anytime during the 12-month follow-up, and HRSD-24  $\geq$  29 at acute phase (at 3 months). The following Table 2 provides a brief summary for D-02 group patients

Table 2. Brief Summary for **D-02** Study, All Sites

	<b>Acute (3 months)</b>	<b>Long-Term (1-year)</b>
Study Design	Double-blind, randomized, parallel, Active VNS versus Sham control, Multi-center (22)	Active VNS $\rightarrow$ VNS Sham control $\rightarrow$ VNS ( <b>Delayed-treatment</b> group)
Follow-up	Baseline (2), Implanted, 2 weeks, ... 3 months	Monthly in the first year, quarterly thereafter
Clinical Outcome	HRSD-24 Score, Primary; IDS-SR Score, Secondary	HRSD-24 score (Per-Protocol)
Primary Endpoint	Comparison of two response proportions	Average rate of change per month (slope)-Repeated-Measures Linear Regression ( <b>RMLR</b> )
Result	<b>HRSD-24</b> (N = 221) 15% (17/111) VNS, 10% Sham (11/110) p = 0.31 (Fisher's exact) <b>IDS-SR</b> (N = 215) 17.4% (19/109)VNS, 7.5% ( 8/106) Sham, p = 0.039 (Fisher's exact)	<b>Evaluable</b> ( N = 205) [Slope = -0.45, Standard error = 0.05, 95% CI: (-0.55, -0.34), p<0.001 to reject the true null hypothesis (slope = <b>0</b> )] <b>12-Month Completers</b> ( N = 177) [Slope = -0.47, Standard error = 0.06, 95% CI: (-0.58, -0.36), p<0.001]

### IV. **D-02 and D-04 Comparison**

#### **Propensity Score (PS) Adjustment**

Since D-04 is an observational study (Standard of Care alone), evaluation of true device effect must control for potential bias or confounding effect in differences

for individual patient demographic characteristics and clinically important baseline covariates between D-02 and D-04 group patients.

The *PS* approach is to derive an overall summary composite score of sponsor's selected 17 patient binary or continuous covariates [age, gender, bipolar versus unipolar depression, lifetime electroconvulsive therapy (ECT) use, length of current major depressive episode (MDE) in months, average number of lifetime episodes of depression, percent of patients received ECT in lifetime, percent of patients received ECT in current MDE, percent of patients with suicide attempt in lifetime or in the past 12 months, and others]. The purpose of *PS* analysis is to reduce bias in non-randomized, observational studies, such as in the D-02/D-04 comparison. Statistical logistic regression is used to predict D-02 treatment assignment conditional on the individual patient's covariates. The resulting individual patient predicted probabilities of receiving active treatment group (D-02) and control (D-04) groups were then ordered to form a 5-subgroup or quintiles based on the estimated propensity scores. For example, the first quintile group contains approximately 20% of the patients with the lowest D-02 *PS* and the last group contains approximately 20% of the patients with the highest D-02 *PS*. *PS* can only adjust for *observed* covariates, not for *unobserved* ones. *PS* analysis may *not* eliminate all selection bias, particularly hidden bias.

In my previous reviews, this reviewer asked the sponsor to provide the following information regarding their *PS* analysis:

Justify selection criteria for fitted logistic regression model:

Graphical display (e.g., Bar chart) of the distribution of *PS* quintile means (for continuous covariates) or quintile proportions (for binary covariates) between D-02 and D-04 patients;

For each selected patient covariates (17), prepare statistical analyses for both before-and-after *PS* adjustment between D-02 and D-04 patients. Explain the degree of covariate unbalance before *PS* adjustment and covariate balance (or unbalance) after adjustment;

Explain 2-way analysis of variance including main effect (treatment group, *PS* quintile) and their interactions.

The sponsor has responded to the above comments in the March 17, 2004 Amendment # 4.

### **Repeated-Measures Linear Regression (RMLR) Analysis**

The RMLR analysis is used to evaluate average rate of change (slope) and average change in IDS-SR scores from baseline to the 12-month follow-up. SAS PROC MIXED was used to analyze the 12-month follow-up data. No missing

data imputation is needed to run SAS PROC MIXED since missing data are assumed to be missing at random (MAR), which means that probability of missing data is independent of future observed data. The last observation carried forward (LOCF) analysis was also prepared by the sponsor for comparison purpose. The patient covariates used in the general mean response mixed model include several fixed-effect study factors [9 *pooled* sites with some pooled sites containing only D-02 patients (see Table 18.2, March 17, 2004 submission), treatment (D-02 versus D-04), 5-level grouped *PS* quintiles, baseline IDS-SR score, indicator variables for follow-up time at 3, 6, 9, and 12 months, and treatment by time interactions]. The spatial power covariance structure is used to count for correlation among different follow-up times.

