CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:
21-336/21-708

<u>STATISTICAL REVIEW(S)</u>

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

STATISTICAL REVIEW AND EVALUATION
CARCINOGENICITY STUDIES

NDA /Serial Number:	NDA 21-336/N_000
Name of Drug:	EMSAM (selegiline HCl, transdermal), 20 mg
Applicant:	Somerset Pharmaceuticals
Indication:	Major Depression
Dates:	April 17, 1998, April 25, 2002, May 21, 2002
Biometrics Division:	Division of Biometrics I (HFD-710)
Statistical Reviewer:	Roswitha Kelly, M.S.
Concurring Reviewers:	Kooros Mahjoob, Ph.D., Deputy Director
Medical Division:	Neuropharmacological Drug Products (HFD-120)
Pharmacologist:	Paul Roney, Ph.D. (HFD-120)
Pharm/Tox Team Leader:	Lois Freed, Ph.D. (HFD-120)
Project Manager:	Doris Bates, Ph.D. (HFD-120)

Keywords: selegiline HCl, transdermal, carcinogenicity bioassay, intercurrent mortality, Kaplan-Meier estimates, exact permutation tests, validity of study, tumorigenic potential

Table of Contents

Table of Tables

Table of Figures

1 Executive Summary

1.1 Conclusions

In this reviewer's opinion, the validity of the rat study is in question. The treatment with selegiline HCl did not produce any increase in mortality or tumor incidences. On the contrary, the high dose animals of both genders had consistently **lower** tumor rates than any other treatment group including the controls. In addition, the mean body weights of the high dose groups were greatly reduced compared to the controls. The length of exposure and the number of animals exposed to treatment were acceptable. However, the very low mean body weights of the high dose rats may have affected tumor development and therefore the lack of any increase in tumor incidences may not reflect a true lack of carcinogenic potential of selegiline HCl.

The validity of the mouse study is even more questionable. It suffered from the same shortcomings as the rat study, namely very low tumor rates and greatly reduced mean body weights of the high dose animals of both genders compared to the controls. In addition, the study lasted only 78 weeks, which may be too short to permit formation of late developing tumors. From the statistical perspective, a carcinogenic potential of selegiline HCl cannot be ruled out despite any increases in tumor findings.

Finally, it is noted that the route of administration is not identical between the rodent bioassays (oral dietary) and the human use (transdermal) in this application.

1.2 Overview of Studies Reviewed

One rat and one mouse bioassay was reviewed. Both studies had been previously submitted to NDA 20-647. However at that time full microscopic histopathology had been done only for the control and high dose animals, for some target organs of all animals and for the low and mid dose animals dying on study. The sponsor had been requested to provide the tumor data for all tissues from all animals. In addition, the sponsor performed a peer review on some of the previous and new findings. Therefore, a new statistical review was warranted.

1.3 Principal Findings

Selegiline HCl was administered in the diet for 104 weeks to Sprague Dawley rats in doses up to 17.5mg/kg/day. Survival was not affected by the administration of the compound and no increase in tumor incidences was observed. However, selegiline had a major effect on reducing mean body weights of the high dose animals. As a matter of fact, the frequency of tumors among the high dose animals was generally lower than the frequency of any other treatment group, including the controls. Excluding the high dose rats from analysis showed numeric increases in some tumors among both the females and males, but p-values at best approached statistical significance.

4

Selegiline HCl was administered in the diet for 78 weeks to CD-1 mice in doses up to 30.0mg/kg/day. As with the rat study, survival among the mice was not affected by the administration of the compound and no increase in tumor incidences was observed. However, selegiline had a major effect on reducing mean body weights of the high dose animals. As a matter of fact, the frequency of tumors among the high dose animals was generally lower than the frequency of any other treatment group, including the controls. Excluding the high dose animals from analysis did not result in any statistically significant or approximately significant findings. Another major concern for the mouse study is its brevity. Seventy-eight weeks may have been too short a duration to allow for the formation of late developing tumors.

2 Introduction

2.1 Overview

2.1.1 Background

These carcinogenicity studies have been previously reviewed for NDA 20-647, Eldepryl, which was an oral indication for selegiline HCl for the treatment of Parkinson's Disease. At that time, the sponsor had evaluated all tissues for all control and high dose animals and for all premature decedent rats or mice of the low and mid doses. In addition, the kidney, lung and liver were examined for all animals. The statistical review of this early (paper) submission was completed in May of 1996 prior to electronic filing. It contains copies of the sponsor's bodyweight graphs, which this reviewer will cite below, but will not include into the current review.

For the submission in support of EMSAM, a transdermal application of selegiline HCL in major depression, the same carcinogenicity studies are used, however all tissues of all animals were microscopically examined and a peer review of all findings was instituted. Therefore, the study in rats had two project numbers: ── Project No. 435507 (original limited histopathological evaluation) and No. 453725 (complete histological evaluation on remaining tissues). The sponsor confirms that the electronic dataset 435507 contains the peer-reviewed results from both ── projects (435507 and 453725).

Groups of 50 male and female Sprague Dawley rats were dosed daily with selegiline HCl in the diet at levels of 0, 0.7, 3.5 and 17.5 mg/kg/day for 104 weeks. This submission presents the peer-reviewed results of the histopathological evaluation of all tissues of all dose groups.

Similarly, the carcinogenicity study in mice had two project numbers: ── Project No. 435664 (original limited histopathological evaluation) and No. 453730 (complete histological evaluation on remaining tissues). Again, the sponsor confirms that the electronic data set 435664 contains the peer-reviewed results from both projects.

Groups of 50 male and female CD-1 mice were dosed daily with selegiline HCl in the diet at levels of 0, 3, 10 and 30 mg/kg/day for 78 weeks. This submission presents the peer-reviewed results of the histopathological evaluation of all tissues of all dose groups.

2.1.2 Major Statistical Issues

The current submission contains the results of the complete microscopic histopathological examination of all tissues from all dose groups and of the peer review of the entire histology findings. The sponsor performed only two-sided Fisher's Exact tests between controls and treated groups. No trend tests, adjustment for intercurrent mortality or for multiplicity was undertaken. Using the methods applied consistently across submissions and detailed in the Draft Guidance[1], this reviewer performed trend tests for intercurrent mortality and for mortality adjusted tumor incidences. These approaches are generally more powerful to detect statistically significant findings than pair-wise comparisons of gross rates.

As there were no statistically significant increases with dose among either species or gender, the validity of each study was assessed. Sufficient numbers of animals were available for rats, but for the mice the study appears to have been too short for a whole life exposure (78 weeks only). It also became apparent that the lack of increase, or actual decrease, in tumor findings among the high dose rats and mice of either gender may be related to the strong suppression of mean body weights. Therefore, in this reviewer's opinion, the true carcinogenic potential cannot be assessed by these studies.

2.2 Data Analyzed and Sources

In their May 21, 2002 BP submission, the sponsor submitted the rat and mouse tumor data as SAS transport files according to the FDA Guideline "Guidance for Industry: Providing Regulatory Submissions in Electronic Format - General Considerations". Data set 435507 contained the complete and peer-reviewed rat carcinogenicity data and data set 435664 contained the complete and peer-reviewed mouse carcinogenicity data. The paths to each are: \\Cdsesub1\n21336\N_000\2002-05-21\FDA Datasets 435507 and \\Cdsesub1\n21336\N_000\2002-05-21\FDA Datasets 435664.

APPEARS THIS WAY
ON ORIGINAL

---

[1] Guidance for Industry: Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (1999)

3 Statistical Evaluation

3.1 Rat Carcinogenicity Study

3.1.1 Sponsor's Results and Conclusions

The sponsor did not report on intercurrent mortality. The sponsor found no increase in the incidence of neoplastic lesions in any organ of either gender of the rats when dosed up to 17.5 mg/kg/day. The extra investigations did not change the original judgement that there was no evidence of a carcinogenic effect at any of the dose levels tested in either sex. The peer review confirmed the findings recorded and found no differences between the evaluation of this extension compared to the original study.

3.1.2 Statistical Methodologies

The sponsor used Fisher's Exact test (two-tailed) comparing the control group with each of the treated groups. This approach does not follow the FDA Draft Guidance, in particular it does not adjust for intercurrent mortality or for multiplicity. This reviewer used the methods applied to all whole life rodent carcinogenicity studies as outlined in the Draft Guidance. Increasing mortality with dose was evaluated by Cox and Kruskal-Wallis tests with a two-sided $\alpha=0.05$ level of significance. Trends with dose in tumor incidences were evaluated separately for incidental and fatal tumors, or for their combination. Levels of statistical significance depended on whether the tumor was considered rare (i.e. $\leq 1.0\%$ among concurrent controls) or common. For rare (common) tumors, trends with $p \leq 0.025$ (0.005) were considered statistically significant. All tumor testing was mortality adjusted regardless of whether or not there was a statistically significant trend in mortality.

3.1.3 Detailed Review of Rat Carcinogenicity Study

This appears to be a standard 2-year bioassay in Sprague Dawley rats. The doses administered had no effect on intercurrent mortality, departure from linear trend, or homogeneity of survival curves (Tables 1,2,5,6). For neither gender did these statistics approach statistical significance. These findings were supported by the Kaplan Meier graphs of survival (Figures 1,2).

None of the tumor findings showed a statistically significant increase in tumor incidences with dose for either gender (Tables 3,7). It is noted however, that the tumor listings are ambiguous. For example, a tumor may be listed alone and also with a TH or TA modifier. Or, the modifier of the tumor, such as 'alveolar/bronchial', may be so long that the actual tumor type, such as adenoma or carcinoma had to be truncated for the tables. One needs to consult the tumor codes or the data listing for exact definitions of the findings. However, appropriate groupings of tumors at a given site are not expected to reach statistical significance. In particular, the reviewing pharmacologist requested the combination of fibroadenocarcinoma with adenocarcinoma in the mammary glands.

These tumor types were not present among the male rats and for the females the combination did not approach statistical significance (results in bold in Table 3).

It is further noted that there is a non-linear relationship between tumor incidence and dose. More specifically, the tumor findings among the high dose animals are rarer than among any other group, including the control groups. This holds for individual tissue sites as well as for the total number of tumors observed: Among the female controls there were 86 tumors reported but only 62 tumors were reported for the high dose females. Among the male controls there were a total of 49 tumors listed but only 35 for the high dose males. In case the high dose is judged too toxic, this reviewer re-analyzed the data without the high dose. No trend in tumor findings among female control, low, and mid doses reached statistical significance, including the combined fibroadenocarcinoma and adenocarcinoma of the mammary gland (Table 4). For the male rats, there emerged an approximate significant finding for histiocytic sarcoma of the lymphoreticular/haemo-poietic system (p=0.034 versus α=0.025) (Table 8). Though the asymptotic test resulted in a statistically significant finding, this approximation is less appropriate because the fatal and incidental tumors did not occur during the same time interval.

As noted in her 1996 review, the administration of selegiline HCl to Sprague Dawley rats had a strong effect on mean body weights, in particular for the high dose groups. The sponsor reported in the Eldepryl submission that mean body weights for the high dose rats were up to 30 and 37 percent less than the male and female controls, respectively. This effect may have had a major effect on tumor formation in these dose groups.

3.1.4 Validity Evaluation of the Rat Study

As there were no statistically significant tumor trends among the male or female rats in this study, the validity of both gender sub-studies needs to be assessed. Two criteria are set up for this purpose (Haseman[23], Chu et al.[4], and Bart et al.[5]):

i)   was a sufficient number of animals exposed long enough to allow for late-developing tumors, and

ii)   did the high dose provide a adequate tumor challenge?

[2] Haseman: Statistical Issues in the Design, Analysis and Interpretation of Animal Carcinogenicity Studies, *Environmental Health Perspectives*, Vol. 58, pp 385-392, 1984.

[3] Haseman: Issues in Carcinogenicity Testing: Dose Selection, *Fundamental and Applied Toxicology*, Vol. 5, pp. 66-78, 1985.

[4] Chu, Cueto, Ward: Factors in the Evaluation of 200 National Cancer Institute Carcinogenicity Bioassays, *Journal of Toxicology and Environmental Health*, Vol. 8, pp 251-280, 1981.

[5] Bart, Chu, Tarone: Statistical Issues in Interpretation of Chronic Bioassay Tests for Carcinogenicity, *Journal of the National Cancer Institute*, pp. 957-974, 1979.

The number of animals and length of exposure can be assessed at weeks 52, 80-90, and at termination, but are generally considered satisfied if 20-30 animals survive through weeks 80-90. In this study, either gender easily satisfied this criterion with at least 72% female and 66% males surviving for 104 weeks. The high dose is expected to be close to the MTD in order to provide an appropriate tumor challenge. In the absence of increased tumor rates, an effect of increased mortality or a roughly 10% decrease in mean bodyweight of the high dose group with respect to their controls are considered markers that the high dose was close to the MTD. For this study, both the high dose females and the high dose males experienced the least mortality. However, average bodyweights were clearly affected. A differential of up to 37% (females) and 30% (males) lower mean body weight of the high dose animals compared to the controls indicates that the high dose exceeded the MTD. The substantially reduced body weight may have resulted in fewer tumor formations among the high dose groups compared to the other dose groups and controls. This reviewer also performed tumor trends for the three lower groups, controls, low dose, mid dose, which did not produce any statistical signals with the possible exception of histiocytic sarcoma of the lymphoreticular/haemopoietic system among the male rats. It needs to be noted the most proper statistical result only approached statistical significance. A decision whether the findings for the high dose can be discarded and the mid dose used as an appropriate top dose group, i.e. whether this latter analysis is meaningful, is left to the expertise of the reviewing pharmacologist.

Table 1: Mortality by Time Interval of Female Rats

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR0	53-78	50	4	46	92.0	8.0
	79-91	46	2	44	88.0	12.0
	92-104	44	6	38	76.0	24.0
	FINALKILL105-105	38	38	0		
LOW	0-52	50	1	49	98.0	2.0
	53-78	49	7	42	84.0	16.0
	79-91	42	2	40	80.0	20.0
	92-104	40	4	36	72.0	28.0
	FINALKILL105-105	36	36	0		
MED	53-78	50	3	47	94.0	6.0
	79-91	47	4	43	86.0	14.0
	92-104	43	7	36	72.0	28.0
	FINALKILL105-105	36	36	0		
HIGH	0-52	50	1	49	98.0	2.0
	53-78	49	4	45	90.0	10.0
	79-91	45	2	43	86.0	14.0
	92-104	43	4	39	78.0	22.0
	FINALKILL105-105	39	39	0		

9

Table 2: Dose Mortality Trend for Female Rats

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Time-Adjusted Trend Test				
Depart from Trend	0.3728	0.8300	0.4021	0.8179
Dose-Mortality Trend	0.3289	0.5663	0.3176	0.5730
Homogeneity	0.7016	0.8728	0.7198	0.8685

Figure 1: Kaplan Meier Curves for Mortality of Female Rats
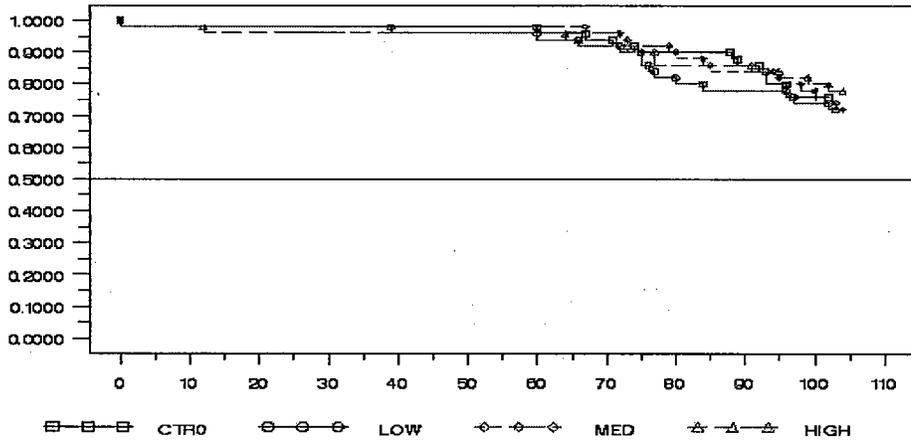
Table 3: Tumor Trends for Female Rats

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR 0	LO W	ME D	HIG H	P-Value (Exact Method)	P-Value (Asymptotic Method)	Link 2XC Table >
10	ADRENALS	14	CORTICAL ADE	1	4	1	0	0.9511	0.9453	1 *
10	ADRENALS	15	PHAEOCHROM OC	4	2	2	1	0.8746	0.8666	2 *
13	BRAIN	162	RETICULOSIS	0	0	0	1	0.2617	0.0568	3 *
13	BRAIN	51	GLIOMA	1	0	2	0	0.6980	0.7922	4 *
14	THYMUS	159	THYMOMA(TA)	0	0	0	1	0.2734	0.0623	5 *
16	PANCREAS	179	ISLET ADENOC	0	0	1	0	0.5034	0.6406	6 *
16	PANCREAS	37	ISLET ADENOM	1	1	1	0	0.8332	0.8544	7 *
16	PANCREAS	68	PANCREAS: IS	1	0	0	0	1.0000	0.8012	8
17	PARATHYRO	26	ADENOMA(TA)	0	0	0	1	0.2617	0.0568	9 *

			IDS							
19	VASCULAR SYS	77	HAEMANGIOSAR	2	0	0	0	1.0000	0.8714	10
2	THYROIDS	2	C-CELL ADENO	2	4	1	4	0.2779	0.2684	11 *
2	THYROIDS	9	FOLLICULAR A	2	0	1	0	0.9054	0.8556	12 *
20	UTERUS	71	STROMAL SARC	1	1	0	2	0.2438	0.1717	13 *
20	UTERUS	73	POLYP	8	7	4	4	0.8633	0.8630	14 *
23	LIVER	69	HEPATOCELLUL	1	1	0	0	0.9362	0.8604	15 *
24	DUODENUM	76	LEIOMYOSARCO	1	0	0	0	1.0000	0.8012	16
26	LUNGS	84	ALVEOLAR/BRO	1	0	0	0	1.0000	0.8012	17
27	LYMPHORETICU	113	LYMPHOMA	0	0	0	1	0.2617	0.0568	18 *
27	LYMPHORETICU	90	LYMPHOMA (TH	1	0	0	0	1.0000	0.8012	19
39	OVARIES	129	GRANULOSA/TH	0	1	0	0	0.7450	0.7730	20 *
39	OVARIES	185	GRANULOSA/TH	0	0	1	0	0.5034	0.6406	21 *
5	PITUITARY	26	ADENOMA(TA)	13	18	22	15	0.6058	0.6128	22 *
5	PITUITARY	40	ADENOCARCINO	2	6	8	1	0.9482	0.9399	23 *
5	PITUITARY	45	CARCINOMA, A	2	0	0	1	0.5947	0.4899	24 *
5	PITUITARY	6	ADENOMA, ANT	25	16	12	17	0.7729	0.7751	25 *
6	SKIN/SUBCUTI	138	SARCOMA(TA)	0	1	0	0	0.7450	0.7730	26 *
6	SKIN/SUBCUTI	29	INTRACUTANEO	0	3	1	0	0.8152	0.8765	27 *
6	SKIN/SUBCUTI	43	LIPOMA(TA)	0	1	1	0	0.6310	0.7696	28 *
6	SKIN/SUBCUTI	7	FIBROMA(TA)	1	2	0	1	0.6405	0.6099	29 *
6	SKIN/SUBCUTI	87	FIBROSARCOMA	1	0	0	1	0.4024	0.2413	30 *
9	MAMMARY GLAN	13	FIBROADENOMA	9	8	11	6	0.8297	0.8328	31 *
9	MAMMARY GLAN	206	ADENO/FIBRO	0	0	0	1	0.2617	0.0568	32 *
9	MAMMARY GLAN	26	ADENOMA(TA)	0	0	0	1	0.2617	0.0568	33 *
9	MAMMARY GLAN	40	ADENOCARCINO	2	5	3	1	0.8862	0.8920	34 *
9	MAMMARY GLAN	64	FIBROADENOCA	4	6	9	2	0.9237	0.9189	35 *
9	MAMMARY GLAN	7	FIBROMA(TA)	0	0	1	0	0.5238	0.6006	36 *
9	**MAMMARY GLAN**	7	**COMBINED FIBROMADENOCAR AND ADENOCARC**	6	10	11	3	**0.9693**	**0.9640**	36 *

Table 4: Tumor Trends for Female Rats with High Dose Excluded

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR 0	LOW	MED	P-Value (Exact Method)	P-Value (Asymptotic Method)
10	ADRENALS	14	CORTICAL ADENOMA(TA)	1	4	1	0.6867	0.7509
10	ADRENALS	15	PHAEOCHROMOCYTOMA(TA)	4	2	2	0.7311	0.7371
13	BRAIN	51	GLIOMA	1	0	2	0.2680	0.1956
16	PANCREAS	179	ISLET ADENOCARCINOMA(TA)	0	0	1	0.3273	0.1409
16	PANCREAS	37	ISLET ADENOMA(TA)	1	1	1	0.5824	0.5649
16	PANCREAS	68	PANCREAS: ISLET ADENOMA(TA)	1	0	0	1.0000	0.8928
19	VASCULAR SYSTEM	77	HAEMANGIOSARCOMA	2	0	0	1.0000	0.9381
2	THYROIDS	2	C-CELL ADENOMA(TA)	2	4	1	0.7991	0.8375
2	THYROIDS	9	FOLLICULAR ADENOMA(TA)	2	0	1	0.8021	0.7581
20	UTERUS	71	STROMAL SARCOMA(TA)	1	1	0	0.8828	0.8875
20	UTERUS	73	POLYP	8	7	4	0.9113	0.9131
23	LIVER	69	HEPATOCELLULAR ADENOMA(TA)	1	1	0	0.8827	0.8848
24	DUODENUM	76	LEIOMYOSARCOMA(TA)	1	0	0	1.0000	0.8928
26	LUNGS	84	ALVEOLAR/BRONCHIOLAR ADENOMA(T	1	0	0	1.0000	0.8928
27	LYMPHORETICULAR/HAEMOPOIETIC T	90	LYMPHOMA (THYMIC)	1	0	0	1.0000	0.8928
39	OVARIES	129	GRANULOSA/THECAL CELL TUMOUR(S	0	1	0	0.6545	0.7817
39	OVARIES	185	GRANULOSA/THECAL CELL TUMOUR	0	0	1	0.3273	0.1409
5	PITUITARY	26	ADENOMA(TA)	13	18	22	0.0725	0.0731
5	PITUITARY	40	ADENOCARCINOMA(TA)	2	6	8	0.0550	0.0564
5	PITUITARY	45	CARCINOMA, ANTERIOR LOBE	2	0	0	1.0000	0.9367
5	PITUITARY	6	ADENOMA, ANTERIOR LOBE	25	16	12	0.9940	0.9938
6	SKIN/SUBCUTIS	138	SARCOMA(TA)	0	1	0	0.6545	0.7817
6	SKIN/SUBCUTIS	29	INTRACUTANEOUS CORNIFYING EPIT	0	3	1	0.4053	0.5170
6	SKIN/SUBCUTIS	43	LIPOMA(TA)	0	1	1	0.3213	0.3273
6	SKIN/SUBCUTIS	7	FIBROMA(TA)	1	2	0	0.8436	0.8933
6	SKIN/SUBCUTIS	87	FIBROSARCOMA(TA)	1	0	0	1.0000	0.8928
9	MAMMARY GLANDS	13	FIBROADENOMA(TA)	9	8	11	0.3174	0.3256
9	MAMMARY GLANDS	40	ADENOCARCINOMA(TA)	2	5	3	0.5637	0.6047
9	MAMMARY GLANDS	64	FIBROADENOCARCINOMA(TA)	4	6	9	0.1240	0.1296
9	MAMMARY GLANDS	7	FIBROMA(TA)	0	0	1	0.4118	0.1927
9	MAMMARY GLANDS	COM	COMBINED FIBROADENOCARCINOMA AND ADENOCARCINOMA	6	10	11	0.2396	0.2481

Table 5: Mortality by Time Interval for Male Rats

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR0	53-78	50	5	45	90.0	10.0
	79-91	45	1	44	88.0	12.0
	92-104	44	9	35	70.0	30.0
	FINALKILL105-105	35	35	0		
LOW	53-78	50	4	46	92.0	8.0
	79-91	46	4	42	84.0	16.0
	92-104	42	9	33	66.0	34.0
	FINALKILL105-105	33	33	0		
MED	0-52	50	2	48	96.0	4.0
	53-78	48	4	44	88.0	12.0
	79-91	44	5	39	78.0	22.0
	92-104	39	4	35	70.0	30.0
	FINALKILL105-105	35	35	0		
HIGH	0-52	50	1	49	98.0	2.0
	53-78	49	2	47	94.0	6.0
	92-104	47	8	39	78.0	22.0
	FINALKILL105-105	39	39	0		

Table 6: Dose Mortality Trend for Male Rats

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Time-Adjusted Trend Test				
Depart from Trend	0.2325	0.8903	0.2521	0.8815
Dose-Mortality Trend	1.8541	0.1733	2.1403	0.1435
Homogeneity	2.0866	0.5546	2.3924	0.4950

13

Figure 2: Kaplan Meier Curves for Mortality of Male Rats
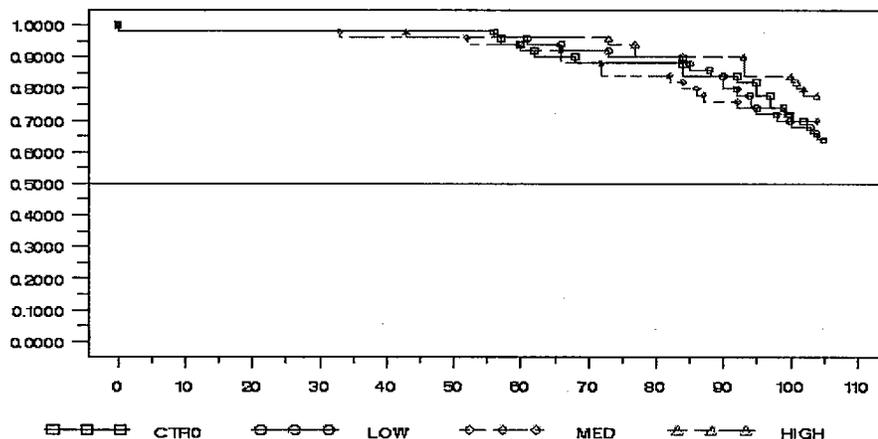
Table 7: Tumor Trends for Male Rats

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR 0	LO W	ME D	HIG H	P-Value (Exact Method)	P-Value (Asymptotic Method)	Link 2XC Table >
1	ZYMBAL'S GLA	1	SQUAMOUS-CEL	1	0	0	0	1.0000	0.7566	1
10	ADRENALS	14	CORTICAL ADE	2	0	0	1	0.6214	0.5157	2 *
10	ADRENALS	15	PHAEOCHROMOC	13	18	8	7	0.9939	0.9926	3 *
12	TESTES	20	INTERSTITIAL	9	7	4	6	0.7438	0.7506	4 *
13	BRAIN	121	GRANULAR CEL	0	1	0	0	0.7535	0.7799	5 *
13	BRAIN	162	RETICULOSIS	0	0	1	1	0.2027	0.1885	6 *
13	BRAIN	25	GRANULAR CEL	1	0	0	0	1.0000	0.8003	7
13	BRAIN	51	GLIOMA	1	2	1	1	0.6351	0.6612	8 *
14	THYMUS	123	THYMOMA	0	1	0	0	0.7328	0.7785	9 *
14	THYMUS	159	THYMOMA(TA)	0	0	1	0	0.5556	0.2963	10 *
16	PANCREAS	37	ISLET ADENOM	3	5	2	2	0.7834	0.7844	11 *
16	PANCREAS	57	EXOCRINE ADE	1	0	1	0	0.7725	0.8003	12 *
17	PARATHYROIDS	143	ADENOMA	0	0	1	0	0.5211	0.6510	13 *
17	PARATHYROIDS	26	ADENOMA(TA)	2	0	0	4	0.0483	0.0246	14 *
19	VASCULAR SYS	58	HAEMANGIOMA	1	0	0	1	0.4753	0.3019	15 *
19	VASCULAR SYS	77	HAEMANGIOSAR	0	0	2	0	0.5347	0.6922	16 *

2	THYROIDS	103	FOLLICULAR C	0	1	0	0	0.7535	0.7799	17 *
2	THYROIDS	105	C-CELL ADENO	0	2	1	0	0.7616	0.8517	18 *
2	THYROIDS	150	C-CELL CARCI	0	0	1	0	0.5211	0.6510	19 *
2	THYROIDS	2	C-CELL ADENO	2	5	2	1	0.8728	0.8735	20 *
2	THYROIDS	22	FOLLICULAR A	1	0	0	0	1.0000	0.8072	21
2	THYROIDS	47	C-CELL CARCI	2	1	1	1	0.6938	0.6774	22 *
2	THYROIDS	9	FOLLICULAR A	2	0	1	0	0.6073	0.5776	23 *
23	LIVER	194	CHOLANGIO CAR	0	0	0	1	0.2746	0.0629	24 *
23	LIVER	69	HEPATOCEL LUL	0	1	1	1	0.3039	0.3704	25 *
23	LIVER	93	HEPATOCEL LUL	0	3	0	0	0.9044	0.8974	26 *
26	LUNGS	142	ALVEOLAR/B RO	0	0	1	0	0.5211	0.6510	27 *
27	LYMPHORET ICU	113	LYMPHOMA	0	1	0	0	0.7535	0.7799	28 *
27	LYMPHORET ICU	152	HISTIOCYTIC	0	0	4	1	0.4468	0.5020	29 *
37	HEART	138	SARCOMA(T A)	0	0	0	1	0.2746	0.0629	30 *
41	SEMINAL VESI	26	ADENOMA(T A)	0	0	1	0	0.5000	0.2566	31 *
42	HEAD	160	AMELOBLAS TOM	0	0	1	0	0.5211	0.6510	32 *
44	KIDNEYS	190	TUBULAR ADEN	0	0	0	1	0.2746	0.0629	33 *
45	BONE	196	OSTEOSARC OMA	0	0	0	1	0.2746	0.0629	34 *
5	PITUITARY	26	ADENOMA(T A)	3	7	8	1	0.9754	0.9675	35 *
5	PITUITARY	32	ADENOMA, INT	2	1	0	1	0.7119	0.6516	36 *
5	PITUITARY	40	ADENOCAR CINO	2	0	0	1	0.6101	0.5049	37 *
5	PITUITARY	45	CARCINOMA , A	1	0	0	0	1.0000	0.8072	38
5	PITUITARY	6	ADENOMA, ANT	30	29	20	23	0.9581	0.9575	39 *
6	SKIN/SUBCU TI	111	CARCINOMA (TA	0	1	0	0	0.7000	0.7562	40 *
6	SKIN/SUBCU TI	115	BASAL CELL T	0	1	1	0	0.6501	0.7807	41 *
6	SKIN/SUBCU TI	118	MYXOMA(TA)	0	1	0	0	0.6667	0.7165	42 *
6	SKIN/SUBCU TI	138	SARCOMA(T A)	0	0	1	1	0.1373	0.0448	43 *
6	SKIN/SUBCU TI	158	SEBACEOUS AD	0	0	1	0	0.5211	0.6510	44 *
6	SKIN/SUBCU TI	18	SCHWANNO MA(T	1	0	0	0	1.0000	0.8072	45
6	SKIN/SUBCU TI	29	INTRACUTA NEO	8	11	5	1	0.9994	0.9980	46 *
6	SKIN/SUBCU TI	34	BASAL-CELL C	1	1	0	0	0.9406	0.8676	47 *

15

6	SKIN/SUBCU TI	43	LIPOMA(TA)	3	1	4	2	0.5968	0.6406	48 *
6	SKIN/SUBCU TI	7	FIBROMA(TA)	5	4	6	1	0.9675	0.9616	49 *
6	SKIN/SUBCU TI	87	FIBROSARCOMA	0	1	0	1	0.4343	0.3534	50 *
6	SKIN/SUBCU TI	96	PAPILLOMA(TA)	0	1	2	0	0.7667	0.8494	51 *
9	MAMMARY GLAN	13	FIBROADENOMA	1	2	0	0	0.9440	0.9070	52 *
9	MAMMARY GLAN	26	ADENOMA(TA)	0	0	0	1	0.2746	0.0629	53 *

Table 8: Tumor Trends for Male Rats Excluding High Dose

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR 0	LOW	MED	P-Value (Exact Method)	P-Value (Asymptotic Method)
1	ZYMBAL'S GLAND	1	SQUAMOUS-CELL CARCINOMA(TA)	1	0	0	1.0000	0.8839
10	ADRENALS	14	CORTICAL ADENOMA(TA)	2	0	0	1.0000	0.9396
10	ADRENALS	15	PHAEOCHROMOCYTOMA(TA)	13	18	8	0.9756	0.9754
12	TESTES	20	INTERSTITIAL CELL TUMOUR	9	7	4	0.9377	0.9373
13	BRAIN	121	GRANULAR CELL TUMOUR	0	1	0	0.6602	0.7871
13	BRAIN	162	RETICULOSIS	0	0	1	0.3398	0.1489
13	BRAIN	25	GRANULAR CELL TUMOUR(S)	1	0	0	1.0000	0.8956
13	BRAIN	51	GLIOMA	1	2	1	0.5829	0.6373
14	THYMUS	123	THYMOMA	0	1	0	0.6277	0.7696
14	THYMUS	159	THYMOMA(TA)	0	0	1	0.5556	0.2963
16	PANCREAS	37	ISLET ADENOMA(TA)	3	5	2	0.7897	0.8150
16	PANCREAS	57	EXOCRINE ADENOMA(TA)	1	0	1	0.5663	0.4679
17	PARATHYROIDS	143	ADENOMA	0	0	1	0.3398	0.1489
17	PARATHYROIDS	26	ADENOMA(TA)	2	0	0	1.0000	0.9396
19	VASCULAR SYSTEM	58	HAEMANGIOMA	1	0	0	1.0000	0.8954
19	VASCULAR SYSTEM	77	HAEMANGIOSARCOMA	0	0	2	0.1133	0.0434
2	THYROIDS	103	FOLLICULAR CELL CARCINOMA	0	1	0	0.6602	0.7871
2	THYROIDS	105	C-CELL ADENOMA	0	2	1	0.3703	0.4757
2	THYROIDS	150	C-CELL CARCINOMA	0	0	1	0.3398	0.1489
2	THYROIDS	2	C-CELL ADENOMA(TA)	2	5	2	0.6920	0.7327
2	THYROIDS	22	FOLLICULAR ADENOCARCINOMA(TA)	1	0	0	1.0000	0.8954
2	THYROIDS	47	C-CELL CARCINOMA(TA)	2	1	1	0.7132	0.6932
2	THYROIDS	9	FOLLICULAR ADENOMA(TA)	2	0	1	0.7166	0.6826
23	LIVER	69	HEPATOCELLULAR ADENOMA(TA)	0	1	1	0.3331	0.3420
23	LIVER	93	HEPATOCELLULAR CARCINOMA(TA)	0	3	0	0.7475	0.8439
26	LUNGS	142	ALVEOLAR/BRONCHIOLA	0	0	1	0.3398	0.1489

			R CARCINOMA					
27	LYMPHORETICULAR/HAEMOPOIETIC T	113	LYMPHOMA	0	1	0	0.6602	0.7871
27	LYMPHORETICULAR/HAEMOPOIETIC T	152	HISTIOCYTIC SARCOMA(TA)	0	0	4	0.0340	0.0118
41	SEMINAL VESICLE(S)	26	ADENOMA(TA)	0	0	1	0.5000	0.2566
42	HEAD	160	AMELOBLASTOMA(TA)	0	0	1	0.3398	0.1489
5	PITUITARY	26	ADENOMA(TA)	3	7	8	0.1174	0.1215
5	PITUITARY	32	ADENOMA, INTERMEDIATE LOBE	2	1	0	0.9630	0.9390
5	PITUITARY	40	ADENOCARCINOMA(TA)	2	0	0	1.0000	0.9339
5	PITUITARY	45	CARCINOMA, ANTERIOR LOBE	1	0	0	1.0000	0.8954
5	PITUITARY	6	ADENOMA, ANTERIOR LOBE	30	29	20	0.9925	0.9924
6	SKIN/SUBCUTIS	111	CARCINOMA(TA)	0	1	0	0.5909	0.7176
6	SKIN/SUBCUTIS	115	BASAL CELL TUMOUR	0	1	1	0.3331	0.3420
6	SKIN/SUBCUTIS	118	MYXOMA(TA)	0	1	0	0.6154	0.7668
6	SKIN/SUBCUTIS	138	SARCOMA(TA)	0	0	1	0.5000	0.2566
6	SKIN/SUBCUTIS	158	SEBACEOUS ADENOMA(TA)	0	0	1	0.3398	0.1489
6	SKIN/SUBCUTIS	18	SCHWANNOMA(TA)	1	0	0	1.0000	0.8954
6	SKIN/SUBCUTIS	29	INTRACUTANEOUS CORNIFYING EPIT	8	11	5	0.8861	0.8905
6	SKIN/SUBCUTIS	34	BASAL-CELL CARCINOMA(TA)	1	1	0	0.8867	0.8898
6	SKIN/SUBCUTIS	43	LIPOMA(TA)	3	1	4	0.1776	0.1683
6	SKIN/SUBCUTIS	7	FIBROMA(TA)	5	4	6	0.2757	0.2799
6	SKIN/SUBCUTIS	87	FIBROSARCOMA(TA)	0	1	0	0.6602	0.7871
6	SKIN/SUBCUTIS	96	PAPILLOMA(TA)	0	1	2	0.3331	0.3420
9	MAMMARY GLANDS	13	FIBROADENOMA(TA)	1	2	0	0.8520	0.8995

3.2 Mouse Carcinogenicity Study

3.2.1 Sponsor's Results and Conclusions

The extra investigations did not change the original judgement that there was no evidence of a carcinogenic effect at any of the dose levels tested in either sex. The peer review confirmed the findings recorded and found no differences between the evaluation of this extension compared to the original study. The sponsor did not report on intercurrent mortality. The sponsor noted the decrease in mean body weights particularly for the high dose animals, but found no evidence of neoplastic findings attributable to dosing with selegiline HCl at doses up to 30 mg/kg/day.

3.2.2 Statistical Methodologies

The sponsor used Fisher's Exact tests (two-tailed) comparing the control group with each of the treated groups. This approach does not follow the FDA Draft Guidance, in particular, it does not adjust for intercurrent mortality or multiplicity, nor did it test for linear trend with dose. This reviewer used the methods applied to all whole life rodent carcinogenicity studies, as outlined in the Draft Guidance. As with the rat study, increasing mortality with dose was evaluated by Cox and Kruskal-Wallis tests with a two-sided $\alpha=0.05$ level of significance. Trends with dose in tumor incidences were evaluated separately for incidental and fatal tumors, or for their combination. Levels of statistical significance depended on whether the tumor was considered rare (i.e. $\leq 1.0\%$) or common ($\alpha= 0.025$ or 0.005, respectively). All tumor testing was mortality adjusted regardless of whether or not there was a statistically significant trend in mortality.

3.2.3 Detailed Review of Mouse Carcinogenicity Study

This study was conducted for only 78 weeks, which is unusually short. The doses of 0, 3, 10, and 30 mg/kg/day in the diet had no effect on mortality of either the female or male mice (Tables 9,10,12,13). Neither tests for linear trend, departure from linear trend, nor test for homogeneity approached statistical significance. These findings are visually apparent in the Kaplan Meier curves (Figures 3,4).

None of the observed tumor findings suggested a linear increase with dose for either gender (Tables 11 and 14). The same comments about the ambiguity of the tumor listing as given above about the rat data apply to the mouse data. Again, it appears that appropriate groupings of tumors would not lead to statistically significant findings. In fact, for each tissue, with the exception when the high dose has the only single tumor for a tissue, no high dose tumor frequency exceeded the largest frequency observed for the other treatment groups. For the female high dose group the total number of observed tumors is 25, for the female controls it is 29. For the male high dose group, the total number of observed tumors is 18, for the male controls it is 28. Therefore, the phenomenon observed among the rats, that the total number of tumors for high dose

animals is lower than the total number of tumors for the controls is also seen among the mice. The reviewing pharmacologist requested to combine hepatocellular adenomas and carcinomas. The results are given in bold in Tables 11 and 14. For the female mice, there were only two tumor bearing animals, one in the control and one in the high dose, and their combination did not approach statistical significance. For the male mice there were 4, 6, 7, 1 hepatocellular adenoma or carcinoma bearing animals in the control, low, mid, and high dose respectively. Of course the trend with increasing dose was not close to statistical significance. As with the rats, this reviewer considered excluding the high dose in an analysis to evaluate the unusually low tumor manifestation in this group. Visual inspection showed that no trend would reach statistical significance. Even the p-value for the combined hepatocellular adenoma and carcinoma did not approach statistical significance when the high dose was excluded.

3.2.4 Validity Evaluation of the Mouse Study

As there were no statistically increased tumor findings, the validity of both the female and the male mouse study needs to be examined. This is done based on the same criteria outlined above for the rats. Briefly, the questions are: were sufficient numbers of animals exposed for a sufficient length of time and was the tumor challenge adequate? Both genders had excellent survival at week 78 (at least 80% for the females and 76% for the males). However, this was the end of the study and may not be a sufficient representation of a whole life exposure, to allow development of tumors, which appear late in the animals' life. With respect to the adequacy of the dosing, the high dose did not influence intercurrent mortality. However, mean body weights of the female and male mice were up to 46 and 48% lower than their respective controls. As with the rat study, these observations indicate that the MTD was exceeded and that the extremely low body weights may have been responsible for the suppressed tumor appearance in the high dose groups. Therefore, the true carcinogenic potential of the compound cannot be evaluated by this study.