In the March 26, 2004 email, FDA asked the sponsor to respond to the following questions:

Provide an analysis the IDS-SR primary efficacy endpoint in the *RMLR* analysis, HRSD-24 secondary efficacy endpoint, and Response/Non-Response (see following section) proportions from only those sites that enrolled for *both* D-02 and D-04 (*overlapping* sites, see **Appendix 1**) for unipolar and bipolar patients **combined**.

Repeat the above analyses for *censored* patients (i.e., additions or changes in either antidepressant drugs or ECT).

The sponsor has responded to the above FDA's comments. For *censored* patients approach in *RMLR* analyses, the sponsor indicated that "IDS-SR raw scores were censored such that the value from the patient's IDS-SR measurement obtained *prior* to their *first* increase in the antidepressant resistance rating (ARR) score was carried forward and **replaced** all of the patient's subsequent, **non-missing**, IDS-SR measurements."

Please note that the current *ECT* addition or change during the follow-up was *not* discussed above.

### **Comparison of Proportions of Response (for 12-Month Completers)**

The comparison of two response proportions (? 50% reduction from baseline in IDS-SR or HRSD-24 scores), defined as one of several other secondary efficacy endpoints, is discussed in this summary review. However, the direct, simple comparison of two response proportions between D-02 and D-04 patients, without adjusting for individual patient baseline IDS-SR or HRSD-24 scores and other clinically important patient covariates, is subject to potential bias. The 12-month completers are the patients who were in close compliance with the scheduled follow-up, may provide the "*best-case*" scenario as compared to

those who did not complete the 12-month study. The selected cutoff point (? 50% from baseline, response; else, non-response) is also subject to measurement error, variability of the IDS-SR or HRSD-24 scores, and serial correlation of repeated-measures data. Statistical logistic regression is more appropriate than simple comparison of two proportions by taking important patient covariates (site, baseline IDS-SR or HRSD-24, and others) into the model building. Sensitivity analysis may be helpful to evaluate robustness of Response/Non-Response outcomes by various cutoff points. Appropriate statistical methods for pooling of multi-center data, such as meta-analysis, stratified categorical data analysis, will provide precise evaluation of true treatment effect for comparison of two response proportions.

### **Concordance Between HRSD-24 and IDS-SR Scores in D-02 Study**

As discussed above, the IDS-SR score was used in the RMLR analyses to estimate the average rate of change per month (slope) for D-02 and D-04 patients. Sponsor's justification is that the HRSD-24 score data were collected only at baseline and at the 12 months in the D-04 study, and that the IDS-SR score had been shown as a "*good predictor*" of the HRSD-24 score from the published literature.

In FDA's major deficiency letter of March 1, 2004 and FDA's E-mail dated March 31, 2004, we do not agree that concordance studies reported in the published literature are sufficient to support IDS-SR as a "*good predictor*" of HRSD-24. Due to wide variability of paired IDS-SR/HRSD-24 scores from patient to patient, *pooled-patient* correlation and regression analyses are not appropriate. FDA asked the sponsor to prepare the following analyses:

Calculate correlation coefficients between IDS-SR and HRSD-24 scores for *each* individual patient and a pooled estimated correlation coefficient over all patients by an appropriate statistical method.

Likewise, calculate corresponding results for estimated slopes and their standard errors for each individual patient.

Provide the fitted linear regression model (intercept, slope, or higher terms), the estimated parameter values, standard errors, 95% confidence intervals, and squared multiple correlation coefficient (R-Square) to show goodness-of-fit of sponsor's regression equation to the observed paired IDS-SR/HRSD-24 data.

Provide graphical displays for all individual patient paired HRSD/IDS-SR data to show all observed individual patient data pairs, and the fitted regression lines.

The sponsor has responded to all of the above FDA's requested issues. However, the linear regression model assumes that all individual patient IDS-SR/HRSD-24 pairs are independent, but they are actually correlated in the more proper

longitudinal data analyses. Nevertheless, this reviewer believes that linear regression analyses based on individual patient paired IDS-SR/HRSD-24 data may still be applied to verify the sponsor's claimed IDS-SR as a "*predictor*" of HRSD-24.