Table 9: Mortality by Time Interval for Female Mice

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR0	0-52	50	1	49	98.0	2.0
	53-78	49	8	41	82.0	18.0
	FINALKILL 79-79	41	41	0		
LOW	0-52	50	2	48	96.0	4.0
	53-78	48	8	40	80.0	20.0
	FINALKILL 79-79	40	40	0		
MED	0-52	50	5	45	90.0	10.0
	53-78	45	3	42	84.0	16.0
	FINALKILL 79-79	42	42	0		
HIGH	0-52	49	3	46	93.9	6.1
	53-78	46	4	42	85.7	14.3
	FINALKILL 79-79	42	42	0		

Table 10: Dose Mortality Trend for Female Mice

| | Method | | | |
| | Cox | | Kruskal-Wallis | |
	Statistics	P-Value	Statistics	P-Value
Time-Adjusted Trend Test				
Depart from Trend	0.1593	0.9234	0.1733	0.9170
Dose-Mortality Trend	0.4382	0.5080	0.4117	0.5211
Homogeneity	0.5975	0.8970	0.5851	0.8998

Figure 3: Kaplan Meier Curves for Mortality of Female Mice
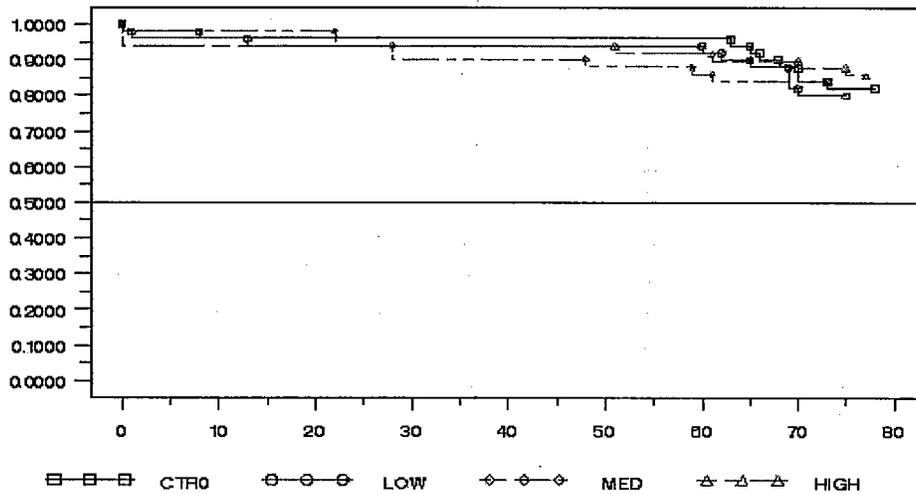
Table 11: Tumor Trends for Female Mice

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR0	LOW	MED	HIGH	P-Value (Exact Method)	P-Value (Asymptotic Method)	Link 2XC Table >
1	LUNGS	1	ALVEOLAR/BRO	10	8	11	9	0.5353	0.5415	1 *
1	LUNGS	7	ALVEOLAR/BRO	2	0	0	0	1.0000	0.9113	2
11	LYMPHORETICU	22	LYMPHOMA	7	4	4	5	0.6229	0.6325	3 *
11	LYMPHORETICU	90	HISTIOCYTIC	0	0	0	1	0.2545	0.0566	4 *
12	VASCULAR SYS	23	HAEMANGIOMA(	0	0	0	1	0.2545	0.0566	5 *

20

12	VASCULAR SYS	74	HAEMANGIOSAR	0	0	0	1	0.2545	0.0566	6 *
13	UTERUS	30	POLYP	3	0	0	2	0.4225	0.4242	7 *
13	UTERUS	43	LEIOMYOMA(TA	1	0	0	1	0.4455	0.3212	8 *
13	UTERUS	46	SARCOMA(TA)	1	0	0	0	1.0000	0.8345	9
17	OVARIES	37	CYSTADENOMA	2	0	0	0	1.0000	0.9113	10
17	OVARIES	60	LUTEOMA	0	1	0	0	0.7515	0.7632	11 *
19	MAMMARY GLAN	39	ADENOCARCINO	2	0	0	1	0.5502	0.4976	12 *
2	LIVER	35	HEPATOCELLUL	1	0	0	0	1.0000	0.8345	13
2	LIVER	85	HEPATOCELLUL	0	0	0	1	0.2545	0.0566	14 *
2	LIVER	35+85	HEPATOCELL ADENO +CARC	1	0	0	1	0.4455	0.3212	
29	SKIN/SUBCUTI	46	SARCOMA(TA)	0	1	0	0	0.6522	0.6813	15 *
43	THYROIDS	77	FOLLICULAR A	0	0	0	1	0.2545	0.0566	16 *
45	PITUITARY	87	ADENOMA, ANT	0	0	0	1	0.2545	0.0566	17 *
46	VAGINA	91	PAPILLOMA(TA	0	0	0	1	0.2545	0.0566	18 *

Table 12: Mortality by Time Interval for Male Mice

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR0	0-52	50	2	48	96.0	4.0
	53-78	48	3	45	90.0	10.0
	FINALKILL 79-79	45	45	0		
LOW	0-52	50	2	48	96.0	4.0
	53-78	48	10	38	76.0	24.0
	FINALKILL 79-79	38	38	0		
MED	0-52	50	3	47	94.0	6.0
	53-78	47	5	42	84.0	16.0
	FINALKILL 79-79	42	42	0		
HIGH	0-52	50	1	49	98.0	2.0
	53-78	49	6	43	86.0	14.0
	FINALKILL 79-79	43	43	0		

Table 13: Dose Mortality Trend for Male Mice

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Time-Adjusted Trend Test Depart from Trend	3.7708	0.1518	3.6738	0.1593
Dose-Mortality Trend	0.1098	0.7403	0.1298	0.7186
Homogeneity	3.8806	0.2746	3.8036	0.2835

Figure 4: Kaplan Meier Curves for Mortality of Male Mice

Figure 4: Kaplan Meier Curves for Mortality of Male Mice
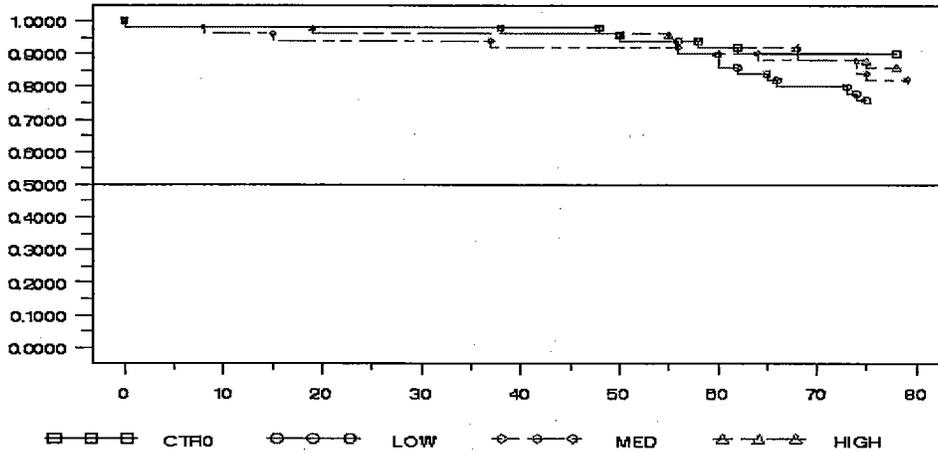
Table 14: Tumor Trends for Male Mice

Organ Code	Organ Name	Tumor Code	Tumor Name	CTR 0	LO W	ME D	HIG H	P-Value (Exact Method)	P-Value (Asymptotic Method)	Link 2XC Table >
1	LUNGS	1	ALVEOLAR/B RO	17	9	14	12	0.6978	0.7026	1 *
1	LUNGS	7	ALVEOLAR/B RO	2	2	2	0	0.9236	0.9176	2 *
11	LYMPHORET ICU	22	LYMPHOMA	2	0	1	2	0.3068	0.2850	3 *
12	VASCULAR SYS	23	HAEMANGIO MA(	1	0	0	0	1.0000	0.8319	4
12	VASCULAR SYS	74	HAEMANGIO SAR	0	0	0	1	0.2560	0.0572	5 *
2	LIVER	18	HEPATOCEL LUL	1	0	2	0	0.6975	0.7398	6 *
2	LIVER	4	HEPATOCEL LUL	4	6	7	1	0.9554	0.9487	7 *
2	LIVER	4+18	HEPATOCEL L ADE AND CARC	4	6	7	1	0.9554	0.9487	
2	LIVER	4+18	HEPATOCEL L ADE AND CARC	4	6	7		0.1963	0.1843	
21	*LYMPHORET I	22	LYMPHOMA	0	1	0	0	0.7444	0.7643	8 *
24	* LUNGS	7	ALVEOLAR/B RO	0	2	0	0	0.8116	0.8355	9 *
26	HARDERIAN GL	64	ADENOMA(T A)	0	0	2	0	0.4940	0.5529	10 *
27	*HARDERIAN	37	CYSTADENO MA	0	1	0	0	0.7459	0.7635	11 *
32	PLEURA	70	MESOTHELI OMA	0	0	1	0	0.5060	0.5457	12 *

22

43	THYROIDS	77	FOLLICULAR A	0	0	0	1	0.2560	0.0572	13 *
44	TESTES	82	INTERSTITIA L	0	0	0	1	0.2560	0.0572	14 *
8	KIDNEYS	15	TUBULAR ADEN	1	0	0	0	1.0000	0.8319	15

4 Conclusions

4.1 Statistical Evaluation of Evidence

There were no statistically significant increases in mortality-adjusted tumor findings among either the male or female rats. The statistical evaluation of the validity of either gender showed that the compound had no effect on mortality and that there were sufficient numbers of animals of either gender exposed to the compound long enough to show late developing tumors. However, the strong effect of selegiline HCl on reducing the average bodyweights of the high dose animals may have contributed to or caused the observed reduction in tumors among these animals. Therefore, based on these statistical considerations this study cannot assess the potential tumorigenicity of selegiline HCl.

The findings among the mice with respect to intercurrent mortality, tumor incidences and suppressed mean body weights were similar to those observed in the rat study. I.e., neither mortality nor tumor findings showed any statistically significant increases and the mean body weights of the high dose animals were greatly reduced compared to the controls. However for the mouse study, the lack of increase in any tumor incidence is not only confounded by the very low mean body weights of the high dose animals but also by the fact that the study lasted only 78 weeks, which does not represent a whole-life exposure. Therefore, though no statistically significant tumor increases were observed, this study is even weaker than the rat study and in this reviewer's opinion cannot rule out a tumorigenic potential of selegiline HCl in the doses administered.

4.2 Conclusions and Recommendations

With respect to the rat study, this reviewer agrees with the sponsor that there were no statistically significant increases in tumor findings with dose for either gender. The compound was not toxic with respect to mortality and sufficient numbers of animals lived long enough to manifest late developing tumors. However, the strong effect of the high dose on average body weights could mask a true tumorigenic effect of the compound.

With respect to the mouse study, this reviewer again agrees with the sponsor that there were no statistically significant increases in tumor findings with dose for either gender. The compound was not toxic with respect to mortality and sufficient numbers of animals were available at the end of the study. However, the study lasted only 78 weeks. Therefore it may not have been sufficiently long to permit the development of late developing tumors. In addition, the strong effect of the high dose on average body weights could have further masked any true tumorigenicity of the compound.

From the statistical perspective, it appears that neither study can provide sufficient evidence to rule out any tumorigenic potential of selegiline HCl at the doses administered.

It is also noted, that these carcinogenicity studies have different routes of administration than the human application. The bioassays were oral dietary studies whereas the human use for this indication will be transdermal. Whether any conclusions based on a dietary study can be transferred to a transdermal application is left to the expertise of the reviewing pharmacologist.

APPEARS THIS WAY
ON ORIGINAL

24

```
/s/
----------------------
```

Roswitha Kelly
1/29/04 09:30:14 AM
BIOMETRICS

Kooros, the PDUFA date is tomorrow, it would be greatly appreciated if you can sign off today.
Thanks!

Kooros Mahjoob
1/30/04 10:48:32 AM
BIOMETRICS

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	21- 336 / N-000 and 21-708 / N-000
Drug Name:	Selegiline Transdermal System
Indication(s):	Acute Therapy and Maintenance for Depression
Applicant:	Somerset
Date:	Submitted: 8/01/2003
Review Priority:	Resubmission (6 months)
Biometrics Division:	Division of Biometrics I, HFD-710
Statistical Reviewer:	Tristan Massie, Ph.D.
Concurring Reviewers:	Kun Jin, Ph.D.,Team Leader
	Kooros Mahjoob, Ph.D., Deputy Director
Medical Division:	Neuropharmacological Drug Products, HFD 120
Clinical Team:	Greg Dubitsky, M.D., Clinical Reviewer
Project Manager:	Doris Bates, Ph.D.

Table Of Contents

List of Tables

Statistical Review and Evaluation

1. Executive Summary

1.1. Conclusions and Recommendations
A 6-week acute study reviewed in the original NDA submission and a new 8-week acute study seem to have provided sufficient evidence of efficacy for the acute therapy for depression claim.

The long term maintenance study of the efficacy of STS for preventing relapse/reappearance of depression is positive at face value. However, a significantly higher proportion of STS patients had "abnormal" HAM-D$_{1-17}$ and CGI-s values (the basis for relapse) and then ceased participation without having a confirmation visit two weeks later, as required by the protocol. Because their "abnormal" values were not confirmed these patients were treated as not-relapsed in the sponsor's analysis. This means that the proportions relapsed and the difference in group proportions relapsed are probably biased in favor of the STS group. In fact, the results are no longer significant if all of these patients truly relapsed. However, the study is still considered positive since most reasonable attempts to re-classify these patients based on reasons for withdrawal, including lack of efficacy and relapsed in the opinion of the investigator, still produce significant results.

1.2. Brief Overview of Clinical Studies
NDA 21708 is a new application for a long-term maintenance of depression claim. NDA 21336 is a resubmission for acute treatment of major depression. It was resubmitted because the original NDA had several deficiencies. One of these deficiencies was that of the four acute treatment studies submitted in the NDA only one was positive, S9303-E106-96B. Study S9303-E106-96B was a 6-week, double-blind, randomized, placebo-controlled, parallel assessment of the safety and efficacy of the selegiline transdermal system (STS) (20 mg /20cm^2) for major depression. One hundred and seventy seven patients were randomized among six centers. The majority of patients completed the study (88.8% for STS and 83% for placebo group). The two groups were compared with respect to the change from baseline in the HAM-D$_{1-17}$ at the end of week 6 (or last available observation) using an analysis of covariance. The statistical review of study S9303-E106-96B and the others in the first submission was done by Dr. Yuan-li Shen. Dr. Shen's review is dated 01-18-2002.

The key studies in this new submission are Study 0052 an acute depression study and Study 9806 a long term maintenance study. Study 0052 was a flexible dose, double-blind, placebo-controlled, multicenter, parallel-group randomized design. Two hundred and sixty five patients were randomized among 3 centers. Patients initiated treatment with a 20 mg/ 20cm^2 STS

(Selegiline patch) or matching placebo. After two weeks the dose could be increased to 30 mg/ 30cm^2 if in the investigator's opinion the patient did not show improvement. Likewise, after 5 weeks the investigator could increase the dose to 30 mg/ 30cm^2 or matching placebo. The investigators performed the HAM-D$_{1-28}$ and other assessments at baseline, week 1, 2, 3, 5, and 8. The two groups were compared with respect to the change from baseline in HAM-D$_{1-28}$ after 8 weeks of double-blind flexible-dose treatment (or last observation) using an analysis of covariance model.

Study 9806 was a randomized, double-blind, placebo-controlled, parallel group study designed to assess the efficacy of selegiline 20 mg administered once daily via the STS for 52 weeks following a 10-week, open-label, remission, lead-in phase in patients with major depression. Patients who had a HAM-D$_{1-17}$ <= 10 at the end of week 6 or 7 and at the end of week 8 in the 10 week open-label phase were considered to have responded and were randomized to 52 weeks of treatment with 20 mg/ 20cm^2 STS or matching placebo. Patients were to have HAM-D$_{1-17}$ and CGI-s assessments taken at baseline and the end of Weeks 2, 4, 6, 8, 10, 12, 14, 18, 22, 26, 34, 42, and 52. Patients who had a HAM-D$_{1-17}$ >= 14 and a CGI-s 2 points above their baseline CGI-s (double-blind phase) at a scheduled visit were instructed to return 2 weeks later to confirm the "abnormal" HAM-D$_{1-17}$ and CGI-s. If the measurements taken two weeks later still exceeded the critical levels just mentioned and the DSM-IV criteria for depression were also confirmed then the patients were considered relapsed and their double-blind study treatment was ceased. Otherwise, the patient continued study treatment and resumed the regular study visit schedule. The two groups were compared with respect to the proportion relapsed at or before week 52, using a center-stratified Cochran-Mantel-Haenszel test. A Cox Proportional Hazards model for the time until relapse, adjusting for treatment, was specified as a secondary analysis.

1.3. Statistical Issues and Findings

In acute treatment study 0052 132 patients were randomized to STS and 133 were randomized to placebo. Since 3 STS and 5 placebo patients did not have any post-baseline HAM-D$_{1-28}$ measures they were excluded from the primary analysis. The mean change from baseline in HAM-D$_{1-28}$ at week 8 (or last available observation) was -11.1 for STS and -8.9 for placebo based on the Modified ITT population, i.e., all patients who had at least one post-baseline HAM-D$_{1-28}$. The STS group was found to be significantly more improved (p = 0.0327) based on the primary analysis: an analysis of covariance of the week 8 (or last available) change scores in the M-ITT[1] population with adjustments for baseline score, center effects, and treatment groups.

---

1 The modified ITT population is the group of all randomized patients with at least one post-baseline HAM-D$_{1-28}$ score.

The study was conducted in only three centers. Center 1 had a noticeably larger difference in treatment group mean changes in HAM-D_{1-28} than the other two centers (3.6 points vs. 1.3 and 1.7) but all centers were numerically better for STS.

The sponsor reported that 110 of 128 (86.0%) Placebo and 100 of 129 (77.5%) STS patients completed the study. The mean change from baseline for STS patients who withdrew (-5.1) was numerically worse than that for Placebo patients who withdrew (-5.8). Therefore, it is not surprising that the group difference is more significant for the Observed Cases analysis which excludes the patients who withdrew. Since the dropout rate was not excessive and both the LOCF and the Observed Cases analyses indicate a significant treatment effect we conclude that there is no obvious bias in the results caused by the dropouts and that the study is positive with regard to efficacy.

In study 9806 163 patients were randomized to placebo and 159 to STS 20 mg/ 20 cm^2. Ten STS patients had no post-baseline efficacy measures and so they were not included in the analysis. According to the protocol patients were to be considered relapsed if they had a HAM-D_{1-17} >=14 and a CGI-s at least 2 points above their baseline CGI-s and the values were still above these levels 2 weeks later at an interim visit. CGI-s >=3 was another condition for relapse but since the CGI-s ranges between 1 and 7 this seems to be covered by the CGI-s 2 points over baseline condition. The sponsor reported that a blinded review of the data revealed that many interim visits took place less than the designated 2 weeks after the initial "abnormal" HAM-D_{1-17} and CGI-s scores so they decided that relapse could be confirmed as little as three days after "abnormal" HAM-D_{1-17} and CGI-s scores were observed. Using this approach the sponsor determined that 50 (30.7%) Placebo patients and 25 (16.8%) STS patients relapsed at or before 1 year. This difference is significant (p=0.0025) based on the primary analysis method, center stratified Cochran-Mantel-Haenszel test. The Cox Proportional Hazards Model of the time until relapse, adjusting only for treatment, was specified as a secondary analysis. The test for a treatment effect based on the Cox model was also significant (p=0.0074). There were 87 placebo and 64 STS patients who simultaneously had a HAM-D_{1-17} >=14 and CGI-s >= 2 +baseline CGI-s during their time in the study, but confirmation 2 weeks later was needed for them to be classified as relapsed. It is important to note that many of these patients were censored (lost or ceased participation) without first having an interim visit. Since these patients did not have a confirmatory visit they were classified as not relapsed although their last measurements were in the "abnormal" range. Furthermore, there was a higher proportion of STS patients, 33 of 64 (52%), than placebo patients, 25 of 87 (29%), that did not have confirmation visits. Thus, the results may be biased in favor of the STS group.

Confirmation from an interim visit only 3 days after "abnormal" values are detected may not be reliable and the protocol required the interim visit to be 2 weeks later. Since visit windows were specified to +/3 days in the protocol we might require confirmatory visits to be at least 11 days after the initial abnormal values. Using this criteria for the timing of interim visits we determine

that 41 (25.2%) placebo and 23 (15.4%) STS patients relapsed. The p-value is 0.0223 for the Cochran-Mantel-Haenszel test and 0.0417 for the test based on the Cox proportional hazards model. In this case we still have a higher proportion of STS patients, 36 of 64 (56%), than Placebo patients, 35 of 87 (40%), that had "abnormal" values and were then censored without first having an interim visit. So again these results may be biased in favor of the STS group.

In summary, although the results are significant at face value they hide the fact that a higher proportion of STS patients had "abnormal" values and then ceased participation in the trial without first having an interim visit. If the criteria for relapse were less stringent and did not require confirmation of the "abnormal" $HAM-D_{1-17}$ and CGI-s values then these patients would be considered relapsed. In fact, the proportions of patients in each group who ever simultaneously had a $HAM-D_{1-17} >=14$ and a CGI-s at least 2 points over baseline are not different based on the center-stratified Cochran-Mantel-Haenszel test (53% vs. 43% p=0.057). On the other hand, the Cox proportional hazards model based test for a treatment group difference in the time until first simultaneous occurrence of $HAM-D_{1-17} >=14$ and CGI-s 2 over baseline, is just significant (p=0.048). Note that this approach counts 6 STS and 12 Placebo patients who had "abnormal" values and then had "normal" values at the interim visit as relapsed.

It is also a little strange that 10 of 159 STS patients did not have any post-baseline efficacy measures, while all 163 Placebo patients did. Seven of these patients were lost to follow up, 2 withdrew consent, and 1 was non-compliant. Although it is very unlikely that all 10 of these STS patients would have relapsed, if they had it would have been enough to alter the outcome of the trial. On the other hand, if none of them relapsed the results would be slightly stronger.

In conclusion, although the results are positive there is extra uncertainty in the estimated proportions relapsed and the group differences in the proportions relapsed because many patients (more in the STS group) were not adequately followed up on after having "abnormal" $HAM-D_{1-17}$ and CGI-s scores.

2. Introduction

2.1. Overview

NDA 21708 is a new application for a long-term maintenance of depression claim. NDA 21336 is a resubmission for acute treatment of major depression. It was resubmitted because the original NDA had several deficiencies. One of these deficiencies was that of the four acute

treatment studies submitted in the NDA only one was positive, S9303-E106-96B. Study S9303-E106-96B was a 6-week, double-blind, randomized, placebo-controlled, parallel assessment of the safety and efficacy of the selegiline transdermal system (STS) (20 mg /20cm^2) for major depression. One hundred and seventy seven patients were randomized among six centers. The majority of patients completed the study (88.8% for STS and 83% for placebo group). The two groups were compared with respect to the change from baseline in the HAM-D$_{1-17}$ at the end of week 6 (or last available observation) using an analysis of covariance. The statistical review of this study and the others in the first submission was done by Dr. Yuan-li Shen.

The key studies in this new submission are Study 0052 an acute depression study and Study 9806 a long term maintenance study. Study 0052 was a flexible dose, double-blind, placebo-controlled, multicenter, parallel-group randomized design. Two hundred and sixty five patients were randomized among 3 centers. Patients initiated treatment with a 20 mg/ 20cm^2 STS (Selegiline patch) or matching placebo. After two weeks the dose could be increased to 30 mg/ 30cm^2 if in the investigator's opinion the patient did not show improvement. Likewise, after 5 weeks the investigator could increase the dose to 30 mg/ 30cm^2 or matching placebo. The investigators performed the HAM-D$_{1-28}$ and other assessments at baseline, week 1, 2, 3, 5, and 8. The two groups were compared with respect to the change from baseline in HAM-D$_{1-28}$ after 8 weeks of double-blind flexible-dose treatment (or last observation) using an analysis of covariance model.

Study 9806 was a randomized, double-blind, placebo-controlled, parallel group study designed to assess the efficacy of selegiline 20 mg administered once daily via the STS for 52 weeks following a 10-week, open-label, remission, lead-in phase in patients with major depression. Patients who had a HAM-D$_{1-17}$ <= 10 at the end of week 6 or 7 and at the end of week 8 in the 10 week open-label phase were considered to have responded and were randomized to 52 weeks of treatment with 20 mg/ 20cm^2 STS or matching placebo. In the double-blind phase, patients were to have HAM-D$_{1-17}$ and CGI-s assessments taken at baseline and the end of Weeks 2, 4, 6, 8, 10, 12, 14, 18, 22, 26, 34, 42, and 52. Patients who had a HAM-D$_{1-17}$ >= 14 and a CGI-s 2 points above their baseline CGI-s (double-blind phase) at a scheduled visit were instructed to return 2 weeks later to confirm the "abnormal" HAM-D$_{1-17}$ and CGI-s. If the measurements taken two weeks later still exceeded the critical levels just mentioned and the DSM-IV criteria for depression were also confirmed then the patients were considered relapsed and their double-blind study treatment was ceased. Otherwise, the patient continued study treatment and resumed the regular study visit schedule. The two groups were compared with respect to the proportion relapsed at or before week 52, using a center-stratified Cochran-Mantel-Haenszel test. A Cox Proportional Hazards model for the time until relapse, adjusting for treatment, was specified as a secondary analysis.

2.2. Data Sources

The data for the new studies, 9806 and 0052, is located at the following address:

\\CDSESUB1\N21336\N_000\2003-08-07

The data for the earlier submission, including the other positive acute treatment study, S9303-E106-96B, is located at the following address:

\\Cdsesub1\n21336\N_000\2001-05-24\CRT\DATASETS\

3. Statistical Evaluation

3.1. Evaluation of Efficacy

3.1.1. Study S9303-P0052
This study started in September 2001 and ended in August 2002. The original protocol was dated July 9, 2001. There were two protocol amendments: one on February 8, 2002 and the other on June 3, 2002.

3.1.1.1. Objective
The objectives of this study are to assess the safety and efficacy of the selegiline transdermal system [STS (20 mg/ 20cm^2), STS (30 mg/ 30cm^2), STS (40 mg/ 40cm^2)] in the treatment of patients with moderate to severe major depressive disorder.

3.1.1.2. Study Design
This is a flexible dose, double-blind, placebo-controlled, multicenter, parallel-group randomized design. After completing screening procedures during a variable screening period of up to 28 days, patients will be randomized to treatment with STS (20 mg/ 20cm^2) or placebo patch. The STS or matching placebo is to be applied to a clean, hairless area on the patient's torso or upper arm. Several sites should be chosen to allow for STS daily rotation. Eligible Patients will be randomized to treatment groups in blocks of four. Randomization schedules will be generated for each site.

Approximately 400 adult patients with major depression will be screened at 3 investigative sites in the US; approximately 250 patients will be randomly assigned to treatment with active drug or placebo (125 patients/treatment group).

Clinical evaluations will be performed during screening, at baseline and at Weeks 1, 2, 3, 5, and 8. Patients who require a dose increase at the Week 5 visit will be evaluated at Week 6. The primary efficacy measure will be the total score of the Hamilton Depression Rating Scale

(HAM-D), items 1 through 28. Secondary efficacy measures include the Montgomery-Asberg Depression Rating Scale (MADRS), the total score of the HAM-D items 1-17, the HAM-D depressed mood item, and the investigator's Clinical Global Impressions of severity and change (CGI-s and CGI-c). Patients will complete the Inventory for Depressive Symptomatology, Self-Report (IDS-SR) at each study visit, including the final study visit.

After 2 weeks of treatment if, in the opinion of the investigator, the patient shows improvement (defined as a CGI-c score of 1 or 2 compared to baseline), the patient will continue treatment at this dose. If in the investigator's opinion the patient has not shown improvement, the dose will be increased to STS (30 mg/ 30cm^2) or to matching placebo patch. After 3 weeks of treatment at this dose (five weeks total treatment), if the patient has not shown improvement according to the investigator's clinical judgment, the dose will be increased to STS (40 mg/ 40cm^2) or the matching placebo patch. Patient's who show satisfactory improvement at a dose level will remain on that dose. Patients who experience adverse experiences due to a dose increase may have their dose decreased by one level. Patients who were not increased to either STS (30 mg/ 30 cm^2) or matching placebo at Week 2 may be increased at Week 3 or 5 if the patient no longer has a CGI-c score of 1 or 2 compared to baseline. Dose increases from STS (20 mg/ 20cm^2) or matching placebo to STS (30 mg/ 30 cm^2) or matching placebo may only occur at scheduled visits (i.e., Weeks 2, 3 and 5). Dose increases from STS (30 mg/ 30 cm^2) or matching placebo to STS (40 mg/ 40cm^2) or matching placebo may only occur at Week 5. Dose increases may not occur after Week 5 regardless of CGI-c score. Patients who continue to show satisfactory improvement (defined as a CGI-c score of 1 or 2 compared to baseline) at a dose will remain on that dose for the remainder of the study.

Patients who have adverse experiences considered by the investigator to be due to a dose increase may have their dose decreased by one dose level any time after a dose increase; an interim visit will be scheduled if the dose decrease does not occur within the next scheduled visit window. All safety and efficacy evaluations will be repeated at the time of the interim visit.

3.1.1.3. Efficacy Assessments

Primary
The primary efficacy will be the total score for items 1 through 28 of the Hamilton Depression Rating Scale (HAM D$_{1-28}$). This includes the 17-item total and item 1 ("depressed mood"), which will be analyzed separately as secondary efficacy parameters. The HAM D$_{1-28}$ will be obtained at Baseline and at Weeks 1, 2, 3, 5 (also at Week 6 for patients who have a dose increase at Week 5) and 8.

Secondary
The secondary efficacy assessments will be:

- Montgomery-Asberg Depression Rating Scale (MADRS) that assesses symptoms of depression such as mood, feelings of sadness, tension, sleep habits, and appetite. The 10-item questionnaire is administered by the investigator or qualified personnel at Baseline and at Weeks 1, 2, 3, 5 (also at Week 6 for patients who have a dose increase at Week 5) and 8; symptoms are rated using a 7-point scale (0 to 6) with 0 better and 6 worse.

- Clinical Global Impression of Severity is the clinician's assessment of severity of illness at Screening, Baseline and Weeks 1, 2, 3, 5 (also at Week 6 for patients who have a dose increase at Week 5) and 8. The·clinician rates the patient's change based on a scale ranging from "very much improved" to "very much worse" compared with baseline.

- Inventory of Depressive Symptoms—Self-Rated (IDS-SR): assesses the patient's perception of disease symptoms as rated by the patient. This questionnaire is completed by the patient at Baseline and subsequent clinic visits.

3.1.1.4.Analysis Methods
Primary Population
The primary population for assessment of efficacy will be the modified intent-to-treat population. The modified intent-to-treat population will consist of all randomized patients who received at least one protocol designated dose of study treatment and who had at least one post-treatment efficacy assessment with the primary outcome variable (HAM-D$_{1-28}$). Datasets will be defined for both 'last observation carried forward' (LOCF) and 'observed cases' (OC).

Other patient populations will be analyzed for efficacy but considered secondary patient populations. These are as follows:
   a) Intent-to-treat population, defined as all randomized patients receiving at least one protocol designated dose of study medication.

   b) Evaluable population (traditional observed cases population), defined as all randomized patients who completed the study without significant protocol violations.

Primary Analysis
For the primary efficacy analysis, a two-way ANOVA model will be fitted using the LOCF HAM-D$_{1-28}$ change from baseline as the response, treatment group and center as main effects, and baseline score as a covariate. Treatment-by-center interaction will be tested using a three way ANOVA model including treatment, center, and treatment-by-center main effects. If the interaction is not statistically significant at the 0.05 level, the above analysis model will be applied. If the treatment-by-center interaction differs significantly from zero at a significance level of 0.05, the interaction term will be included in the above analysis model and treatment-by-center plots will be examined to characterize the form of the interaction. If the

interaction effect is quantitative (that is the magnitude of the effect differs among centers but not the direction) then nothing further will be done. If the interaction is qualitative (that is, the direction of the treatment effect differs among centers), further exploratory analyses will be undertaken to understand the interaction.

These ANOVAs will be conducted for the change from baseline in the HAM-D$_{1-28}$ at weeks 1, 2, 3, 5, 8, and LOCF endpoint analyses.

Secondary analyses will be conducted on the primary endpoint using the OC data set, as well as LOCF analyses at visits prior to the last visit, to understand the behavior of the medication over the course of the study. For continuous secondary efficacy variables, the same sequence of statistical analyses described above will be conducted. For categorical secondary efficacy variables a Cochran-Mantel-Haenszel (CMH) Type 2 (ANOVA mean score) statistic using Center as stratum will be used for ordinal variables (e.g., CGI-s, HAM-D item 1) while a CMH Type 1 will be used for nominal variables (e.g., response rates).

This clinical trial is to be done at exactly three centers so there will be no need for center pooling.

Sample Size
Based on previous data the standard deviation is expected to be approximately 9.5. The study is powered to detect a difference between groups of approximately 3.5 based on the HAM-D$_{1-28}$. The hypothesis test will be two-tailed. To achieve 80% power, approximately 110 subjects per group will be required. Approximately 125 patients per group will be used to allow a small margin for unevaluable cases.

3.1.1.5.Sponsor's Results

A total of 265 patients were randomized into this study (n=132 for STS and n=133 for placebo). The study was completed by 100 (75.8%) STS patients and 106 (79.7%) placebo patients. The most frequent reason for withdrawal was lost to follow up for the placebo group, while adverse event and withdrew consent were the two most frequent reasons for STS patients.

Table 3.1 Completion Status

COMPLETION STATUS	TREAT		Total
	Placebo	STS	
Complete	106 79.70	100 75.76	206
Withdraw	27 20.30	32 24.24	59
Total	133	132	265
Reason for Withdrawal			
ADVERSE EVENT	3 11.11	9 28.13	12
LACK OF EFFICACY	3 11.11	5 15.63	8
NONCOMPLIANCE	3 11.11	6 18.75	9
LOST TO FOLLOW-UP	14 51.85	2 6.25	16
WITHDREW CONSENT	2 7.41	7 21.88	9
PROTOCOL VIOLATION	1 3.70	1 3.13	2
OTHER	1 3.70	2 6.25	3

Patient demographics and baseline characteristics appeared to be comparable between the treatment groups. Age, Race, Gender and baseline HAM-D_{1-28} are presented in the following table.

Table 3.2 Patient Demography and Baseline by Treatment Group (All Randomized Patients)

		STS (N=132)	Placebo (N=133)
AGE, YRS	N	132	133
	MEAN (SD)	41.83 (12.36)	41.56 (11.61)
	MEDIAN	42	43
	MIN, MAX	19,70	18,68
Race, N (%)	Caucasian	106 (80.3)	108 (81.2)
	Black	6 (4.55)	6 (4.51)
	Asian	1 (0.76)	2 (1.50)
	Hispanic	12 (9.09)	13 (9.77)
	Other	7 (5.30)	4 (3.01)
Gender, N (%)	Male	52 (39.39)	63 (47.37)
	Female	80 (60.61)	70 (52.63)
HAM-D$_{1-28}$ -Baseline	N	132	133
	MEAN (SD)	28.34 (3.73)	28.50 (3.94)
	MEDIAN	28	28
	MIN, MAX	21,39	21,40

There were no apparent differences between the groups in terms of Age or Race. Although there were 8% more males in the placebo group than the STS group this difference is not significant. The mean baseline HAM-D$_{1-28}$ Total scores were also comparable.

Primary Analysis

Table 3.3 Change from Baseline in HAM-D$_{1-28}$

	Treat		Baseline	Week 8
Modified ITT *	Placebo	N	128	128
		Mean (SD)	28.6 (4.0)	-8.9 (9.1)
	STS	N	129	129
		Mean (SD)	28.3 (3.7)	-11.1 (8.6)
	Between	P-VALUE	0.6150	0.0327
Observed Cases	Placebo	N	128	110
		Mean (SD)	28.6 (4.0)	-9.5 (9.3)
	STS	N	129	100
		Mean (SD)	28.3 (3.7)	-12.8 (8.0)
	Between	P-VALUE	0.6150	0.0053

* 8 patients were excluded from the modified ITT population because they lacked a post-baseline HAM D

For the modified ITT population, the HAM-D$_{1-28}$ mean (+/- SD) scores at baseline were 28.3 (±3.7) and 28.6(±4.0) for the STS and placebo groups, respectively. At end of treatment, the

LOCF HAM-D$_{1-28}$ scores decreased to 17.2 ($\pm$8.6) and 19.8($\pm$9.2) for the STS and placebo groups, respectively. Significant differences for mean change from baseline in HAM-D$_{1-28}$ for LOCF data, were observed between STS and placebo treatment by Week 5 (p=0.0255) continuing until end of treatment (p=0.0327). By observed cases (OC) analysis, significant differences for mean change from baseline in HAM-D$_{1-28}$ were observed between STS and placebo treatment at Week 8 (p=0.0053).

Secondary Analyses

For the modified ITT population, the HAM-D$_{1-17}$ mean ($\pm$SD) scores at baseline were 23.4 ($\pm$2.5) and 23.7 ($\pm$2.7) for the STS and placebo groups, respectively. At end of treatment, LOCF HAM-D$_{1-17}$ mean scores decreased to 14.7 ($\pm$7.2) and 16.2 ($\pm$7.5) and the OC HAM-D$_{1-17}$ mean scores decreased to 13.4 ($\pm$6.9) and 15.6 ($\pm$7.6) for the STS and placebo groups, respectively. A significant difference was observed between treatment groups for the mean change from baseline in HAM-D$_{1-17}$ scores at Week 8 (p=0.0310 for OC data), but not for LOCF data at any timepoint.

Table 3.4 Secondary Endpoints Modified ITT-LOCF

		Placebo	STS	Between p-value
HAM-D$_{1-17}$	N	128	129	
	Base Mean (SD)	23.7(2.7)	23.4(2.5)	
	Change Mean (SD)	-7.4(7.4)	-8.7(7.0)	0.13
HAM-D-Bech	N	128	129	
	Base Mean (SD)	12.6(1.3)	12.4(1.3)	
	Change Mean (SD)	-4.1(4.2)	-5.5(4.3)	0.004
HAM-D Item 1 Depressed Mood	0=Absent	19.0(14.8)	28.0(21.7)	0.004
	1	29.0(22.7)	42.0(32.6)	
	2	29.0(22.7)	28.0(21.7)	
	3	47.0(36.7)	30.0(23.3)	
	4	4.0(3.1)	1.0(0.8)	
CGI-c N(%)	Very Much Improved	23.0(18.0)	29.0(22.5)	0.04
	Much Improved	22.0(17.2)	31.0(24.0)	
	Minimally Improved	31.0(24.2)	35.0(27.1)	
	Unchanged	45.0(35.2)	27.0(20.9)	
CGI-c N(%) (continued)	Minimally Worse	7.0(5.5)	7.0(5.4)	
	Much Worse	0.0(0.0)	0.0(0.0)	
	Very Much Worse	0.0(0.0)	0.0(0.0)	
CGI-s N(%)	Normal	14.0(10.9)	21.0(16.3)	0.15
	Borderline Ill	16.0(12.5)	16.0(12.4)	
	Mildly Ill	19.0(14.8)	26.0(20.2)	
	Moderately Ill	59.0(46.1)	47.0(36.4)	
	Markedly Ill	19.0(14.8)	19.0(14.7)	
	Severely Ill	1.0(0.8)	0.0(0.0)	
	Among Most Ill	0.0(0.0)	0.0(0.0)	
MADRS	N	128	129	
	Base Mean (SD)	29.3(4.2)	29.3(4.2)	

		Placebo	STS	Between p-value
	Change Mean (SD)	-8.6(10.3)	-11.6(9.8)	0.02

3.1.1.6.Reviewer's Comments

This reviewer verified the sponsor's primary analysis which yielded a significant difference between the treatment groups in the change from baseline in the HAM-D_{1-28} at Week 8 (or LOCF). The changes from baseline in the HAM-D_{1-28} were -11.1 for STS and -8.9 for placebo (p=0.0327). The analysis was based on an analysis of covariance model with center and treatment effects and the baseline HAM-D_{1-28} as a covariate. There were a total of 8 patients (5 placebo ; 3 STS) with no post-baseline HAM-D_{1-28} measures. Baseline measures were not carried forward for these patients so the primary analysis was based on 128 placebo patients and 129 STS patients.

The sponsor noted that although the change from baseline in the HAM-D_{1-17} was significant at Week 8 for the Observed Cases population it was not significant at Week 8 (LOCF) for the Modified Intent-to-Treat population (p=0.13). Insomnia, agitation, and weight loss were reported as adverse events by STS patients more than by placebo patients (p=0.0044, 0.0662, 0.0295 respectively). The treatment effect on the LOCF HAM-D_{1-17} may have failed to achieve significance because HAM-D_{1-17} items 4, 5, and 6 refer to Insomnia, item 9 concerns agitation, and item 16 concerns weight loss. In fact, if we analyze these items individually using proportional odds models we find that the treatment effects favor placebo but are not significant. These items would have less influence on the larger HAM-D_{1-28} scale and this might explain why the treatment effect on the change in HAM-D_{1-28} was significant while for the change in HAM-D_{1-17} it was not.

Center Differences

The difference in treatment group means for the change from baseline in the HAM-D_{1-28} was larger in center 1 than in center 2 or 3. The differences were 3.6, 1.3, and 1.7 respectively. The mean change for the STS group in center 1 was more than a point lower than the mean change for the STS group in center 2 or center 3. The largest placebo group mean was also observed in center 1. Center 1 also had the smallest baseline HAM-D_{1-28} scores. Still since there was not a significant center by treatment interaction and the overall result was positive we do not worry too much about these center differences.

Table 3.5 Change in HAM-D_{1-28} by Center

Center	Treat	N	Baseline HAM-D_{1-28} Mean (SD)	Change HAM-D_{1-28} Mean (SD)
1	Placebo	39	27.1 (4.1)	-8.3 (9.2)
1	STS	42	27.3 (3.7)	-11.9 (9.9)

Center	Treat	N	Baseline HAM-D$_{1-28}$ Mean (SD)	Change HAM-D$_{1-28}$ Mean (SD)
2	Placebo	40	28.5 (3.1)	-9.5 (7.8)
2	STS	40	27.9 (3.0)	-10.8 (7.8)
3	Placebo	49	30.0 (4.0)	-8.9 (10.0)
3	STS	47	29.7 (4.0)	-10.6 (8.1)

Effect of Missing Data and Observed Cases Analysis

In Table 3.1 reports that 106 (80%) placebo and 100 (76%) STS patients completed the study. It is important to consider how the dropouts may have impacted the results. One way to do this is to check for agreement between the LOCF analysis, which includes all randomized patients with at least one post-baseline primary efficacy assessment, and the Observed Cases analysis, which only includes patients with primary efficacy assessments in week 8. In Table 3.3 we saw that the sponsor found a significant treatment effect on the change from baseline in the HAM-D$_{1-28}$ for both the LOCF and the Observed Cases analyses. This reviewer found fewer Observed Cases patients than the sponsor. This may be because this reviewer only included patients that had a visit falling in the protocol specified window for week 8, i.e., between days 53 and 59. At any rate, this reviewer's results based on the smaller group are consistent with the sponsor's results.

Table 3.6 Reviewer's Observed Cases Results for Change in HAM-D$_{1-28}$

Treat	N	Baseline	Change	p-value
Placebo	98	28.3 (3.9)	-9.3 (9.3)	0.006
STS	94	28.4 (3.8)	-12.7 (7.9)	

The following table shows that patients who withdrew had less improvement than those who completed. This is not surprising since on average patients in both groups improved over time. This also helps to explain why the Observed Cases analysis is more significant than the Last Observation Carried Forward analysis. In particular, STS patients who withdrew were not more improved than placebo patients. Therefore, the treatment effect is diminished by including these patients in the analysis on the basis of their last post-baseline observations. One might expect that if there were no dropouts the results would be somewhere between the LOCF results and the Observed Cases results. Since both the LOCF and the Observed Cases analyses indicate a significant treatment effect we conclude that there is no evidence that the results are biased because of dropouts and that the study is positive.

Table 3.7 Change in HAM-D$_{1-28}$ by Completion Status

Status	Treat	N	Baseline HAM-D$_{1-28}$	Change in HAM-D$_{1-28}$
Completed	Placebo	106	28.3(3.9)	-9.5(9.3)
Completed	STS	100	28.4(3.8)	-12.8(8.0)
Withdrew	Placebo	22	30.0(4.0)	-5.8(7.5)
Withdrew	STS	29	28.1(3.6)	-5.1(8.0)

3.1.2. Study 9303-P9806

The original protocol was dated June 21, 1999. There were two amendments the first on August 17, 1999 and the second on May 15, 2000.

3.1.2.1.Objective

The objective of this study is to assess the safety and efficacy of selegiline versus placebo with regard to the reappearance of symptoms associated with major depression in patients with a stable HAM D_{1-17} score of <= 10.

3.1.2.2.Study Design

This randomized, double-blind, placebo-controlled, parallel group study is designed to assess the safety and efficacy of selegiline 20 mg, administered once daily (i.e., every 24 hours) via the STS for 52 weeks following a 10-week, open-label, remission, lead-in phase in patients with major depression.

Approximately, 700 patients will be enrolled into the 10-week, open-label, remission, lead-in phase with an estimated 300 patients potentially eligible for randomization in the double-blind phase of the study. All patients will receive 20 mg of selegiline administered via a STS once daily for 10 weeks. The STS is to be applied to a clean, hairless area on the patient's torso or upper arm. STS application should be alternated among several chosen sites. Patients whose symptoms associated with depression remit as defined by two HAM-D_{1-17} scores of <= 10 within the last 2 weeks of open-label treatment (i.e., either visits 6 or 7 and Visit 8) will then be randomly allocated into the double-blind phase of the study. The DSM-IV assessment for a major depressive episode will be performed prior to randomization to confirm remission of the current major depressive episode. Patients who meet the DSM-IV criteria for a major depressive episode will not be randomly allocated into the double-blind phase.

This study will consist of a screening and a baseline visit. Patients who meet entrance criteria at the screening visit will discontinue all prohibited medications and will discontinue any psychotherapy sessions. Patients will return to the clinic within 2 weeks of screening for a baseline visit.

Patients enrolled into the open-label phase of the study at baseline will return for bi-weekly visits (i.e., every 14 days ± 3 days) for the first 8 weeks of STS treatment and weekly visits for the last 2 weeks (i.e., Visits 7 and 8). Patients who are responders to open-label treatment will be randomly allocated into the double-blind phase and will receive STS or matching placebo for 52 weeks. Patients will be assessed at weeks 2, 4, 6, 8, 10, 12, 14, 18, 22, 26, 34, 42, and 52 of the double-blind phase. So the length between visits gets longer at weeks 14, 26, and 42. If at a scheduled visit during double-blind treatment, a patient has a HAM D_{1-17} score >= 14 and a CGI-s score of >= 3 with at least a 2-point increase in CGI-s from visit 8, he/she must return to the clinic 2 weeks later. At the 2-week visit HAM D_{1-17} and CGI-s must be administered to confirm

reappearance of depression. In addition, reappearance must be confirmed by the presence of
DSM-IV criteria for a major depressive episode at the interim 2-week visit. Patients who are not
considered to have a reappearance of depression at the interim visit will remain in the study and
will return for the next scheduled visit.

Patients who meet the criteria for reappearance during the first 6 months of double-blind
treatment will be given the opportunity to enter an open-label rescue phase with STS 20 mg/ 20
cm^2 for a maximum of 6 weeks.

3.1.2.3. Efficacy Assessments
Primary
The primary efficacy comparison will be a between-group comparison of the cumulative
proportion of patients in each group experiencing reappearance of depression over 12 months.
Reappearance of depression is defined in this study as a HAM-D_{1-17} score $\geq$ 14 and a CGI-s
score of $\geq$ 3 with an increase in CGI-s of at least 2 points from Visit 8 during the double-blind
phase of the study and measured at two visits performed over a 2-week time period.