### **Unipolar and Bipolar Patients**

In Question # 9 of FDA's major deficiency letter of March 1, 2004, we asked the sponsor to prepare separate and combined analyses for unipolar and bipolar patients for both Response/Non-Response secondary efficacy endpoint, and *RMLR* analyses for the primary efficacy endpoint. The sponsor's responded in their March 17, 2004 submission. The unipolar/bipolar subgroup analyses were not discussed in the original study design. The sample size for bipolar patients is too small to provide any statistically valid conclusion. The distribution of sample size in D-02 and D-04 comparison is shown in Table 3.

Table 3. Distribution of Number of Patients in the D-02 and D-04 Comparison, by Unipolar/Bipolar Patients, IDS-SR (HRSD-24) Scores

Group	D-02	D-04
<b><u>Unipolar</u></b>		
Evaluable	163 (164)	97 (91)
12-Month Completer	156 (157)	97 (91)
<b><u>Bipolar</u></b>		
Evaluable	17 (17)	15 (13)
12-Month Completer	17 (17)	15 (13)
<b><u>Combined</u></b>		
Evaluable	180 (181)	112 (104)
12-Month Completer	173 (174)	112 (104)

## V. Statistical Analyses Results

### PS Analyses

All graphical displays for each of 17 covariates for D-02 and D-04 comparisons are shown in Attachment 20 of March 17, 2004 submission. The graphical displays appear to be acceptable to examine comparability of D-02 and D-04 patient populations with respect to these 17 covariates after *PS* adjustment.

The before-and-after *PS* comparisons for 17 covariates are also shown. The statistically significant differences in some covariates between D-02 and D-04 group patients *before PS* adjustment were non-significant *after PS* adjustment. For example, a statistically significant p-value <0.001 in percent of patients who received ECT in current MDE between D-02 and D-04 groups became non-significant (p = 0.434) after *PS* adjustment. Although some *PS* quintile by treatment interaction is also shown for percent of patients who received ECT in their lifetime and length of current MDE and treatment, the sponsor's *PS* adjustment procedures via logistic regression model appear to be acceptable. The final *PS* quintile by treatment frequency distribution is shown in Table 4.

Table 4. Treatment (D-02/D-04) by *PS* Quintile Frequency Distribution (Evaluable Patients)

PS Quintile Group	D-02 ( N = 205)	D-04 ( N = 124)	
1	22 (10.9%)	43 (34.7%)	
2	39 (19.4%)	26 (21.0%)	
3	36 (17.9%)	29 (23.4%)	
4	48 (23.9%)	17 (13.7%)	
5	56 (27.9%)	9 ( 7.3%)	
Total	201 (100%)*	124 (100%)	

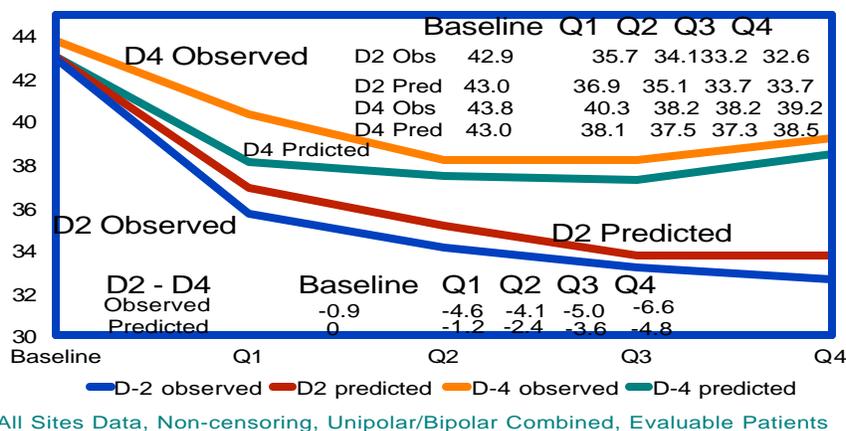
[\*: 4 patients excluded from *PS* analysis]

The above frequency distributions are statistically acceptable. Patients belong to each of the above 5 *PS* quintile groups were coded as the categorical variable (5 levels) in the *RMLR* analyses.