Secondary
Secondary efficacy assessments include a between-group comparison of time-to-reappearance of
depression and the mean time without reappearance during the trial. In addition, the within-
group and between-group mean changes in CGI-s, Clinical Global Improvement of Change
(CGI-C), Montgomery-Asberg Depression Rating Scale (MADRS) scores, and HAM-D_{1-28}
scores from the time of randomization into the double-blind phase of the study to study
discontinuation will be assessed.

3.1.2.4.Protocol Specified Statistical Analysis Plan
Sample Size
It is estimated that of the 700 patients initially enrolled in the open-label phase of the study at
least 60% will be responders to treatment, providing 300 potential patients for enrollment in the
double-blind phase of the study. It is estimated from placebo-controlled studies with other
antidepressants that over 6 months of double-blind treatment approximately 35% of placebo
patients and 15% of selegiline patients will experience a reappearance of disease. Due to the
length of the study, it is difficult to estimate how many patients who are randomly allocated to
double-blind treatment will complete therapy. The target patient population of 150 patients per
treatment group will provide more than adequate power (97%) to detect a difference between the
groups. However, if drop-out rates are high, samples of 107 or 83 patients in each group during
the double-blind phase of the study will have 90% or 80% power, respectively to detect a
difference between active drug group and placebo group, using a Mantel-Haenszel test with a
0.05 two-sided significance level.

While the primary analysis will be based on overall treatment group totals stratified by center,

additional analyses will be performed by levels of various potential confounding variables to determine the consistency of the response across subgroups of patients. The following variables will be examined in this manner: treatment, investigator, age, gender, race, underlying condition, duration of current depressive episode prior to enrollment, and medication compliance. A regression model will be used to test the effect of each of these variables and to test the treatment effect for each of these variables. Additional regression models will be run on an exploratory basis for other independent variables to rule out a possible relationship to response. Sites with a small number of patients will be pooled for the regression analyses.

Patient Population
Analysis of the primary efficacy measures will be carried out on the ITT and modified ITT populations. Additional analyses will be performed using both ITT and modified ITT patients.

In addition, two analyses will be performed on the ITT population for the proportion of patients with reappearance of depression and the time to reappearance analyses. In the first analysis, patients who discontinue due to lack of efficacy will be designated as marked on the CRF, and in the second analysis they will be designated (i.e., re-categorized by the sponsor) as patients with a reappearance of depression.

Analysis Methods
- Center stratified Mantel-Haenszel test for ordinal variables (SAS Proc FREQ). If covariates must be in the test, a linear modeling test for ordinal variables will be used (SAS Proc CATMOD).
- Covariance analysis for continuous variable using treatment group and center as main effects and baseline measurements as covariates (SAS Proc GLM).

- Time to reappearance of depression based on HAM-D$_{1-17}$ will be summarized by Kaplan-Meier method (SAS Proc LIFETABLE) and Cox proportional hazards methods will be applied to test statistical difference between the two treatment groups (SAS Proc PHREG).

  The analyses will be performed:
  a) for the 1-year double-blind treatment, and
  b) separately for the first 6 months of double-blind treatment

3.1.2.5.Patient Completion Status and Demography

In this study 675 patients were enrolled into the open label phase. One patient was not dispensed study drug. The mean age of the 675 patients enrolled in the open label phase of the study was

approximately 43 years. Sixty eight percent of the patients were female and about 83% were Caucasian. 366 (54.2%) patients completed the open label portion of the study and 309 (45.8%) discontinued. Remission of depression (defined by HAM-D$_{1-17}$ scores $\leq$ 10 at week 6 or 7 and week 8 of open label treatment) occurred in 342 (53.0%) of the 645 patients enrolled in the initial phase (open label) of this study. Three Hundred and Twenty Two patients were randomized in the double-blind phase. The following table gives the various reasons that the other 353 (52.4%) patients were not randomized into the double blind portion of the study.

Table 3.8 Reasons not Randomized into Double Blind Phase

Reason Not Randomly Allocated	Frequency	Percent
Patient was not a Responder	90	25.50
Adverse Event	63	17.85
Lost to Follow-Up	60	17.00
Withdrew Consent	59	16.71
Lack of Efficacy	41	11.61
Noncompliance	22	6.23
Other	8	2.27
Protocol Violation	5	1.42
Pregnancy	5	1.42

Of the 322 patients randomized into the double-blind phase 159 were randomized to STS and 163 were randomized to placebo. 10 STS patients were excluded from the Modified Intent to Treat population because they did not have at least one post-baseline HAM-D$_{1-17}$ assessment. Seven of these 10 patients were lost to follow up, 1 was non-compliant, 1 withdrew consent, and 1 was not dispensed study medication.

The two treatment groups were well matched with respect to age, gender, race, CGI-s and HAM-D$_{1-17}$ score at the baseline for the double-blind phase.

Table 3.9 Patient Demography and Baseline Characteristics- Double Blind Phase*			
		Placebo	STS
Age, yrs	N	149	163
	MEAN (SD)	42.4(12.5)	43.9(11.3)
	MEDIAN	44.0	45.0
	MIN, MAX	19.0, 73.0	18.0, 81.0
Gender N(%)	Female	106.0(71.1)	106.0(65.0)
	Male	43.0(28.9)	57.0(35.0)
Race N(%)	Asian	1.0(0.7)	2.0(1.2)
	Black	8.0(5.4)	9.0(5.5)

Table 3.9 Patient Demography and Baseline Characteristics- Double Blind Phase*

		Placebo	STS
	Caucasian	124.0(83.2)	134.0(82.2)
	Hispanic	11.0(7.4)	14.0(8.6)
	Other	5.0(3.4)	4.0(2.5)
Baseline CGI-s N(%)	1	66.0(44.3)	81.0(49.7)
	2	60.0(40.3)	59.0(36.2)
	3	21.0(14.1)	22.0(13.5)
	4	2.0(1.3)	1.0(0.6)
	5	0.0(0.0)	0.0(0.0)
	6	0.0(0.0)	0.0(0.0)
HAM-D$_{1-17}$-Baseline	MEAN (SD)	5.3(2.7)	5.3(2.7)
	MEDIAN	5.0	5.0
	MIN, MAX	0.0, 10.0	0.0, 10.0

*10 STS patients had no post-baseline efficacy measures.

The following table gives the completion status for the double blind phase.

Table 3.10 Double Blind Period Completion Status

Double-blind phase completion status	Treatment group		Total
	Placebo	STS	
Completed Study	39	41	80
	23.93	25.79	
Relapsed*	51	30	81
	31.29	18.87	
Withdrew Prematurely	73	88	161
	44.79	55.35	
Total	163	159	322

* Based on the double-blind termination CRF page (not the primary efficacy endpoint)

Forty of 163 (24.5%) Placebo patients and 24 of 149 (16.1%) STS patients entered the rescue phase. Although this was not specified as an efficacy assessment and rescue was only an option if a patient had been in the double-blind phase less than six months the difference is not significant (p=0.07).

APPEARS THIS WAY
ON ORIGINAL

3.1.2.6.Sponsor's Results

The primary efficacy endpoint, the proportion of patients experiencing relapse/recurrence by
Type I Criteria at Week 52, was defined utilizing all of the following criteria:
- HAM-D$_{1-17}$ score >=14
- CGI-s score >=3 (this seems to be covered by the next condition since CGI-s >=1)
- At least a 2-point increase in CGI-s from Visit 8 (double-blind baseline visit)
- All of the above three conditions met again (i.e., relapse/recurrence conditions met at
  both the last two consecutive visits) to confirm relapse/recurrence
- A visit window interval of >= 3 days apart for the last two consecutive visits (initial and
  confirmatory visits)

Note that the protocol required a visit window interval of 2 weeks for the last two
consecutive visits. The sponsor stated that this was changed to 3 days after a blinded review
of the data.

Patients were defined as having Type II relapse/recurrence if they met all of the Type I criteria or
had one of the following investigator recorded determinations of reappearance:
- Investigator indicated reappearance of depression on the relapse or reappearance CRF
  page
- Investigator indicated reappearance of depression on the "termination" CRF page
- Investigator indicated withdrawn due to lack of efficacy on the "termination" CRF page
  and 80%-120% compliant with study drug

The primary efficacy endpoint was the relapse/recurrence event characterized by Type I Criteria.
In the 52-week double-blind phase, statistically significant differences in the percentage of
patients with relapse/recurrence, by Type I Criteria, were observed between the treatment groups
at Week 52, the primary endpoint, (p=0.0025)

Table 3.11 Patients with Type I Criteria for Relapse/Recurrence

		STS	Placebo	P-value *
Reappearance in 1 yr (Primary)	Yes	25.0(16.8)	50.0(30.7)	
	No	124.0(83.2)	113.0(69.3)	0.0025
Reappearance in ½ yr	Yes	25.0(16.8)	48.0(29.5)	
	No	124.0(83.2)	115.0(70.6)	0.0051

* based on a Cochran-Mantel-Haenszel test stratified by center w/ centers 9,19,20,24 pooled

Statistically significant differences between the treatment groups in the Kaplan-Meier cumulative rate of relapse/recurrence by Type I criteria were found using the Log-Rank test, a secondary analysis, at Week 26 (p=0.0115) and at Week 52 (p=0.0061) of the double blind phase.

The sponsor's definition of Type II relapse was not defined in the protocol. Type II relapse patients include patients who satisfied Type I relapse criteria or who the investigator indicated as relapsed on the CRF, or who withdrew for lack of efficacy were also considered relapsed. The sponsor's results for Type II relapse were also significant.

Table 3.12 Patients with Type II Criteria for Relapse/Recurrence

		STS	Placebo	P-value *
Reappearance in 1 yr	Yes	42.0(28.2)	69.0(42.3)	
	No	107.0(71.8)	94.0(57.7)	0.0057
Reappearance in ½ yr	Yes	38.0(25.5)	66.0(40.5)	
	No	111.0(74.5)	97.0(59.5)	0.0035

* based on a Cochran-Mantel-Haenszel test stratified by center w/ centers 9,19,20,24 pooled

3.1.2.7.Reviewer's Comments

Curiously, 10 of the 159 STS patients had no post-baseline HAM-D_{1-17} or CGI-s measurements and were therefore excluded from the analysis, but all 163 placebo patients did have post-baseline HAM-D_{1-17} and CGI-s measurements. Seven of these 10 STS patients were lost to follow up, 2 withdrew consent, and 1 was non-compliant. We note in passing that although we have no reason to believe that any of these patients relapsed the results are not robust to the very unlikely event that all 10 of them relapsed. On the other hand, if none of them relapsed the results would be slightly stronger.

The mean time that patients were in "remission" (HAMD-17 total score <= 10) in the open-label phase prior to randomization was defined in the following way. For each ultimately randomized patient determine the first time at which the patient had a HAMD-17 <=10 and no HAMD-17 scores > 10 after that time. Then take the difference between this time and the last open label visit time. The average value for this over the 322 randomized patients was 24.6 +/- 0.9 days (standard error of the mean) and the standard deviation was 16.0. It is important to note that some observations in the open label period went beyond the visit 8 (week 10) window (one as far out as 84 days). Also, if one strictly imposes the protocol specified visit windows (+/- 3 days) then 47 ultimately randomized patients did not have a HAM-D_{1-17} <=10 at visit 6 or 7 and 8. For example, one randomized patient only had a HAMD-17 <=10 at visit 8. Another randomized patient only had HAMD-17<=10 at visits 7 and 8, and these were only 2 days apart. Apparently, the protocol requirements for eligibility for randomization were not strictly observed.

This reviewer checked the sponsor's primary analysis of the proportions relapsed at 1 year using

Type I criteria. The analysis method was the center stratified Cochran-Mantel-Haenszel general association test. Four centers (09, 19, 20, 24) with a small number of patients [<2 patients in each treatment group] were pooled for the analysis. The number of patients relapsed at or before 1 year was significantly higher for placebo than for STS (50 (30.7%) vs. 25 (16.8%) p=0.0025).

If the HAM-D$_{1-17}$ and CGI-s criteria for reappearance of depression were met then the DSM-IV was supposed to be administered at the interim visit to get further confirmation. It appears that the DSM-IV confirmation was not recorded for 18 out of 50 (36%) placebo cases and 10 out of 25 (40%) STS cases. However, there were still significantly more placebo patients than STS patients that met all of the criteria (i.e., including DSM-IV criteria) for reappearance of depression.

The Type I criteria allowed interim visits as early as 3 days after the initial visit at which "abnormal" HAM-D$_{1-17}$ and CGI-s values were recorded to be used for relapse confirmation. The sponsor states that this was decided on after a blinded review of the data which revealed that a number of patients were not able to wait the protocol specified 2 weeks. However, it appears that this change was not submitted as a protocol amendment. Furthermore, Dr. Greg Dubitsky, the FDA clinical reviewer, believes that an 11+ day interval would be more meaningful and in keeping with the visit windows specified in the protocol. Therefore, we consider events confirmed after 11 or more days to be of primary importance rather than the sponsor's Type I or Type II events. The sponsor reported the following proportions of relapse using the 11+ day criteria for interim visits.

Table 3.13 Sponsor's Relapse/Reappearance Results (interim 11+ days later)

Reappearance by 1 year ?	STS	Placebo	Cochran-Mantel-Haenszel P-value	Cox PH Model P-value
Yes	21.0(14.1)	39.0(23.9)		
No	128.0(85.9)	124.0(76.1)	0.0183	0.0347

\# uses >= 11 day criteria for interim visit
* based on a Cochran-Mantel-Haenszel test stratified by center w/ centers 9,19,20,24 pooled

The sponsor's Cochran-Mantel-Haenszel analysis of the proportions relapsed was significant (p=0.0183). The sponsor also analyzed the time to relapse using a Cox Proportional Hazards Model with treatment as a factor and found that the STS : Placebo hazard ratio was 0.565 and the test for a difference in the survival curves was significant (p=0.0347). At 52 weeks the estimated survival was 80% for STS and 68% for Placebo.

It appears that STS patient 28035 and placebo patient 30013 who met the criteria for relapse using the >= 3 day criteria for the interim visit were mistakenly not classified as relapsed by the sponsor using the >= 11 day criteria. This might have happened because each patient had two visits within the 14 days after the initial relapse signal and each visit was less than 11 days after the preceding one. However, since the conditions were confirmed at both of these visits and the last one was exactly 14 days after the original signal it seems that these patients should be

classified as relapsed. Two other patients (04026 and 28037) satisfied the conditions which indicated the need for an interim visit and although they did not have a confirmatory visit, their next measurements, taken in the rescue phase, satisfied the relapse criteria. However, these two patients were not classified as relapsed by the sponsor. Thus, this reviewer believes that 4 more patients (2 placebo and 2 STS) should be classified as relapsed using the 11+ day interval criteria for the interim visit. When this is done the proportions become 15.4% and 25.2% and the Cochran-Mantel-Haenszel test gives a p-value of 0.0223. The test for a difference in time to relapse based on the Cox Proportional Hazards Model yields a p-value of 0.0417 and a STS : Placebo hazard ratio of 0.588.

Table 3.14 Reviewer's Relapse/Reappearance Results (interim 11+ days later)

Reappearance by 1 year ?	STS	Placebo	Cochran-Mantel-Haenszel P-value	Cox PH Model P-value
Yes	23.0(15.4)	41.0(25.2)		
No	126.0(84.6)	122.0(74.8)	0.0223	0.0417

The following table shows the number of patients at risk and the cumulative number relapsed at 30 day increments. It is apparent that almost all of the relapses happened in the first half year of patients' participation.

Table 3.15 Number at Risk and Number Relapsed Over Time (interim 11+ days later)

	STS(N=149)			Placebo(N=163)		
Day	# At Risk	# Relapsed	# Censored	# At Risk	# Relapsed	# Censored
30	110	13	26	117	20	26
60	87	21	41	88	30	45
90	74	23	52	68	37	58
120	63	23	63	63	38	62
150	59	23	67	53	40	70
180	55	23	71	49	40	74
210	47	23	79	34	41	88
240	43	23	83	33	41	89
270	38	23	88	29	41	93
300	35	23	91	29	41	93
330	32	23	94	28	41	94
360	25	23	101	22	41	100
365	23	23	103	16	41	106

It appears that the sponsor has defined the time to relapse as the time from double-blind randomization until relapse is confirmed. The time to event analysis was considered a secondary analysis in the protocol and the definition of the time to event was not specified. This reviewer believes it may be more reasonable to define the time to relapse as the time from randomization until the visit which triggered the interim visit, assuming that relapse was confirmed at the interim visit. However, this distinction seems to have little effect on the results. This reviewer also noted that the sponsor's censoring time was sometimes larger than the reviewer's censoring time and the reviewer could find no data to support the longer censoring time. For example, this reviewer determined a censoring time of 182 days for the patient with screen number='05040' and could find no data beyond this time, yet the sponsor's censoring time was 279 days for this

patient. However, these discrepancies seemed to have little impact on the results.

Potential Source of Bias: Patients that were inadequately followed up on

Although isolated occurrences of HAM-D_{1-17} >=14 and CGI-s 2 points above baseline were not enough for relapse it is of interest to check how many patients ever simultaneously had "abnormal" values, i.e., HAM-D_{1-17} >=14 and CGI-s 2 points above baseline.This reviewer found that 64 of 149 (43%) STS patients and 87 of 163 (53%) placebo patients simultaneously met the HAM-D_{1-17} $\geq$ 14 and 2 point increase from baseline in the CGI-s at least once.
The following table shows how patients may have met the criteria at least once but not ultimately gotten classified as relapsed. Note that, here, like the sponsor, we have counted as not relapsed the two patients mentioned earlier who could be considered relapsed based on "abnormal" values in the rescue phase.

Table 3.16 Relapse Status for patients who ever simultaneously had abnormal HAM-D_{1-17} and CGI-s

Relapse Status for patients who ever simultaneously had HAM-D_{1-17} >= 14 and CGI-s>= 2 +baseline			
Status	Reason	STS (n=64)	Placebo (n=87)
Relapsed	Had HAM-D_{1-17} >= 14 and CGI-s >= 2+baseline CGI-s then "abnormal" values confirmed >= 11 days later at interim visit	22.0(34.4)	40.0(46.0)
Not Relapsed	Had HAM-D_{1-17} >= 14 and CGI-s >= 2+baseline CGI-s then censored	32.0(50.0)	22.0(25.3)
	Had HAM-D_{1-17} >= 14 and CGI-s >= 2+baseline CGI-s then "abnormal" values confirmed < 11 days later at interim visit then censored	4.0(6.3)	13.0(14.9)
	Had HAM-D_{1-17} >= 14 and CGI-s >= 2+baseline CGI-s then "normal" score at interim visit	6.0(9.4)	12.0(13.8)

* using >= 11 day criteria for interim visits
In row 2, 5 placebo and 4 STS relapsed (investigator) and 8 placebo and 9 STS withdrew for LOE
In row 3, 5 placebo and 3 STS relapsed (investigator) and 7 placebo and 1 STS withdrew for LOE

Thirty two (32) STS and 22 placebo patients had a HAM-D_{1-17} >=14 and CGI-S >= 2 +baseline CGI-S and then were censored without having an interim visit. Another 4 STS and 13 placebo had an interim visit which confirmed the earlier "abnormal" values but the interim visit was less than 11 days later. So, adding these patients, it seems that there are 36 STS and 35 placebo patients that may have relapsed but weren't followed long enough for us to be sure. This may

have biased the results. The following table shows the completion status and reasons for withdrawal for these 71 patients.

APPEARS THIS WAY
ON ORIGINAL

Table 3.17 Completion Status for patients who withdrew before having appropriately timed interim visit

		STS	Placebo	Total
Completed Study		3.0(8.3)	1.0(2.9)	4
Relapsed (Investigator)		7.0(19.4)	10.0(28.6)	17
Withdrew:		26.0(72.2)	24.0(68.6)	50
	Reason Withdrew			
	Adverse Event	7.0(26.9)	3.0(12.5)	10
	Lack of Efficacy	10.0(38.5)	15.0(62.5)	25
	Noncompliance	2.0(7.7)	1.0(4.2)	3
	Lost to Follow-Up	1.0(3.9)	0.0(0.0)	1
	Withdrew Consent	3.0(11.5)	5.0(20.8)	8
	Other	3.0(11.5)	0.0(0.0)	3

Of these patients 7 STS and 10 placebo were classified as relapsed by the investigator, but not by the sponsor. Another 10 STS and 15 placebo withdrew for lack of efficacy. If we consider these patients relapsed then we have a total of 39 (26%) STS and 65 (40%) placebo relapse cases. The difference is significant using the Cochran-Mantel-Haenszel test (p=0.0061) of the proportions or the Cox PH model for the time until the event (p=0.0197). So the results are still significant, but there are still more STS patients (19 STS vs. 10 placebo) who weren't adequately followed up on and that means there is extra uncertainty in the proportions relapsed. It is an exploratory analysis but if we were to eliminate the need for confirmation then we would have 64 (43%) STS and 87 (53%) placebo cases and the results would be mixed: Cochran-Mantel-Haenszel test p=0.0570 and Cox PH model p=0.0478 with a hazard ratio of 0.721. Note that this approach counts 6 STS and 12 Placebo patients who had "abnormal" values and then had "normal" values at the interim visit as relapsed.

Patients that were inadequately followed up on using 3+ day criteria for interim visit
Note that if we allowed interim visits 3 or more days later then we would have a greater imbalance: 33 STS and 25 placebo patients that may have relapsed but weren't followed long enough for us to be sure. Another reason to prefer the 11+ day criteria for the interim visit is that there was one patient who had "abnormal" values confirmed at an interim visit 6 days later but

continued in the trial and had "normal" values at the next visit (20 days later) and there was one
patient who had "abnormal" values confirmed 8 days later but continued in the trial and had
"normal" values at the next visit (15 days later). This suggests that an interim visit as little as 3
days after the first "abnormal" values are observed may not give a reliable confirmation.

Center Differences

There was some variation among centers but it doesn't seem to have had much effect on the
results. The number of randomized patients per center ranged from 1 to 24. Centers 08, 01, 14,
15, 18 had the largest differences between treatments in proportion relapsed.