### RMLR Analyses, D-02 and D-04 Comparison, IDS-SR Scores

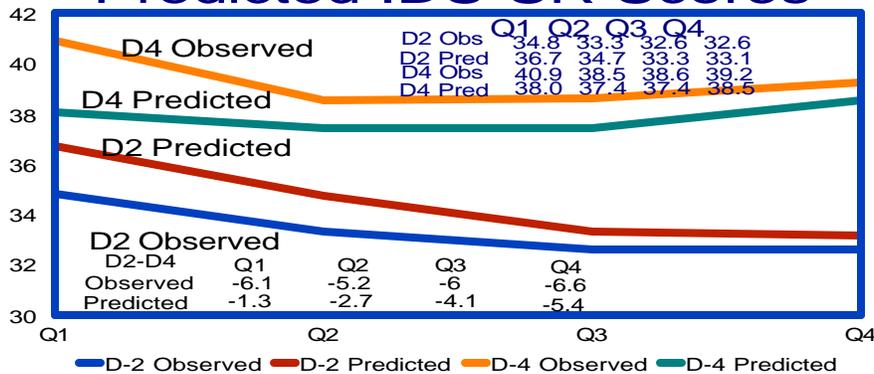
The following Figure 1 shows the observed and predicted (by *RMLR*) mean IDS-SR scores by baseline, and each of 4 quarters (9 pooled sites, evaluable patients). The predicted mean IDS-SR scores appear to be close to the observed scores. The predicted differences in mean IDS-SR scores between D-2 and D-4 patients showed smaller improvement than these observed scores between D-02 and D-04 patients. For example, at Quarter 4, the predicted difference (D2 – D4) is -4.8 (33.7 – 38.5) and the observed difference (D2 – D4) is -6.6 (32.6 – 39.2).

### Figure 1. The Observed and Predicted Mean IDS-SR Scores



The following Figure 2 shows the corresponding mean IDS-SR scores for 12-month completers

### Figure 2. The Observed and Predicted IDS-SR Scores



For primary effectiveness endpoint (IDS-SR), the differences of average rate of change per month (slope) and their 95% confidence intervals (CI) for Evaluable and 12-month completers are shown in Table 5.

Table 5. Difference of slope (D2 – D4) and the 95% CI, All Sites, Unipolar/Bipolar Patients Combined, All Sites

Patient Population	Difference (Std Error)	95% CI for Difference
Evaluable	-0.397 (0.1)	(-0.59, -0.21)
12-Month Completers	-0.452 (0.1)	(-0.65, -0.26)

Clinical interpretation is needed to decide whether or not the above results are clinically acceptable.

The sample size for the above D-2 and D-4 comparison is shown in Table 6

Table 6. Sample Size (N) by D-2/D-4 Comparison, All Sites

Group	Baseline	Q 1	Q 2	Q 3	Q 4
D2 (N)	201*	200	195	183	177
Missing	0	1	6	18	24
D4 (N)	124	120	119	116	112
Missing	0	4	5	8	12

\*: 4 evaluable patients did not have IDS-SR and/or *PS* scores available

I have revised part of sponsor's reported proportions of Response/Non-Response for IDS-SR and HRSD-24 scores, for 12-month completers, as shown in sponsor's Tables 3 and 4, Volume 19, Clinical Summary (See Tables 7-A and 7-B below)

Table 7-A. **FDA's Revised** Proportions of Response for 12-Month Completers, IDS-SR Scores, All Pooled Sites

<b>12-Month Data</b>	D-02	D-04	p-value <sup>a</sup>
Response <sup>b</sup>	22 % (38/173)	12% (13/112)	0.027
LOCF Response	22% (39/176)	12% (13/112)	0.027
Complete Response <sup>c</sup>	15% (27/180)	4% (4/112)	0.001
LOCF Complete Response	13% (27/204)	3% (4/124)	0.003

a. Fisher's two-sided exact test

b. ? 50% decreasing change from baseline; c. IDS-SR ? 14

Table 7-B. **FDA's Revised** Proportions of Response for 12-Month Completers, HRSD-24 Scores, All Pooled Sites

<b>12-Month Data</b>	D-02	D-04	p-value <sup>a</sup>
Response <sup>b</sup>	30 % (54/181))	13% (13/104)	0.001
LOCF Response	27% (55/205)	13% (13/104)	0.004
Complete Response <sup>c</sup>	17% (31/181)	7% (7/104)	0.018
LOCF Complete Response	16% (32/205)	7% (7/104)	0.029

- a. Fisher's two-sided exact test;
- b. ? 50% decreasing change from baseline
- c. HRSD-24 ?9

Please note that, in Tables 7-A and 7-B, p-values calculated from direct pooling of all cell frequencies (Response/Non-Response by treatment group) over all sites (non-overlapping and overlapping), without preparing appropriate statistical modeling approach or meta-analysis, may be invalid.

Under the section for **RMLR**, in the FDA's March 26, 2004 email, we asked the sponsor to reanalyze the IDS-SR and HRSD-24 score data from only these sites that enrolled **both** D-2/D-4 (**overlapping** sites) and for **censored** patients (i.e., additions or changes in either antidepressant drugs or ECT). The following **FDA's revised** Tables 8-A through 8-D are for Tables 24-1 through 24-4 shown in the sponsor's responses of April 2, 2004 to the FDA's email dated March 26, 2004 (Amendment # 6).

**FDA's Revised Table 8-A**

**IDS-SR Scores-D-02/D-04 Comparisons (*Overlapping* sites for both D-02 and D-04 patients only), **Evaluable** Patient Population (Unipolar and Bipolar patients combined)**

	D-02	D-04	P-value	95% CI (Least square mean)	RR(95% CI) <sup>b</sup> D2 – D4
N at Baseline	147	120			
Baseline Average	42.7	43.6			
<b>12 Month Data</b>	N = 131	N = 108			
Average	33.8	39.4			
Average change from baseline (SD)	-8.9 (13.3)	-4.2 (12.1)	0.003	(-8.9, -1.8)	
LOCF average change from baseline (SD)	-8.4 (12.8)	-4.6 (12.2)	0.021	(-7.1, -0.6)	
Response (% of Subjects) <sup>c</sup>	19.8 (26/131)	11.1 (12/108)	<b>0.076<sup>a</sup></b>		1.8 (0.94, 3.4)
LOCF Response (% of Subjects)	17.7 (26/147)	11.7 (14/120)	<b>0.227</b>		1.5 (0.83, 2.8)
Complete Response (% Subjects) <sup>d</sup>	13.0 (17/131)	2.8 (3/108)	<b>0.0045</b>		4.7 (1.4, 15.5)
LOCF Complete Response (% Subjects)	11.6 (17/147)	2.5 (3/120)	<b>0.0048</b>		4.6 (1.4, 15.4)

a By Fisher's two-sided exact test

b Risk Ratio (RR) = [P(Response) for D-02]/[P(Response) for D-04]  
Example: for Response, the estimated RR = (26/131)/ (12/108) = 1.78

c Response: ? 50% decreasing change from baseline

d Complete Response: IDS-SR ? 14

**Primary effectiveness endpoint** [Difference in two slopes, D2 – D4 and 95% CI: -0.32 per month, 95% CI: (-0.52, -0.12), p = 0.002 to reject true null hypothesis (difference = 0)] (see April 2, 2004 Amendment # 6)

**FDA's Revised Table 8-B**

**HRSD-24 Scores-D-02/D-04 Comparisons (Overlapping sites  
for both D-02 and D-04 patients only), Evaluable Patient Population**

	D-02	D-04	P-value	95% CI (Least square mean)	RR (95% CI) <sup>b</sup>
N at baseline	147	120			
Baseline Average	27.4	27.7			
<b>12 Month Data</b>	N = 130	N = 100			
Average	19.7	23.0			
Average change from baseline (SD)	-7.7 (8.8)	-4.7 (7.6)	0.020	(-5.1, -0.4)	
LOCF Average change from baseline (SD)	-6.9 (8.9)	-4.7 (7.6)	0.113	(-4.0, 0.4)	
Response (% of Subjects) <sup>c</sup>	27.7 (36/130)	11.0 (11/100)	<b>0.0018<sup>a</sup></b>		2.5 (1.3, 4.7)
LOCF Response (% of Subjects)	25.2 (37/147)	11.0 (11/100)	<b>0.0055</b>		2.3 (1.2, 4.2)
Complete Response (% Subjects) <sup>d</sup>	16.9 (22/130)	5.0 (5/100)	<b>0.0065</b>		3.4 (1.3, 8.6)
LOCF Complete Response (% Subjects)	15.6 (23/147)	5.0 (5/100)	<b>0.013</b>		3.1 (1.2, 7.9)

a. By Fisher's two-sided exact test

b. Risk Ratio (RR) = [P(Response) for D-02]/[P(Response) for D-04]

c. Response: ? 50% decreasing change from baseline

d. Complete Response: HRSD<sub>24</sub> ? 9

**FDA's Revised Table 8-C**

**IDS-SR Scores After D-02 Censoring Only-D-02/D-04 Comparisons  
(Overlapping sites for both D-02 and D-04 patients only), Evaluable  
Patient Population**