Table 3.18 Centers with largest group differences for % Relapsed (11+ day interim)

	Placebo		STS		
	Relapsed	Not Relapsed	Relapsed	Not Relapsed	% Difference
Center	n (%)	n (%)	n (%)	n (%)	
01	4.0(50.0)	4.0(50.0)	0.0(0.0)	9.0(100.0)	50.0
08	4.0(80.0)	1.0(20.0)	0.0(0.0)	4.0(100.0)	80.0
14	6.0(66.7)	3.0(33.3)	1.0(11.1)	8.0(88.9)	55.6
15	1.0(11.1)	8.0(88.9)	4.0(40.0)	6.0(60.0)	-28.9
18	4.0(36.4)	7.0(63.6)	0.0(0.0)	11.0(100.0)	36.4

3.2. Evaluation of Safety

See Clinical Review by Dr. Greg Dubitsky.

APPEARS THIS WAY
ON ORIGINAL

APPEARS THIS WAY
ON ORIGINAL

4. Findings in Special/Subgroup Populations

4.1. Gender, Race, and Age
4.1.1. Gender

The treatment differences appear reasonably consistent across gender subgroups.

Table 4.1Study 0052: Change from Baseline in HAM-D$_{1-28}$

		FEMALE	MALE
Placebo	N	68.0	60.0
	Mean Base	28.9(3.8)	28.3(4.1)
	Mean Change	-8.2(9.8)	-9.6(8.2)
STS	N	78.0	51.0
	Mean Base	28.7(4.0)	27.8(3.3)
	Mean Change	-11.0(8.8)	-11.2(8.3)
Between	P-value	0.09	0.22

Table 4.2 Study 9806: Number (Percent) Relapsed at 1 year for Gender

	Relapsed	STS	Placebo	P-value*
Female	No	90.0(84.9)	75.0(70.8)	0.0200
	Yes	16.0(15.1)	31.0(29.3)	
Male	No	34.0(79.1)	38.0(66.7)	0.1863
	Yes	9.0(20.9)	19.0(33.3)	

* based on Fisher's Exact test

APPEARS THIS WAY
ON ORIGINAL

4.1.2. Race

The vast majority of patients were Caucasian so it is difficult to say anything about the treatment effects in the other race categories. The negative finding in the Hispanic subgroup based on only 22 patients, in the acute therapy study 0052, is probably due to chance.

Table 4.3 Study 0052: Change from baseline in HAM-D$_{1-28}$ for Race

		ASIAN	BLACK	CAUCASIAN	HISPANIC	OTHER
Placebo	N	2.0	5.0	105.0	12.0	4.0
	Mean Base	25.5(3.5)	28.6(1.9)	28.5(4.1)	29.7(3.2)	31.5(3.1)
	Mean Change	-8.0(12.7)	-8.4(11.5)	-8.9(9.1)	-10.9(9.6)	-4.0(5.0)
STS	N	1.0	6.0	105.0	12.0	5.0
	Mean Base	23.0(.)	27.0(1.3)	28.3(3.7)	29.2(4.2)	30.8(4.0)
	Mean Change	-6.0(.)	-15.0(6.4)	-11.6(8.9)	-5.2(5.4)	-10.2(5.8)
Between	P-value	.	0.53	0.02	0.07	0.11

Table 4.4 Study 9806: Number (Percent) Relapsed at 1 year for Race

	Relapsed	STS	Placebo	P-value*
Asian	No	1.0(100.0)	1.0(50.0)	1.000
	Yes	0.0(0.0)	1.0(50.0)	
Black	No	7.0(87.5)	9.0(100.0)	0.471
	Yes	1.0(12.5)	0.0(0.0)	
Caucasian	No	103.0(83.1)	88.0(65.7)	0.002
	Yes	21.0(16.9)	46.0(34.3)	
Hispanic	No	9.0(81.8)	11.0(78.6)	1.000
	Yes	2.0(18.2)	3.0(21.4)	
Other	No	4.0(80.0)	4.0(100.0)	1.000
	Yes	1.0(20.0)	0.0(0.0)	

* based on Fisher's Exact test

4.1.3. Age

Table 4.5 Study 0052: Change from Baseline in HAM-D$_{1-28}$ for Age

Treatment		<=30	31-40	41-50	>50
Placebo	N	24.0	35.0	39.0	30.0
	Mean Base	28.4(4.0)	28.6(3.5)	28.5(4.6)	29.0(3.8)

	Mean Change	-10.9(11.0)	-7.6(7.5)	-9.8(8.7)	-7.5(9.5)
STS	N	30.0	27.0	36.0	36.0
	Mean Base	27.7(3.8)	27.9(3.0)	29.8(4.3)	27.8(3.3)
	Mean Change	-8.0(8.1)	-11.3(7.7)	-12.6(9.3)	-12.1(8.7)
Between	P-value	0.36	0.06	0.37	0.03

Notice that Placebo was numerically better than STS in the <=30 age group. This is suggestive of a qualitative interaction between age and treatment. A test of the interaction between treatment and age group yields a p-value of 0.096. The p-value for interaction is more significant (p=0.043) if age is not divided into groups and is assumed to have a linear relationship with the change in the HAM-D$_{1-28}$. In the earlier positive study E106 STS was numerically better (-9.9 STS -5.8 Placebo) in the <= 30 age group, but the difference was not significant probably because of the small group size. Therefore, although there may be an age by treatment interaction in this study, suggesting that STS is only effective for higher ages, it is not a serious concern since it was not observed in the earlier study.

Table 4.6 Study 9806: Number (Percent) Relapsed at 1 year for Age

Age Group	Relapsed	STS	Placebo	P-value*
<=30	No	23.0(76.7)	20.0(80.0)	1.000
	Yes	7.0(23.3)	5.0(20.0)	
31-40	No	26.0(86.7)	28.0(77.8)	0.523
	Yes	4.0(13.3)	8.0(22.2)	
41-50	No	38.0(79.2)	38.0(64.4)	0.133
	Yes	10.0(20.8)	21.0(35.6)	
> 50	No	37.0(90.2)	27.0(62.8)	0.004
	Yes	4.0(9.8)	16.0(37.2)	

* based on Fisher's Exact test

In study 9806 in the age <=30 category a numerically larger percentage of STS patients relapsed than placebo patients. This was accommodated for in the other age groups, particularly in the age > 50 group where a significantly larger percentage of placebo patients relapsed (37% placebo vs. 10% STS).

4.2. Other Special/Subgroup Populations

No other special subgroups were examined.

5. Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

Acute Therapy

In the original submission of NDA 21336 there were four acute therapy studies. Three had an 8-week double-blind phase, like the new study 0052, and one had a 6-week double-blind phase. Only the 6-week study was found to be positive by the previous statistical reviewer, Dr. Yuan-li Shen. The following table summarizes the results for that 6 week randomized, placebo-controlled, double-blind, single-dose (STS 20 mg/ 20 cm^2) acute therapy study.

Table 5.1 Study S9303-E106-96B

	Treat		Baseline	Change
HAMD-17 (Primary)	Placebo	N	88	88
		Mean (SD)	23.30 (2.90)	-6.10 (6.67)
	STS	N	88	88
		Mean (SD)	22.86 (2.05)	-8.73 (7.53)
	Between	P-VALUE	0.1285	0.0130
HAMD-28 (Secondary)	Placebo	N	88	88
		Mean (SD)	30.78 (5.77)	-7.59 (8.75)
	STS	N	88	88
		Mean (SD)	29.69 (3.85)	-11.23 (9.87)
	Between	P-VALUE	0.1487	0.0039

Two important differences between the two acute therapy studies are that study E-106 had a 6-week double blind phase, while study 0052 had an 8-week double-blind phase and E-106 used the HAM-D$_{1-17}$ as primary endpoint, while 0052 used the HAM-D$_{1-28}$ as primary endpoint.

The new 8-week acute therapy study 0052 was also found to be positive for efficacy as summarized in the following table. The effect size for the HAM-D$_{1-28}$ was slightly smaller than that observed in study E-106 despite the larger sample size and longer duration. This might be because of the larger placebo response: -8.9 (0052) compared to -7.6 (E-106).

Table 5.2 Study 0052: Change from Baseline in HAM-D$_{1-28}$

	Treat		Baseline	Week 8
Modified ITT *	Placebo	N	128	128
		Mean (SD)	28.6 (4.0)	-8.9 (9.1)
	STS	N	129	129
		Mean (SD)	28.3 (3.7)	-11.1 (8.6)
	Between	P-VALUE	0.6150	0.0327
Observed Cases	Placebo	N	128	110
		Mean (SD)	28.6 (4.0)	-9.5 (9.3)
	STS	N	129	100
		Mean (SD)	28.3 (3.7)	-12.8 (8.0)
	Between	P-VALUE	0.6150	0.0053

* 8 patients were excluded from the modified ITT population because they lacked a post-baseline HAM D

Study 0052 was conducted in only three centers. Center 1 had a noticeably larger difference in treatment group mean changes in HAM-D$_{1-28}$ than the other two centers (3.6 points vs. 1.3 and

1.7) but all centers were numerically better for STS.

The sponsor reported that 110 of 128 (86.0%) and 100 of 129 (77.5%) STS patients completed the study. The mean change from baseline for STS patients who withdrew (-5.1) was numerically worse than that for Placebo patients who withdrew (-5.8). Therefore, it is not surprising that the group difference is more significant for the Observed Cases analysis which excludes the patients who withdrew. Since the dropout rate was not excessive and both the LOCF and the Observed Cases analyses indicate a significant treatment effect we conclude that there is no obvious bias in the results caused by the dropouts and that study 0052 is positive with regard to efficacy.

Long Term Maintenance

In study 9806, 163 patients were randomized to placebo and 159 to STS 20 mg/ 20 cm^2. Ten STS patients had no post-baseline efficacy measures and so they were not included in the analysis. Seven of these patients were lost to follow up, 2 withdrew consent, and 1 was non-compliant. According to the protocol patients were to be considered relapsed if they had a HAM-D_{1-17} >=14 and a CGI-s at least 2 points above their baseline CGI-s and the values were still above these levels 2 weeks later at an interim visit. CGI-s >=3 was another condition for relapse but since the CGI-s ranges between 1 and 7 this seems to be covered by the CGI-s 2 points over baseline condition. The sponsor reported that a blinded review of the data revealed that many interim visits took place less than the designated 2 weeks after the initial "abnormal" HAM-D_{1-17} and CGI-s scores so they decided that relapse could be confirmed as little as three days after "abnormal" HAM-D_{1-17} and CGI-s scores were observed. Using this approach the sponsor determined that 50 (30.7%) Placebo patients and 25 (16.8%) STS patients relapsed at or before 1 year. This difference is significant (p=0.0025) based on the primary analysis method, center stratified Cochran-Mantel-Haenszel test. The Cox Proportional Hazards Model of the time until relapse, adjusting only for treatment, was specified as a secondary analysis. The test for a treatment effect based on the Cox model was also significant (p=0.0074). There were 87 placebo and 64 STS patients who simultaneously had a HAM-D_{1-17} >=14 and CGI-s >= 2 +baseline CGI-s during their time in the study, but confirmation 2 weeks later was needed for them to be classified as relapsed. It is important to note that many of these patients were censored (lost or ceased participation) without first having an interim visit. Since these patients did not have a confirmatory visit they were classified as not relapsed although their last measurements were in the "abnormal" range. Furthermore, among the patients that ever had "abnormal" HAM-D_{1-17} and CGI-s scores there was a higher proportion of STS patients, 33 of 64 (52%), than placebo patients, 25 of 87 (29%), that did not have confirmation visits. Thus, the results may be biased in favor of the STS group.

Confirmation from an interim visit only 3 days after "abnormal" values are detected may not be reliable and the protocol required the interim visit to be 2 weeks later. Since visit windows were specified to +/3 days in the protocol we might require confirmatory visits to be at least 11 days

after the initial abnormal values. Using this criteria for the timing of interim visits we determine that 41 (25.2%) placebo and 23 (15.4%) STS patients relapsed. The p-value is 0.0223 for the Cochran-Mantel-Haenszel test and 0.0417 for the test based on the Cox proportional hazards model. In this case we still have a higher proportion of STS patients, 36 of 64 (56%), than Placebo patients, 35 of 87 (40%), that had "abnormal" values and were then censored without first having an interim visit. So again these results may be biased in favor of the STS group.

In summary, although the results are significant at face value they hide the fact that a higher proportion of STS patients had "abnormal" values and then ceased participation in the trial without first having an interim visit. If the criteria for relapse were less stringent and did not require confirmation of the "abnormal" HAM-D_{1-17} and CGI-s values then these patients would be considered relapsed. In fact, the proportions of patients in each group who ever simultaneously had a HAM-D_{1-17} >=14 and a CGI-s at least 2 points over baseline are not different based on the center-stratified Cochran-Mantel-Haenszel test (53% vs. 43% p=0.057). On the other hand, the Cox proportional hazards model based test for a treatment group difference in the time until first simultaneous occurrence of HAM-D_{1-17} >=14 and CGI-s 2 over baseline, is just significant (p=0.048). Note that this approach counts 6 STS and 12 Placebo patients who had "abnormal" values and then had "normal" values at the interim visit as relapsed.

It is also a little strange that 10 STS patients did not have any post-baseline efficacy measures but all Placebo patients did. Although it is very unlikely that all 10 STS patients would have relapsed, if they had it would have been enough to alter the outcome of the trial. On the other hand, if none of them relapsed the results would be slightly stronger.

In summary, although the results are positive there is extra uncertainty in the estimated proportions relapsed because many patients (more in the STS group) were not adequately followed up on.

5.2 Conclusions and Recommendations
With this resubmission of NDA 21336 the sponsor has added a new 8-week acute therapy study, 0052, to the 6-week acute therapy study that was previously found to be positive for efficacy. Two differences between these studies are the difference in lengths of the double-blind phases and the closely related but different primary endpoints HAM-D_{1-17} (study E-106) vs. HAM-D_{1-28} (study 0052). Study E-106 also had a smaller sample size (N=177 vs. N=265) and a larger treatment effect on the change in HAM-D_{1-28} (-3.6 vs. -2.2 based on the modified ITT-LOCF analysis). This may be because of a larger placebo effect in study 0052 and/or a lower completion rate. Seventy Six percent (76%)of STS patients completed study 0052 compared to 80% of placebo patients. Completion rates in study E-106 were slightly higher than study 0052: 89% for STS and 83% for placebo. Although there were differences in the studies, both studies were found to be positive for efficacy.

The long term maintenance study (9806) is more problematic. The study had two phases: a 10 week open label phase in which all patients received 20 mg/ 20 cm^2 STS, followed by a 52 week double blind, placebo controlled, double blind phase. Only those who had responded in the open-label phase, i.e, had a HAM-D$_{1-17}$ <= 10 at the end of visit 6 or 7 <u>and</u> 8 (weeks 8,9, and 10 respectively), were to be randomized to treatment with STS 20 mg/ 20cm^2 or matching placebo in the double-blind phase. Three hundred and forty two of the 675 patients enrolled into the open-label phase responded and 322 were randomized into the double-blind phase. The primary objective was to compare the treatment groups with respect to the proportions relapsed at or before 1 year of double-blind treatment. A patient was considered to have relapsed if

  a) he/she had a HAM-D$_{1-17}$ >= 14 and a CGI-s 2 points above his/her double-blind baseline CGI-s and

  b) the HAM-D$_{1-17}$ and CGI-s scores still exceeded these levels two weeks later at an "interim" visit.

It was also necessary that CGI-s >=3 but this seems to be redundant in light of condition a) since the CGI-s ranges from 1 to 7. It turned out that many patients did not wait to have an interim visit 2 weeks after the abnormal HAM-D$_{1-17}$ and CGI-s values. The sponsor addressed this by changing the necessary interval for an interim visit from 2 weeks to 3 days or more. They found a significant difference in the proportions relapsed (p=0.0025). However, this was a change from the protocol and it does not solve the problem because there are still many patients, in fact, a significantly higher proportion of STS patients that had "abnormal" values and then ceased participation in the trial without having any interim visit. For this reason, these results are likely biased in favor of the STS group. An analysis that requires interim visits to be at least 11 days after the visit at which "abnormal" values were detected is closer in spirit to the protocol. The results are still significant with these criteria but the STS group still has a higher proportion that were not adequately followed up on. Although it is an exploratory analysis we might remove the need for confirmation and consider as relapsed anyone who simultaneously had a HAM-D$_{1-17}$ >=14 and a CGI-s 2 above baseline in up to a year of participation. Under this new definition of relapse there are 87 (53%) placebo relapse cases and 64 (43%) STS cases. This difference is not significant using the primary analysis method, Cochran-Mantel-Haenszel test (p=0.057), but is significant using the Cox proportional hazards model for the time until the event (p=0.048). It is important to note that 12 of the 87 placebo patients and 6 of the 64 STS cases had normal values at an interim visit. So they were sufficiently followed up on under the original definition of relapse and were found not to meet that definition. Another curious finding is that 10 STS patients had no-post baseline efficacy measures whereas all placebo patients did. Therefore, 10 STS patients were not included in the analysis. Although it is very unlikely that all 10 of these patients would have relapsed, 10 additional STS cases would have been enough to make the difference in proportions insignificant. On the other hand, if none of them relapsed the results would have been slightly stronger.

In summary, it seems that the results of this study are more modest than the sponsor's p-values

suggest. The reason for this is that a significantly higher proportion of STS patients had "abnormal" $HAM-D_{1-17}$ and CGI-s scores and then ceased participating in the study, so we can not be sure whether they relapsed or not. The sponsor's analysis treats these patients as not relapsed which favors the STS group. Since there are many of these patients it is likely that some of them would have been found to be relapsed if they had the necessary interim visit. This means that there is extra uncertainty in the estimated proportions relapsed and the group differences. In fact, the results are no longer significant if all of these patients truly relapsed. However, the study is still considered positive since most reasonable attempts to re-classify these patients based on reasons for withdrawal, including lack of efficacy and relapsed in the opinion of the investigator, still produce significant results.

Tristan Massie, Ph.D.
Statistical Reviewer

Concurring reviewers:

Kun Jin, Ph.D.
Team Leader

Kooros Mahjoob, Ph.D.
Deputy Director, Division of Biometrics

I

/s/

Tristan Massie
12/19/03 02:53:17 PM
BIOMETRICS

Kun Jin
12/19/03 03:26:58 PM
BIOMETRICS

Kooros Mahjoob
12/22/03 03:13:41 PM
BIOMETRICS

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

STATISTICAL REVIEW AND EVALUATION

Medical Division: Division of Neuropharm Drug Products (HFD-120)
Biometrics Division: Division of Biometrics I (HFD-710)

NDA NUMBER:	21-336
SERIAL NUMBER:	0
DATE RECEIVED BY CENTER:	5/30/01
DRUG NAME:	Selegiline Transdermal System
INDICATION:	Major Depression
SPONSOR:	Sumerset
DOCUMENTS REVIEWED:	NDA submission
STATISTICAL REVIEWER:	Yuan-Li Shen (HFD-710)
STATISTICAL TEAM LEADER:	Kun Jin (HFD-710)
BIOMETRICS DIVISION DIRECTOR:	George Chi (HFD-710)
CLINICAL REVIEWER:	Greg Dubitsky (HFD-120)
PROJECT MANAGER:	Doris Bates (HFD-120)

Distribution: NDA 21-411
HFD-120/Doris Bates
HFD-120/Greg Dubitsky
HFD-120/Tom Laughren
HFD-120/Russell Katz
HFD-710/Yuan-Li Shen
HFD-710/Kun Jin
HFD-710/George Chi
HFD-700/Chuck Anello

File Directory: C:/NDA2001/selegiline-trns dermal_depression/
selegilline-trns dermal_depression.stat.doc

1

Table of Contents

1. BACKGROUND

Four Phase III trial results were submitted in this submission. Only one of the studies was a double-blind, 6-week, placebo-controlled study (S9303-E106-98B). The rest of the studies were 8-week studies.

2.   Study S9303-E106-96B

The study period was between January 29, 1997 to October 31, 1997. The final protocol was signed off on October 31, 1996. One amendment (February 11,1997) and one addendum (February 12, 1997) were made to the protocol. Most of the amendment items were related to study operation, e.g. timing of patch application was increased from 4-hour window (8 a.m. to 12 p.m.) to 6-hour window (6 a.m. to 12 p.m.), expanding exclusion criteria, timing of drug inventory/collection were changed, etc. The addendum was to allow patients who satisfactorily completed 6-week of double-blind treatment to continue to receive STS during a 3-month open-label extension period.

2.1   Study Design

This was a 6-week, double-blind, randomized, placebo-controlled, parallel assessment of the safety and efficacy of the selegiline transdermal system (STS) (20 mg /20cm^2) in study included a 1-week, single-blind, placebo run-in period. Patients who met the entry criteria were randomized to STS, once daily (i.e. every 24 hours) or placebo.

2.2   Objectives

The primary objective of this study was to evaluate the safety and efficacy of the STS. The secondary objective was to assess the pharmacokinetic parameters of selegiline, administered via the STS.

2.3   Efficacy Variables

Total score of Items 1 to 17 in the Hamilton Depression Scale (HAMD) was the primary efficacy assessment. HAMD includes 28 items designed to assess mood, feelings of guilt, suicide, agitation, anxiety, and the effect on work and related activities using the rating scale from 0 to 2 or 0 to 4 with 0 being the best rating, 2 or 4 being the worst rating.

The secondary assessment include total score of HAMD (items 1-28), Clinical Global Impression of Severity (CGI-s), Clinical Global Impression of Change (CGI-c), Montgomery-Asberg Depression Rating Scale (MADRS) and MED-D. Note MED-D score was treated as the safety variables in the study report.

2.4   Sample Size

A total of 176 patient (88 per group) was planned with 80% power to detect 3-unit difference in HAM-D$_{1-17}$ change from baseline score between active drug group and placebo. This calculation was based on the assumption that the standard deviation of the mean change from baseline in each group was 7 and the significance level was 0.05.

2.5   Population and Statistical Analysis

In the primary efficacy analysis section of the protocol (XIII. Statistical Considerations and Data Analysis), it indicated that the evaluable patient subset would be used for the primary efficacy analysis. The evaluable patient subset is defined as those patients who do not violate the protocol and have completed the study.

Several populations were indicated in the protocol for the efficacy evaluation :

- Intent to treat (ITT): defined as all patients randomized who received at least one dose of medication and had baseline and at least one follow-up efficacy assessment.
- Classical Intent to treat: included patients who never received the assigned treatment.
- Traditional Observed Cases (OC): included patients who did not discontinue prematurely and were available for assessment at week 6.

However, in the study report, the protocol defined ITT population was called as "Modified ITT" population. The sponsor's new ITT population was defined as all randomized patients who were administered at least one dose of double-blind study drug. The new ITT population was used for the primary efficacy analysis in the study report.

The primary efficacy analysis for the HAM-D$_{1-17}$ was based on an analysis of covariance model with the week 6 value as the dependent variable. The independent variables included the baseline HAM-D$_{1-17}$, treatment group and center. The LOCF method was used to impute the data if outcomes at week 6 were unavailable. Two-sided test at the 5% level of significance was used.