	D-02	D-04	P-value	95% CI (Least square mean)	RR (95% CI) <sup>b</sup>
N (at Baseline)	147	120			
Baseline Average	42.7	43.6			
<b>12 Month Data</b>	N = 131	N = 108			
Average	36.0	39.4			
Average change from baseline (SD)	-6.7 (13.3)	-4.2 (12.1)	0.026	(-7.5,-0.5)	
LOCF Average change from baseline (SD)	-6.1 (12.8)	-4.6 (12.2)	0.160	(-5.5, 0.9)	
Response (% of Subjects) <sup>c</sup>	16.8 (22/131)	11.1 (12/108)	<b>0.26<sup>a</sup></b>		1.5 (0.78, 2.9)
LOCF Response (% of Subjects)	15.0 (22/147)	11.7 (14/120)	<b>0.47</b>		1.3 (0.7, 2.4)
Complete Response (% Subjects) <sup>d</sup>	8.4 (11/131)	2.8 (3/108)	<b>0.095</b>		3 (0.86, 10.6)
LOCF Complete Response (% Subjects)	7.5 (11/147)	2.5 (3/120)	<b>0.097</b>		3 (0.85, 10.5)

a. By Fisher's two-sided exact test

b. Risk Ratio (RR) = [P(Response) for D-02]/[P(Response) for D-04]

c. Response: ? 50% decreasing change from baseline

d. Complete Response: IDS-SR ? 14

**Primary effectiveness endpoint** [Difference in two slopes, D2 – D4 and 95% CI: -0.18 per month, 95% CI: (-0.38, 0.02), p = 0.079 to reject true null hypothesis (difference = 0)] (see April 2, 2004 Amendment # 6)

**RMLR** predicted mean IDS-SR (Table 6.2.37, Amendment # 6)

Quarter	D-2 (SE)	D-4 (SE)	D2-D4	95% CI*
1	37.93 (0.59)	38.47 (0.60)	-0.54	(-1.70, 0.64)
2	37.01 (0.65)	38.09 (0.69)	-1.08	(-2.38, 0.24)
3	36.35 (0.71)	37.96 (0.81)	-1.61	(-3.10, -0.12)
4	36.58 (0.77)	38.72 (0.98)	-2.14	(-3.84, -0.54)

(\*The average of two standard errors was used as pooled SE for 95% CI, N unknown)

### FDA's Revised Table 8-D

**HRSD-24 Scores-After D-02 Censoring Only -D-02/D-04**  
**Comparisons (Overlapping sites for both D-02 and D-04 patients only),**  
**Evaluable Patient Population**

	D-02	D-04	P-value	95% CI (Least square mean)	RR (95% CI) <sup>b</sup>
N (at 12-Month)	147	120			
Baseline Average	27.4	27.7			
<b>12 Month Data</b>	N = 130	N = 100			
Average	22.5	23.0			
Average change from baseline (SD)	-4.9 (9.1)	-4.7 (7.6)	0.581	(-3.9, 1.7)	
LOCF Average change from baseline (SD)	-4.3 (8.9)	-4.7 (7.6)	0.910	(-2.1, 2.4)	
Response (% of Subjects) <sup>c</sup>	18.5 (24/130)	11.0 (11/100)	<b>0.14<sup>a</sup></b>		1.7(0.86,3.3)
LOCF Response (% of Subjects)	16.3 (24/147)	11.0 (11/100)	<b>0.27</b>		1.5(0.76,2.3)
Complete Response (% Subjects) <sup>d</sup>	7.7 (10/130)	5.0 (5/100)	<b>0.59</b>		1.5(0.54,4.3)
LOCF Complete Response (% Subjects)	6.8 (10/147)	5.0 (5/100)	<b>0.79</b>		1.3(0.48,3.8)

- By Fisher's two-sided exact test
- Risk Ratio (RR) = [P(Response) for D-02]/[P(Response) for D-04]
- Response: ? 50% decreasing change from baseline
- Complete Response: HRSD<sub>24</sub> ? 9

In Tables 8-A through 8-D, the 95% confidence interval (CI) for the difference (D2 – D4) of two average changes from baseline, rather than the calculated p-values, would allow better clinical evaluation, because p-values are simply used to reject the true null hypothesis that two average changes = 0 against the alternative hypothesis that two average changes ? 0.