2.6   Sponsor's Results

A total of 177 patients were randomized in this study (n=89 for STS and n=88 for placebo group). The majority of patients completed the study (88.8% for STS and 83% for placebo group) (Table 2.A.1). The most frequently report reason for premature discontinuation was lack of efficacy.

The demographic and baseline characteristics appeared to be compatible between treatment groups. There were more female ($\approx$60%) than male patients ($\approx$40%) in the

population. The majority of patients were Caucasians (96% for the STS group and 90% for the placebo).

Table 2.A.1   Patient Disposition by Treatment Group

	Selegiline N (%)	Placebo N (%)	Total N
Number of patients:			
Randomized[a]	89	88	177
Efficacy Analysis			
ITT ♣	89 (100)	88 (100)	177
Modified ITT ♦	88 (98.9)	88 (100)	176
Evaluable population	79 (88.8)	73 (83.0)	152
Who had w0 visit	89 (100)	88 (100)	177
Who had w1 visit	86 (96.6)	85 (96.6)	171
Who had w2 visit	85 (95.5)	86 (97.7)	171
Who had w3 visit	76 (85.4)	77 (87.5)	153
Who had w4 visit	84 (94.4)	76 (86.4)	160
Who had w6 visit	79 (88.8)	74 (84.1)	153
Who had follow-up visit (v9)	7 (7.9)	8 (9.1)	15
Who completed study	79 (88.8)	73 (83.0)	152
Who discontinued	10 (11.2)	15 (17.0)	25
Reason for discontinued			
Adverse Event	3	0	3
Withdrew consent	2	3	5
Noncompliance	0	2	2
Lost to follow-up	0	1	1
Other [b]	5	9	14

Note : a : All randomized patients : received study drug and was referred as the
  "safety population".
  b: These 14 patients were considered by the sponsor to have early discontinuation due
  to lack of efficacy.
  ♣: ITT : the sponsor's ITT stated in the report
  ♦ : modified ITT : the protocol specified ITT

Table 2.A.2   Patient Demography and Baseline Characteristics

	Selegiline (n=149)	Placebo (n=152)	p-value
Age			
N	89	88	0.286
Mean ± STD	41.4 ± 10.9	43.2 ± 10.8	
Median	42	43.0	
Range	19.0-62.0	20.0-65.0	
Sex			0.841
Female	53 (59.6)	53 (60.2)	
Male	36 (40.4)	35 (39.8)	
Race			0.338
Asian	0 (0)	3 (3.4)	
Black	2 (2.2)	2 (2.3)	
Caucasian	85 (95.5)	79 (89.8)	
Hispanic	2 (2.2)	4 (4.5)	
Weight (KG)			0.822
N	89	88	
Mean ± STD	80.9 ± 20.2	80.3 ± 19.5	
Median	77.6	78.9	
Height (cm)			0.489
N	89	88	
Mean ± STD	170.7 ± 10.0	169.9 ± 9.7	
Median	170.2	170.2	
HAMD 1-17 baseline			
N	89	88	0.299
Mean ± STD	22.9 ± 2.0	23.3 ± 2.9	
Median	22.0	22.0	
Range	20.0-28.0	20.0-35.0	
HAMD 1-28 baseline			
N	89	88	0.149
Mean ± STD	29.7 ± 3.8	30.8 ± 5.8	
Median	29.0	30.0	
Range	21.0-41.0	21.0-58.0	
MDRS baseline			
N	89	88	0.418
Mean ± STD	28.9 ± 5.3	29.5 ± 4.4	
Median	28.0	30.0	
Range	18.0-43.0	19.0-41.0	
CGI baseline	0	0	
Borderline ill	70 (78.7)	73 (83.0)	0.583
Moderately ill	19 (21.3)	14 (15.9)	
Markedly ill	0	1 (1.1)	
Severely ill	0	0	
Most Extremely ill			
Primary diagnosis			
Major Depressive Disorder			
Recurrent	57 (64.0)	62 (70.5)	0.288
Single	32 (36.0)	26 (29.5)	

There were no statistically significant differences observed with respect to vital signs, medical history and abnormal physical examination results. The patients' systolic/diastolic mean supine/standing blood pressure values at baseline were compatible. The most frequently reported medical history abnormality was "other" for the STS group and "urological/renal" for the placebo group.

With respect to prior medication history, 37 (41.65%) STS treated patients and 35 (39.8%) placebo patients had taken at least one medication within 90 days prior to the start of study. The most often used medication was nervous system agents, such as psychoanaleptic agents fluoxetine and paroxetine hydrocholride.

For the concomitant medication, 67 STS treated patients (75.3%) and 62 placebo patients (70.5%) used at least one medication during the course of the study. Similar to the prior medication history, nervous system agents, such as acetaminophen, were the most commonly used.

The sponsor calculated the treatment compliance based on the overall number of patches used divided by time in days (multiplied by 100), assuming that all patches not returned used the medication. The sponsor recognized that the assumption tended to inflate the percentage of compliance. Based on the calculation, 93.3% of STS treated patients and 87.5% of placebo patients were classified as compliant patients (by definition, compliance rate between 80% and 120% were compliant patients).

The sponsor presented their primary efficacy analysis based on different ITT population (defined as all randomized patients who had at least one dose of treatment, which was not the same as the protocol defined ITT population). The secondary analysis of the primary efficacy endpoint used the modified ITT population (which was actually the protocol defined ITT population) and evaluable population (defined as the population for the primary analysis). Since only one additional patients in STS group based on the sponsor's new defined ITT population (i.e. n=89 for STS group based on the sponsor's new defined ITT; n=88 for STS group based on the protocol defined ITT), the final results from two populations were very similar. Only results based on the protocol defined ITT population was presented (Table 2.A.3). The STS group showed more reduction (-8.7) based on the mean change from baseline to endpoint in the $HAMD_{1-17}$ total score as compared with placebo (-6.1) (p=0.013) using the protocol defined ITT population. The result was very similar to the finding based on the sponsor's ITT population (p=0.018) (Table 2.A.3). The statistical significance results based on $HAMD_{1-17}$ were also found at weeks 1 through 4 based on the protocol specified ITT population.

Based on the secondary efficacy variables, the STS group also showed statistically significant reduction on the symptom scores, such as $HAM-D_{1-28}$ (p=0.004) and MADRA (p=0.005). In addition, the STS group had more favorable responses based on CGI-s (p=0.015) and CGI-c (p=0.007). The sponsor also concluded that the sexual activity was not compromised within the STS treated group as compared with placebo based on the MED-D scores (p=0.063). However, this analysis was not based on ITT population.

The sponsor did not provide efficacy based on subgroup population.

Table 2.A.3 Efficacy Analysis (protocol defined ITT-LOCF) : Mean Scores and Mean Changes From Baseline to Endpoint

	Selegiline (n=88)	Placebo (n=88)	p-value #
HAM-D$_{1-17}$			
N	88	88	0.013
Mean ± STD	14.14 ± 7.87	17.19 ± 7.17	
Chg Mean ± STD	-8.73 ± 7.53	-6.10 ± 6.67	
HAM-D$_{1-28}$			
N	88	88	0.004
Mean ± STD	18.47 ± 10.31	23.19 ± 9.41	
Chg Mean ± STD	-11.23 ± 9.67	-7.59 ± 8.75	
MADRS			
N	88	88	0.005
Mean ± STD	19.08 ± 11.03	23.89 ± 10.28	
Chg Mean ± STD	-9.77 ± 11.52	-5.65 ± 9.07	
CGI-s (severity of illness)			
Normal, not ill	9 (10.2)	8 (9.1)	0.015
Borderline ill	18 (20.5)	7 (8.0)	
Mildly ill	24 (27.3)	16 (18.2)	
Moderately ill	29 (33.0)	45 (51.1)	
Markedly ill	6 (6.8)	10 (11.4)	
Severely ill	1 (1.1)	2 (2.3)	
Among the most	1 (1.1)	0 (0.0)	
CGI-c (change in severity of illness)			
Very much improved	17 (19.3)	8 (9.1)	0.007
Much improved	20 (22.7)	16 (18.2)	
Minimally improved	21 (23.9)	21 (23.9)	
Unchanged	23 (26.1)	30 (34.1)	
Minimally worse	3 (3.4)	9 (10.2)	
Much worse	4 (4.5)	4 (4.5)	
MED-D ♣			
N for the last visit	76	75	0.063
Mean ± STD	13.6 ± 7.3	14.4 ± 8.0	
N for the change from BL	71	65	
Chg Mean ± STD	-0.6 ± 6.3	1.1 ± 6.8	

Note : ♣ : MED-D, which grades the symptoms that described patients' feeling, activity during the past week to 10 days, was treated as an efficacy variable in the protocol, but was treated as safety variable in the study report.

: The p-values for the continuous variables were based on the ANCOVA model = treatment+center+baseline score.

2.7   Reviewer's Evaluation and Comments

Since there was only one more patient in the sponsor's ITT population than that in the protocol specified ITT population, the analysis result based on the protocol specified ITT population was almost identical to the sponsor's ITT population. In this review, only the results based on the protocol specified ITT population were presented.

This reviewer had confirmed the sponsor's finding that the selegiline group had significant improvement (mean change from baseline of HAM-D $_{1-17}$ =-8.74) over the placebo group (mean change from baseline of HAM-D $_{1-17}$ =-6.16) (p=0.015). This p-value was slightly different from the sponsor's p-value (p=0.013) due to the three different records (for patients 0104, 1302 and 2004) between the sponsor's and the reviewer's analysis dataset. Similar reason as illustrated in Appendix I may be applied.

The significant treatment effect was also demonstrated based on the HAM-D$_{1-28}$ total scores and MADRA total scores (p=0.005 and 0.002 for HAM-D$_{1-28}$ and MADRA, respectively).

The effect of subgroup was evaluated by including age (continuous or dichotomized : 42 years old or >42 years old), gender, racial group (Caucasian or non-Caucasian) in the ANCOVA model, individually. All the results supported the primary efficacy analysis result in favor of selegiline (Table 2.B.2).

Table 2.B.1   P-values for the treatment effect based on the ANCOVA model using the protocol specified ITT population

Change from baseline measures	P-value for treatment effect (Protocol specified ITT) STS : n=88 Placebo : n=88
HAM-D$_{1-17}$	0.0147
HAM-D$_{1-28}$	0.0047
MADRA	0.002

Note : @ : model = treatment+site+baseline score

Table 2.B.2 P-values for treatment effect based on ANCOVA model, adjusted for age, gender and race subgroup, individually

Model for Change from baseline in HAM-D $_{1-17}$ total scores	P-value for treatment effect from Sponsor's ITT population STS : n=88 Placebo : n=88
Treatment+site+baseline score + age group ♣	0.0155
Treatment+site+baseline score + age (continuous variable)	0.0184
Treatment+site+baseline score + race group ♠	0.0143
Treatment+site+baseline score + gender	0.0140

Note : ♣: age group was dichotomized as <=42 years old and > 42 years old.
        ♠ : race group was defined as Caucasian and non-Caucasian.
        ♦ : p=0.0184 if continuous age variable was included in the model.

3. Study S9303-P9804

This study started from November, 1998 and ended in December, 1999. The original protocol was finalized on October 16, 1998 with one amendment on January 15, 1999. The most significant change in the amendment was the change of the primary efficacy endpoint from using all HAMD total scores (items 1-28) to using the first 17 of the 28 items in calculating HAMD total scores.

3.1   Study Design

This is an 8-week, double-blind, placebo-controlled, parallel assessment of the safety and efficacy of single daily doses of the Selegiline Transdermal System (STS) in patients with major depression. Following a 1-week, single-blind, run-in period, eligible patients were randomized to either 1.0 mg/cm^2 of selegiline or placebo. Treatment patch was applied between 8:00 a.m. and 12:00 p.m. each day for 8 weeks. The STS was supplied as a 20 cm^2 patch containing 1.0 mg/cm^2 which is anticipated to produce a steady-state plasma concentration of 2.3 ng/mL.

3.2   Objectives

The objective of this study is to evaluate efficacy and safety of the STS in patients with depression.

3.3   Efficacy variables

The primary efficacy variable was the change from baseline to week 8, Hamilton Depression Rating Score Total 17 items (HAM-D$_{1-17}$). The secondary efficacy variables

included change from baseline in HAM-D$_{total}$, HAM-D depressed mood item, MADRS and MED-D scale. Additional secondary efficacy variables were proportion of patients who improved in CGI (score 5-7 vs. 1-4) and patients who did not worsen in CGI (scores 4-7 vs. 1-3). Note that MED-D scale was treated as a safety variable in the study report.

3.4   Sample Size

A total of 125 patient per group was planned with 80% power to detect 2.5-unit difference in HAM-D$_{1-17}$ change from baseline score between active drug group and placebo, assuming 0.05 significance level and the standard deviation of the mean HAM-D$_{1-17}$ change from baseline scores as 6.8.

3.5   Population and Statistical Analysis

The protocol specified that two type of populations will be used for efficacy analysis :

1. Intent-to-treat (ITT) patients : all randomized patients who took at least one dose of study drug and had at least one post-baseline measurement of HAM-D$_{1-17}$;
2. Evaluable patients : all ITT patients who met all inclusion criteria and had no significant exclusion criteria present.

Both the observed data and the Last Observation Carry Forward data would be used for efficacy analyses.

The primary analysis was ANCOVA with center and treatment as main effects and baseline HAM-D$_{1-17}$ as a covariate. However, an additional covariate : age was added in the ANCOVA model in the report.

For the categorical variables, the sponsor indicated in the protocol that linear modeling test for ordinal variables using center and treatment as main effects and baseline measurements as covariate (SAS Proc CATMOD) would be used. However, in the study report, the sponsor indicated that for categorical efficacy variables, CMH test controlling for center and baseline scores was used for testing treatment difference. For percentage of worsen score in CGI-c (defined as 5: minimally worse; 6: much worse vs. 1: very much improved; 2: much improved; 3: minimally improved; 4: unchanged) or improved score in CGI-c (defined 1: very much improved; 2: much improved 3: minimally improved vs. as 4: unchanged; 5: minimally worse; 6: much worse vs.) the center stratified Mantel-Haenszel mean score (CMH) test will be used.

The sponsor indicated that a treatment-by-center interaction would be evaluated by fitting interaction term in the ANCOVA. If no significant interaction was detected ($p<0.05$), the interaction term would be excluded from the model, otherwise a treatment-by-center plot will be presented to evaluate the nature of the interaction. The sponsor indicated that no further evaluation would be performed if the interaction was found to be quantitative

(differences in magnitude). If the interaction was qualitative (difference in direction), exploratory analyses will be performed for further evaluation of the confounding effect.

3.6   Sponsor's Results

A total of 365 patients were screened for this study. 310 of these patients were randomized to STS or placebo (153 patients for STS and 157 for placebo). Nine patients from site 12 were excluded from all efficacy and safety analyses due to legal/regulatory matters currently under active federal investigation. The sponsor indicated that the exclusion of these patients would not affect the power since additional 9 patients were enrolled at other sites. There were 301 patients included in the efficacy and safety analysis. Of these patients, 219 patients completed the study and 82 patients were prematurely discontinued from the study. The most frequently reported reason for early discontinuation was lost to follow-up for STS patients and withdrew consent and lost to follow-up for placebo patients (Table 3.A.1).

Patients' demographic data such as age, race, gender, appeared to be compatible between treatment groups. The placebo treated patients (43.5 years old) were about 2 year older than the STS treated patients (41.2 years old), although it was not statistically significant. The majority of the patients were female (63% and 65% for STS and placebo group, respectively) and Caucasian (77% and 88% % for STS and placebo group, respectively).

The baseline severity based on HAM-D$_{1-17}$, HAM-D$_{1-28}$, MDRS or CGI seem compatible between treatment groups. The majority of patients were classified as moderately ill based on CGI scores.

According to the sponsor, the vital sign, medical history and physical examination results between treatment groups were also similar. There were 63.8% STS treated patients and 71.1% placebo treated patients who had at least one medication within 90 days prior to the start of the study. The most commonly used medication was the analgesic paracetamol (acetaminophen). With regard to concomitant medications, there were 75.2% STS treated patients and 81.6% treated patients used at least one medication during the course of the study. The most often used medication were alimentary tract, metabolism agents and nervous system agents, e.g. multivitamins, paracetamol.

The sponsor's primary efficacy analysis was based on ITT population defined as all randomized patients who had at least one dose of treatment (which was not the same as the protocol defined ITT population). The secondary analysis of the primary efficacy endpoint (change from baseline to endpoint of HAM-D$_{1-17}$ total scores) used the modified ITT population which was actually the protocol defined ITT population.

The sponsor presented the primary efficacy analysis results were based on ANCOVA model with treatment, center, baseline scores and age in the model (note that age as a covariate was added in the study report and was not specified in the protocol).

Table 3.A.1 Patient Disposition by Treatment Group

	Selegiline N (%)	Placebo N (%)	Total N
Number of patients:			
Screened			365
Randomized	153	157	310
Randomized at	4	5	
center 12			9
Efficacy Analysis			
ITT ♣	149 (100)	152 (100)	301
Modified ITT ♦	145 (97.3)	144 (94.7)	289
Evaluable population	108 (72.5)	111 (73.0)	219
Who had w0 visit	149 (100)	152 (100)	301
Who had w1 visit	137 (91.9)	141 (92.8)	278
Who had w2 visit	137 (91.9)	139 (91.4)	276
Who had w4 visit	121 (81.2)	124 (81.6)	245
Who had w6 visit	112 (75.2)	115 (75.7)	227
Who had w8 visit	107 (71.8)	109 (71.7)	216
Who completed study	108 (72.5)	111 (73.0)	219
Who discontinued	41 (27.5)	41 (27.0)	82
Reason for discontinued [a]			
Adverse Event	10 (6.7)	8 (5.3)	18
Non-compliance	1 (0.7)	2 (1.3)	3
Lost to follow-up [b]	18 (12.1)	10 (6.6)	28
Withdrew consent	5 (3.4)	10 (6.6)	15
Protocol violation	3 (2.0)	2 (1.3)	5
Pregnancy	1 (0.7)	0 (0.0)	1
Other	3 (2.0)	9 (5.9)	12

Note [a] : 11 patients who were prematurely discontinued from study were considered by the sponsor to
Have been discontinued due to lack of efficacy (3 STS patients, 8 placebo patients).

[b] : One patient (patient 605) in the STS group was noted as "lost to follow-up" on the study
completion page, but was noted as having discontinued due to AE.

♣: ITT : the sponsor's ITT stated in the report

♦ : modified ITT : the protocol specified ITT

Note : Patients were classified as having had a particular visit based on an algorithm that accounted for
visit window. Patients were classified as having completed study based on the study completion
page of the CRF. So the number of patients who completed the study may not necessarily be the
same as the number of patients who had a week 8 visit.

Table 3.A.2 Patient Demography and Baseline Characteristics

	Selegiline (n=149)	Placebo (n=152)	p-value
Age			
N	149	152	0.073
Mean ± STD	41.2 ± 11.6	43.5 ± 10.0	
Median	42	43.5	
Range	19.0-64.0	19.0-65.0	
Sex			
Female	94 (63.1)	99 (65.1)	0.664
Male	55 (36.9)	53 (34.9)	
Race			
Asian	1 (0.7)	1 (0.7)	0.326
Black	20 (13.4)	8 (5.3)	
Caucasian	114 (76.5)	134 (88.2)	
Hispanic	13 (8.7)	8 (5.3)	
Other	1 (0.7)	1 (0.7)	
Weight (KG)			
N	148	152	0.277
Mean ± STD	81.8 ± 21.9	84.7 ± 27.2	
Median	76.2	79.0	
Range	43.3-181.4	46.3-223.8	
Height (cm)			
N	149	152	0.928
Mean ± STD	168.8 ± 9.2	168.7 ± 8.8	
Median	167.6	167.6	
Range	149.9-193.0	147.3-188.0	
HAMD 1-17 baseline			
N	149	152	0.833
Mean ± STD	22.8 ± 3.0	22.9 ± 3.0	
Median	22.0	22.0	
Range	16.0-34.0	17.0-32.0	
HAMD 1-28 baseline			
N	149	152	0.166
Mean ± STD	29.0 ± 4.5	29.7 ± 5.1	
Median	29.0	28.0	
Range	18.0-42.0	19.0-45.0	
MDRS baseline			
N	149	152	0.845
Mean ± STD	28.3 ± 6.1	28.3 ± 6.0	
Median	29.0	28.0	
Range	8.0-42.0	13.0-48.0	
CGI baseline			
Mildly ill	5 (3.4)	5 (3.3)	0.254
Moderately ill	113 (75.8)	107 (70.4)	
Markedly ill	31 (20.8)	39 (25.7)	
Severely ill	0 (0.0)	1 (0.7)	

Based on the sponsor's "ITT" population with LOCF (Table 3.A.3), the sponsor obtained a marginally significant p-value (0.046) for the primary analysis of change from baseline of HAM-D$_{1-17}$ total scores. The result showed a larger reduction of HAM-D$_{1-17}$ total scores in Selegiline (mean= -7.87) as compared with placebo group (mean=-6.32).

However, this result became non-significant (p>0.05) if the "modified ITT" population was used (p=0.069) (see Table 3.A.4) (Note that this "modified ITT" population was the protocol specified analysis).

Table 3.A.3   Efficacy Analysis (sponsor's ITT-LOCF) : Mean Scores and Mean Changes from Baseline to Endpoint

	Selegiline (n=137)	Placebo (n=146)	p-value #
HAM-D$_{1-17}$			
N	149	152	0.046
Mean ± STD	14.95 ± 7.39	16.47 ± 7.06	
Chg Mean ± STD	-7.87 ± 7.29	-6.32 ± 6.78	
HAM-D$_{1-28}$			
N	149	152	0.022
Mean ± STD	18.98 ± 10.51	21.61 ± 9.33	
Chg Mean ± STD	-10.01 ± 9.05	-8.08 ± 8.61	
MADRS			
N	149	152	0.001
Mean ± STD	18.34 ± 10.11	21.95 ± 9.79	
Chg Mean ± STD	-9.93 ± 9.86	-6.37 ± 9.71	
CGI-s (severity of illness)			
Normal, not ill	12 (8.1)	5 (3.3)	0.024
Borderline ill	26 (17.4)	23 (15.1)	
Mildly ill	37 (24.8)	26 (17.1)	
Moderately ill	57 (38.3)	72 (47.4)	
Markedly ill	16 (10.7)	26 (17.1)	
Severely ill	1 (0.7)	0 (0.0)	
CGI-c (change in severity of illness)			
Very much improved	22 (14.8)	13 (8.6)	0.157
Much improved	40 (26.8)	33 (21.7)	
Minimally improved	41 (27.5)	40 (26.3)	
Unchanged	35 (23.5)	45 (29.6)	
Minimally worse	6 (4.0)	11 (7.2)	
Much worse	1 (0.7)	2 (1.3)	
MED-D ♣			
N for the last visit	109	115	0.559
Mean ± STD	10.6 ± 7.3	10.6 ± 7.2	
N for the change from BL	100	107	
Chg Mean ± STD	-0.7 ± 6.6	-0.4 ± 6.6	

Note : ♣ : MED-D, which grades the symptoms that described patients' feeling, activity during the past week to 10 days, was treated as an efficacy variable in the protocol, but was reported as a safety variable in the report.