In Tables 8-C and 8-D for censored patients (*overlapping* sites), as described previously, that "IDS- SR or HRSD-24 raw scores were *censored* such that the

value from the patient’s IDS-SR or HRSD-24 measurement obtained *prior to* their first increase in the antidepressant resistance rating score was carried forward and replaced all of the patient’s subsequent, *non-missing*, IDS-SR measurements.

In Tables 8-C and 8-D (*Censored* patients from *overlapping* sites, see Appendix 1), most of statistical results *failed* to support that D-2 patients showed superior IDS-SR or HRSD-24 results to those for D-4 patients, except for average change from baseline comparison at 12 months for IDS-SR only.

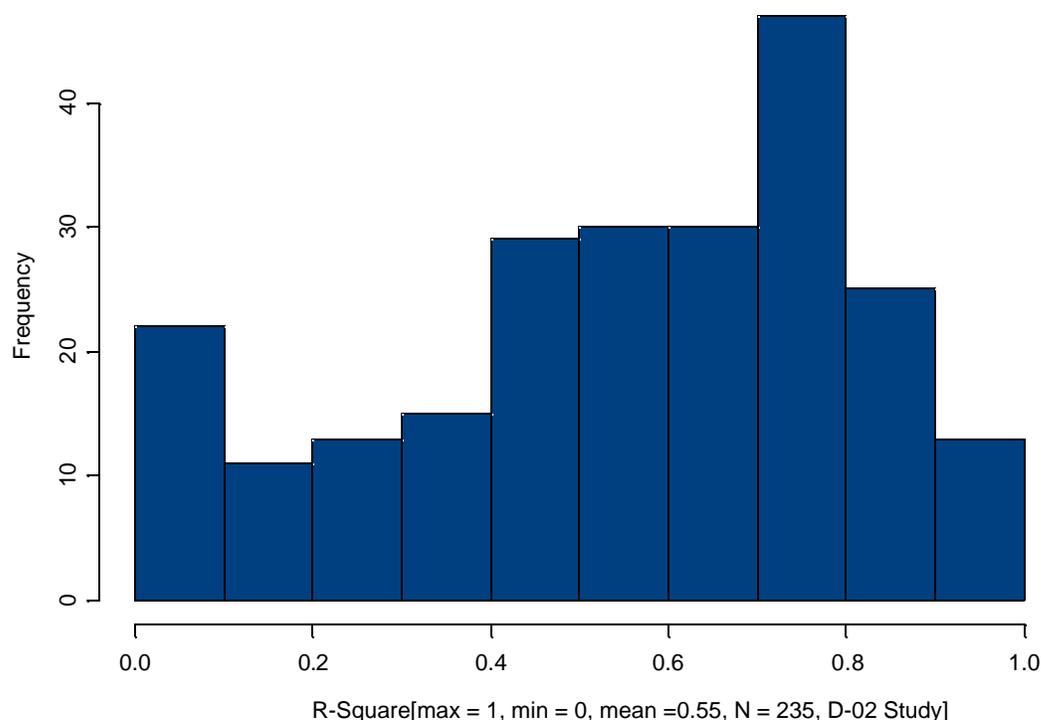
### Concordance between IDS-SR and HRSD-24 Scores

Due to wide variability of paired IDS-SR/HRSD-24 scores from patient to patient, FDA requested the sponsor to calculate the estimated correlation coefficient and its 95% confidence interval (CI), the estimated regression intercept and slope and their 95% CI, the unadjusted (for degrees of freedom) R-Square ( $R^2$ ), which measures the “proportion of total variation about the mean HRSD-24 explained by the fitted regression equation”, from each *individual* patient. The  $R^2$  evaluates how well IDS-SR, predicts HRSD-24 score.  $R^2$  ranged from 0% (worst prediction fit) to 100% (perfect prediction fit). In Figure 3, the histogram of  $R^2$  shows relatively poor to fair prediction with mean  $R^2$  of 0.55, ranging from 0 to 1, for 235 evaluable patients from the D-2 study. In Table 9, the average simple Pearson correlation coefficient of 0.7 with 95% CI (0.67, 0.73) between IDS-SR and HRSD-24 scores, again indicates that IDS-SR is *not* a “*good predictor*” of HRSD-24.

Table 9. Summary of Correlation Coefficient and Regression Slope, for 235 Evaluable Patients, Paired IDS-SR/HRSD-24 Scores, All Sites

Parameter	N	Mean	SD	Median	Min	Max	Lower 95% CL	Upper 95% CL
Pearson Correlation Coefficient	235	0.7	0.25	0.77	-0.26	1.0	0.67	0.73
Slope	235	0.55	0.25	0.56	-0.76	1.49	0.51	0.58

Figure 3, Histogram for Regression R-Square, IDS versus HAM-D



## VI. Conclusion

The IDS-SR does *not* to be a “*good predictor*” of the HRSD-24 based on the distribution of  $R^2$  values (Figure 3) and sample correlation coefficients (Table 9) of 235 D-2 evaluable patients.

Due to small sample size (Table 3) for bipolar patients, no valid statistical analyses can be prepared, but a clinical decision is needed.

The required sample size analyzed in this PMA was based on neither the comparison of two true slopes (**Primary effectiveness** endpoint) nor mean responses in this repeated-measures/longitudinal data analyses. No minimum clinically detectable difference in two slopes or mean HRSD-24 or IDS-SR was defined at the study design stage in order to estimate the required sample size with pre-specified power, type I error, estimated variability of the data, number of follow- up visits, and correlation among repeated measures.

A clinical decision is also required to decide several important issues, such as pooled sites, potentially important *hidden covariates* in the *PS* analysis, non-overlapping sites or *overlapping* sites, and *censored* or *non-censored* analyses.

The validity of statistical inferences from comparison of two proportions of Responses pooled over all non-overlapping and overlapping sites, without any appropriate statistical modeling approach, such as meta-analysis, is highly questionable. Clinically important patient covariates, such as patient baseline IDS-SR or HRSD-24 measurements, clinical site, and others, must be considered in the comparison of two proportions via statistical modeling approach or stratified categorical data analysis.

For *censored* and *overlapping* sites, *no* statistically significant differences in primary effectiveness endpoint (Difference in two slopes, IDS-SR, Table 8-C) and secondary effectiveness endpoint (Difference in two proportions of responses or difference in average change from baseline, HRSD-24, Table 8-D) were found.

Due to above statistical issues, such as questionable concordance between HRSD-24 and IDS-SR, questionable pooling of multi-center data for comparison of proportions of responses, statistically insignificant findings from censored and overlapping sites (Tables 8-C and 8-D) for IDS-SR primary effectiveness endpoint (Slope) and HRSD-24 secondary effectiveness endpoint (Response proportions), it is unclear whether the effectiveness claim of D-02 over D-04 group patients has been demonstrated.

## Appendix 1

Table 10. Distribution of Number of Patients by Clinical Site  
(10 Non-overlapping and 12 *Overlapping* Sites), D-02 and D-04 Study

Site	Long-Term, D-02	Evaluable D-02	12-Month D-02	Evaluable D-04	12-Month D-04
040	<b>16</b>	<b>15</b>	<b>15</b>	<b>6</b>	<b>5</b>
041	<b>10</b>	<b>9</b>	<b>7</b>	<b>3</b>	<b>1</b>
042	10	9	7	-	-
043	<b>13</b>	<b>12</b>	<b>10</b>	<b>8</b>	<b>7</b>
044	<b>17</b>	<b>13</b>	<b>10</b>	<b>12</b>	<b>12</b>
045	<b>18</b>	<b>18</b>	<b>15</b>	<b>13</b>	<b>10</b>
046	<b>13</b>	<b>11</b>	<b>10</b>	<b>2</b>	<b>2</b>
047	9	7	7	-	-
048	7	7	7	-	-
049	<b>12</b>	<b>10</b>	<b>9</b>	<b>11</b>	<b>11</b>
050	<b>10</b>	<b>9</b>	<b>8</b>	<b>8</b>	<b>6</b>
051	9	7	3	-	-
052	9	8	7	-	-
053	4	3	2	-	-
054	<b>13</b>	<b>12</b>	<b>12</b>	<b>15</b>	<b>14</b>
055	8	5	4	-	-
056	9	9	9	-	-
057	<b>6</b>	<b>5</b>	<b>5</b>	<b>16</b>	<b>14</b>
058	<b>19</b>	<b>17</b>	<b>14</b>	<b>14</b>	<b>14</b>
059	<b>18</b>	<b>16</b>	<b>13</b>	<b>12</b>	<b>12</b>
060	3	3	3	-	-
071	-	-	-	4	4

Total (All)	233	205	177	124	112
<b>Total</b>	<b>165</b>	<b>147</b>	<b>128</b>	<b>120</b>	<b>108</b>
<i>(Overlapping)</i>					