: The p-values for the continuous variables were based on the ANCOVA model = treatment+center+baseline score+age; the p-values for the categorical variables were based on the CMH test , controlling for center.

Table 3.A.4 Efficacy Analysis (protocol defined ITT-LOCF) : Mean Scores and Mean Changes from Baseline to Endpoint

	Selegiline (n=145)	Placebo (n=144)	p-value #
HAM-D$_{1-17}$			
N	145	144	0.069
Mean ± STD	14.71 ± 7.29	16.32 ± 7.15	
Chg Mean ± STD	-8.08 ± 7.27	-6.67 ± 6.80	
HAM-D$_{1-28}$			
N	145	144	0.039
Mean ± STD	18.67 ± 9.41	21.26 ± 9.37	
Chg Mean ± STD	-10.28 ± 9.02	-8.53 ± 8.62	
MADRS			
N	145	144	0.001
Mean ± STD	18.05 ± 10.06	21.75 ± 9.93	
Chg Mean ± STD	-10.21 ± 9.85	-6.72 ± 9.86	
CGI-s (severity of illness)			
Normal, not ill	12 (8.3)	5 (3.5)	0.055
Borderline ill	26 (17.9)	23 (16.0)	
Mildly ill	37 (25.5)	26 (18.1)	
Moderately ill	55 (37.9)	66 (45.8)	
Markedly ill	14 (9.7)	24 (16.7)	
Severely ill	1 (0.7)	0 (0.0)	
CGI-c (change in severity of Illness)			
Very much improved	22 (15.2)	13 (9.0)	0.157
Much improved	40 (27.6)	33 (22.9)	
Minimally improved	41 (28.3)	40 (27.8)	
Unchanged	35 (24.1)	45 (31.3)	
Minimally worse	6 (4.1)	11 (7.6)	
Much worse	1 (0.7)	2 (1.4)	

: The p-values for the continuous variables were based on the ANCOVA model = treatment+center+baseline score+age; the p-values for the categorical variables were based on the CMH test , controlling for center for CGI-C and controlling for center and baseline CGI-S for CGI-S. .

In the analysis of the secondary endpoints, the sponsor obtained statistical significant results in favor of STS from HAM-D$_{1-28}$ (p=0.022), MADRA (p=0.001) and CGI-s (p=0.024) based on their "ITT" analysis with LOCF. Based on the "modified ITT", only HAMD-D$_{1-28}$ (p=0.039) and MADRA (p=0.001) showed statistical significance.

The sponsor did not have subgroup analysis to demonstrate the consistency among the subgroups, such as age, gender or status of whether patients had history of antidepressant use.

3.7   Reviewer's Evaluation and Comments

Since the sponsor's efficacy analysis deviated slightly from the protocol specified
analysis, this reviewer performed the analysis based on the following specification :

1. This reviewer performed the protocol specified ITT analysis for the efficacy analysis
    along with the sponsor's ITT analysis.

    Note that the protocol specified ITT was defined as all randomized patients who took
    at least one dose of study drug and had at least one post-baseline measurement of
    HAM-D_{1-17} (indicated as "modified ITT" in the sponsor's study report). The
    sponsor's ITT was defined as all randomized patients who had at least one dose of
    treatment.

    There were more patients in the sponsor's ITT analysis (149 STS treated
    patients and 152 placebo patients) than that in the protocol specified ITT analysis
    (145 STS treated patients and 144 placebo patients). The discrepancy was due to that
    12 patients did not have post baseline HAM-D_{1-17} measures. However, the sponsor
    carried over the baseline value to the endpoint for these 12 patients to form their ITT
    analysis.

2. The reviewer performed the ANCOVA analysis based on models with or without the
    age in the model since the sponsor used the ANCOVA model adjusted for age which
    was not pre-specified in the protocol.

These analysis results were shown in Table 3.B.1. The result indicated that if the
protocol specified analysis was used (not including age in the model and using the
protocol specified ITT population), the selegiline treated group would not show
statistically significant improvement over the placebo treated group (p=0.0844; the mean
change from baseline for the HAM-D $_{1-17}$ scores was –8.08 and –6.71 for the STS group
and placebo group, respectively). With the age in the model, the p-value for the treatment
effect would be 0.0781 (as compared with p=0.069 based on the sponsor's analysis) if the
protocol specified ITT population was used. This reviewer does not agree with the
sponsor's approach of carrying over the baseline data to endpoint in their ITT analysis for
the 12 patients (8 placebo patients and 4 STS treated patients). Potential bias may be
introduced due to the imbalance of the patient distribution. This reviewer adopted the
protocol specified ITT population for the analysis. The results based on the protocol
specified ITT population did not support the sponsor's marginal significant finding
(p=0.046), regardless of whether age was included in the ANCOVA model or not.

The results in Table 3.B.1 from the reviewer's analysis with baseline values carried over
and based on ANCOVA model adjusted for age were a little different from the sponsor's
results, since the endpoint value for patient 0202 obtained by this reviewer was different
from that obtained by the sponsor. More detailed of this discrepancy was described in
Appendix I.

18

In addition to the non-significant endpoint result (i.e. at week 8 with LOCF), there was not any nominal significant results found in any endpoint prior to week 8 (the results were shown in the sponsor's Table 10.7 of study report based on the modified ITT-LOCF) : p-values were 0.54, 0.74, 0.36 and 0.25 for weeks 1, 2, 4, 6 respectively.

With regard to the secondary analysis, selegiline treated group showed marginal significant improvement over the placebo group based on the change from baseline of HAM-D$_{1-28}$ score using the protocol specified ITT population (p=0.043 based on ANCOVA model including age and p=0.053 based on model without age). Selegiline treated group showed nominal significant improvement based on MADRA scales using the protocol specified ITT population (see Table 3.B.1; p=0.0013 and p=0.0020 for model with or without age, respectively). For the evaluation of CGI, instead of using Weighted-least-squared method(as indicated in the protocol section 13.6.2 Efficacy Analysis : linear modeling test for ordinal variables using center and treatment as main effects and baseline measurements as covariate [SAS Proc CATMOD]), the sponsor used CMH test adjusting for center, baseline score (for CGI-s) and age (see November 16, 2001 submission). Since CMH test stratified by continuous variables is not a valid test, this reviewer used CMH test adjusting for center only and performed ranked ANCOVA to confirm the results (Table 3.B.2). The results showed nominal significant treatment effect in favor of STS.

Another secondary efficacy variables were based on dichotomized CGI-c scores, i.e. improved and worsen scores (Table 3.B.2). The results showed that the association of treatment by worsen/improved scores was not significant when center adjusted CMH test was used.

This reviewer also fitted ANCOVA model for HAM-D $_{1-17}$ scores with treatment, site and baseline score in the model, adjusted for age group (<=42 years old or >42 years old), gender and racial group (Caucasian and non-Caucasian), individually, using protocol specified ITT population (see Table 3.B.3). There were no significant treatment effect found based on these models.

Table 3.B.1 P-values for the treatment effect based on the protocol specified ITT / The Sponsor's ITT analysis and for the ANCOVA model with or without age in the model

Change from baseline measures	ANCOVA Model @	P-value for treatment effect (Protocol specified ITT) STS : n=145 Placebo : n=144	P-value for treatment effect (Sponsor's ITT) STS : n=149 Placebo : n=152
HAM-D$_{1-17}$	Not including age	0.0884	0.0586
	Including age	0.0781	0.0519
HAMD-D$_{1-28}$	Not including age	0.053	0.025
	Including age	0.043	0.031
MADRA	Not including age	0.002	0.0014
	Including age	0.0013	0.0009

Note : @ : model = treatment+site+baseline score

Table 3.B.2 P-values for treatment effect based on Cochran-Mantel-Haenszel Statistics and ranked ANCOVA, adjusting for center

Variable	Analysis	P-value for treatment effect (Protocol specified ITT) STS : n=145 Placebo : n=144
CGI-s	CMH, stratified by center	0.0078
	Ranked ANCOVA, adjust for center	0.0116
CGI-c	CMH, stratified by center	0.0139
	Ranked ANCOVA, Adjust for center	0.0231
	Improved vs. non-improved ♣	0.055
	Non-worsen vs. worsen ♦	0.186

Note : ♣ : improved score in CGI-c: improved : 1=very much improved; 2=much improved 3=minimally improved vs.
non-improved : 4=unchanged; 5=minimally worse; 6=much worse.
♦ : worsen score in CGI-c : Not-worsen : 1=very much improved; 2=much improved; 3=minimally improved; 4=unchanged
vs. worsen : 5=minimally worse; 6=much worse.
The p-values for the two analyses were based on center-stratified CMH

Table 3.B.3 P-values for treatment effect based on ANCOVA model, adjusted for age, gender and race subgroup, individually

Model for Change from baseline in HAM-D $_{1-17}$ total scores	P-value for treatment effect from protocol specified ITT population STS : n=145 Placebo : n=144
Treatment+site+baseline score + age group ♣	0.0767
Treatment+site+baseline score + race group ♠	0.1255
Treatment+site+baseline score + gender	0.0786

Note : ♣: age group was dichotomized as <=42 years old and > 42 years old.
   ♠ : race group was defined as Caucasian and non-Caucasian.

4.   Study S9303-E114

The study period was between March 11, 1999 to April 10, 2000. The final protocol was signed off on January 19, 1999. There were no protocol amendments for this study.

4.1   Study Design

This was an 8-week, double-blind, randomized, placebo-controlled, parallel-group, multicenter study to assess the safety and efficacy of the selegiline transdermal system 10 mg and 20 mg (per 20cm^2) in patients aged 18-65 with major depression. This study included a 1-week, single-blind, placebo run-in period prior to randomization to exclude early placebo responders. Patients were randomized to STS 10 mg, 20 mg and placebo in 1:1:1 ratio. A total of 19 US centers enrolled patients in this study.

4.2   Objectives

The primary objective of this study was to evaluate the safety and efficacy of the STS. The secondary objective was to assess the pharmacokinetic parameters of selegiline, administered via the STS.

4.3   Efficacy Variables

Similar to previous studies, the total score of Items 1 to 17 in the Hamilton Depression Scale (HAMD) was the primary efficacy assessment.

The secondary assessment include total score of HAMD (items 1-28), Clinical Global Impression of Severity (CGI-s), Clinical Global Impression of Change (CGI-c),

Montgomery-Asberg Depression Rating Scale (MADRS) and MED-D. Note MED-D score was treated as the safety variables in the study report.

4.4   Sample Size

One hundred and thirty seven patients per treatment group was planned with 80% power to detect 2.4-unit difference in HAM-D_{1-17} change from baseline score between two groups. This calculation was based on one-way analysis of variance for three groups, assuming standard deviation of 6.8 and the significance level of 0.05.

4.5   Population and Statistical Analysis

The protocol specified that ITT population, included all randomized patients who took at least one dose of study medication and had at least one on-treatment measurement of the primary efficacy variable, HAM-D_{1-17}, would be evaluated. The evaluable population (defined as all ITT population who met all inclusion criteria and had no significant exclusion criteria present) would also be evaluated. In the study report, one additional criterion: patients who had no influential deviation from the protocol was added for the evaluable population.

Similar to the previous two studies, the study report used different ITT population (defined as all randomized patients who were dispensed double-blind therapy) for the primary efficacy analysis. The modified ITT population (which is actually the protocol specified ITT population) was used for the secondary efficacy analysis. Only protocol specified ITT population would be presented in this review.

Safety will be assessed based on the safety population (defined as all randomized patients who received at least one dose of study medication).

Similar to study S9303-E106-96B, an ANCOVA model was used to compare the change from baseline of HAMD $_{1-17}$ between treatment groups. The model included treatment and center as main effects and the baseline measurement as a covariate. In the protocol, it indicates that the difference between each dose group versus placebo and between the two dose groups would be evaluated. However, no multiplicity adjustment was specified in the protocol.

The sponsor also indicated that treatment by center interaction would be evaluated. Since no evidence of statistical significance was found (p>0.1), the interaction term was not included in the final model.

4.6   Sponsor's Results

A total of 446 patients were randomized in this study (n=149 for 20 mg STS, n=151 for 10 mg STS and n=146 for placebo group). There were more than 70% of the patients

completed the study. The distributions of patients who discontinued from the study were compatible across treatment groups (Table 4.A.1).

The demography and patient characteristics appear to be compatible between treatment groups (Table 4.A.2). The only exception is that 20 mg STS group and placebo were heavier (mean values of 84 kg and 81 kg for 20 mg STS and placebo groups, respectively) than the 10 mg STS group (77 kg). The distributions of severity of illness at baseline (e.g. primary diagnosis, $HAMD_{1-17}$, $HAMD_{1-28}$, MADRAS, CGIS) also seem similar between treatment groups (Table 4.A.2).

The distributions of baseline vital signs were also compatible between treatment groups except that a significant treatment difference was found for supine systolic heart rate.

The only significant findings in distributions of the percentages of patients with abnormal medical history were in the incidence of musculoskeletal disorders (p=0.001) and allergies (p=0.046) with the 10 mg STS having the lowest incidence and 20 mg STS and placebo having similar higher incidences. The distribution in percentages of patients with abnormal musculoskeletal findings was also found to be significantly different between treatments (p=0.015).

The percentages of prior medication use and concomitant medication use were comparable between treatments. The most commonly prior medication was nervous system agent. The most commonly used medication among patients in all three-treatment groups was acetaminophen.

From the sponsor's efficacy analysis, there was not any significant overall treatment effect (Table 4.A.3) based on the primary (p=0.357) or secondary efficacy endpoint analysis (all p-values were greater than 0.1). The mean change from baseline in $HAMD_{1-17}$ was similar across treatment groups: -9.18, -9.02 and –8.12 for 20 mg, 10 mg STS and placebo, respectively. Using Fisher's protected approach, any pair-wise testing should not be performed (Note, no multiplicity adjusting method was proposed in the original protocol).

Table 4.A.1 Patient Disposition by Treatment Group

	Selegiline 20 mg N (%)	Selegiline 10 mg N (%)	Placebo N (%)	Total N
Number of patients:				
Randomized [a]	149	151	146	446
Efficacy Analysis				
ITT ♣	149 (100.0)	151 (100.0)	146(100.0)	446
Modified ITT ♦	142 (95.3)	151 (100.0)	142 (97.3)	435
Evaluable population	109 (73.2)	109 (72.2)	109 (74.7)	327
Who had w0 visit	149 (100.0)	151 (100.0)	146 (100.0)	446
Who had w1 visit	140 (94.0)	142 (94.0)	137 (93.8)	419
Who had w2 visit	136 (91.3)	150 (99.3)	134 (91.8)	420
Who had w4 visit	126 (84.6)	134 (88.7)	127 (87.0)	387
Who had w6 visit	116 (77.9)	120 (79.5)	116 (79.5)	352
Who had w8 visit	109 (73.2)	112 (74.2)	110 (75.3)	331
Who completed study	109 (73.2)	109 (72.2)	109 (74.7)	327
Who discontinued	40 (26.9)	42 (27.8)	37 (25.3)	119
Reason for discontinued				
Adverse Event	17 (11.4)	8 (5.3)	8 (5.5)	33
Lack of efficacy	5 (3.4)	13 (8.6)	7 (4.8)	25
Noncompliance	3 (2.0)	4 (2.7)	1 (0.7)	8
Lost to follow-up	7 (4.7)	4 (2.7)	8 (5.5)	19
Withdrew consent	6 (4.0)	10 (6.6)	9 (6.2)	25
Protocol violation	2 (1.3)	2 (1.3)	2 (1.4)	6
Other	0 (0.0)	1 (0.7)	2 (1.4)	3

Note : a : All randomized patients : received study drug and was referred as the
"safety population".
♣: ITT : the sponsor's ITT stated in the report
♦ : modified ITT : the protocol specified ITT

Table 4.A.2 Patient Demography and Baseline Characteristics

	Selegiline		Placebo	p-value
	20 mg (n=149)	10 mg (n=151)	(n=146)	
Age				
N	149	151	146	0.428
Mean ± STD	42.05 ± 11.19	40.36 ± 12.58	40.79 ± 11.11	
Median	44	39.00	42.0	
Range	19.0-66.0	17.00-64.00	19.0-63.0	
Sex				
Female	98 (65.77)	97 (64.24)	100 (68.49)	0.625
Male	51 (34.23)	54 (35.76)	46 (31.51)	
Race				
Asian	1 (0.67)	1 (0.66)	3 (2.05)	0.550
Black	10 (6.71)	8 (5.3)	7 (4.79)	
Caucasian	129 (86.58)	130 (86.09)	130 (89.04)	
Hispanic	7 (4.7)	7 (4.64)	5 (3.42)	
Other	2 (1.34)	5 (3.31)	1 (0.68)	
Weight (KG)				
N	149	151	146	0.011
Mean ± STD	84.03 ± 20.78	77.01 ± 19.78	81.02 ± 22.09	
Median	80.70	75.30	77.00	
Range	47.20-158.80	44.00-150.60	45.40-165.60	
Height (cm)				
N	149	151	146	0.174
Mean ± STD	169.48 ± 9.71	167.47 ± 9.48	168.39 ± 9.80	
Median	167.6	167.60	167.60	
	135.90-198.90	135.10-190.50	127.00-190.50	
HAMD 1-17 baseline				
N	149	151	146	0.254
Mean ± STD	23.30 ± 3.08	22.73 ± 3.18	23.06 ± 2.93	
Median	23.0	22.0	23.0	
Range	15.0-31.0	17.0-33.0	17.0-34.0	
HAMD 1-28 baseline				
N	149	151	146	0.441
Mean ± STD	29.64 ± 4.3	29.05 ± 5.07	29.55 ± 4.85	
Median	29.0	28.00	29.0	
Range	21.0-44.0	18.0-42.0	19.0-45.0	
MADRS baseline				
N	149	151	146	0.474
Mean ± STD	27.52 ± 5.15	26.93 ± 5.69	27.43 ± 5.51	
Median	28.00	27.00	28.00	
Range	13.00-41.00	8.00-40.00	14.0-43.0	
CGI baseline				
Mildly ill	5 (3.36)	6 (3.97)	4 (2.74)	0.083
Moderately ill	100 (67.11)	115 (76.16)	106 (72.60)	
Markedly ill	42 (28.19)	29 (19.21)	33 (22.60)	
Severely ill	2 (1.34)	0 (0.00)	3 (2.05)	
Unknown	0 (0.0)	1 (0.66)	0 (0.00)	
Primary diagnosis				
Recurrent	86 (57.72)	92 (60.93)	98 (67.12)	0.220
Single	63 (42.28)	59 (39.07)	48 (32.88)	

Table 4.A.3   Efficacy Analysis (protocol defined ITT-LOCF) : Mean Scores and Mean Changes From Baseline to Endpoint

	Selegiline		Placebo	p-value #
	20 mg (n=142)	10 mg (n=151)	(n=142)	
HAM-D$_{1-17}$				
N	142	151	142	Overall : 0.357
Mean ± STD	14.11 ± 7.70	13.71 ± 7.56	14.94 ± 7.80	10 mg vs plc : 0.314
Chg Mean ± STD	-9.18 ± 7.39	-9.02 ± 7.51	-8.12 ± 7.81	20 mg vs plc : 0.569
HAM-D$_{1-28}$				Overall : 0.449
N	142	151	142	10 mg vs plc : 0.299
Mean ± STD	17.72 ± 9.78	17.54 ± 9.55	18.85 ± 9.85	20 mg vs plc : 0.251
Chg Mean ± STD	-11.99 ± 10.06	-11.52 ± 9.88	-10.75 ± 10.18	
MADRS				Overall : 0.144
N	142	151	142	10 mg vs plc : 0.421
Mean ± STD	16.26 ± 10.45	17.35 ± 10.63	18.49 ± 10.96	20 mg vs plc : 0.051
Chg Mean ± STD	-11.16 ± 10.58	-9.58 ± 9.63	-8.96 ± 10.97	
CGI-s (severity of illness)				Overall : 0.156
Normal, not ill	22 (15.5)	20 (13.2)	15 (10.6)	10 mg vs plc : 0.159
Borderline ill	33 (23.2)	35 (23.2)	33 (23.2)	20 mg vs plc : 0.072
Mildly ill	32 (22.5)	33 (21.9)	24 (16.9)	
Moderately ill	42 (29.6)	52 (34.4)	51 (35.9)	
Markedly ill	11 (7.7)	11 (7.3)	17 (12.0)	
Severely ill	2 (1.4)	0 (0.0)	2 (1.4)	
CGI-c (change in severity of illness)				Overall : 0.156
Very much improved	33 (23.2)	33 (21.9)	26 (18.3)	10 mg vs plc : 0.141
Much improved	44 (31.0)	44 (29.1)	31 (21.8)	20 mg vs plc : 0.048
Minimally improved	28 (19.7)	28 (18.5)	35 (24.6)	
Unchanged	27 (19.0)	35 (23.2)	39 (27.5)	
Minimally worse	7 (4.9)	8 (5.3)	9 (6.3)	
Much worse	3 (2.1)	3 (2.0)	2 (1.4)	

Note :   # : The p-values for the continuous variables were based on the ANCOVA model = treatment+center+baseline score.

4.7   Reviewer's Evaluation and Comments

This reviewer confirmed that this study was a failed study based on the overall treatment effect from the primary endpoint (p=0.312; means change from baseline of the HAM-D$_{1-17}$ were -9.18, -9.02 and -8.06 for 20 mg, 10 mg and placebo, respectively). Note that the discrepancy of the reviewer's result and the sponsor's result was similar to the reason that illustrated in appendix I.

None of overall treatment effect was significant over time based on the primary endpoint. In addition, none of the secondary endpoints, such as $HAMD_{1-28}$, MADRA, CGI-s and CGI-c, showed statistical significance in the sponsor's analysis.

The sponsor did not provide any explanation of why the trial failed. This reviewer inspected the data and found that the effect size based on the change from baseline in $HAMD_{1-17}$ was relative larger than that from studies S9303-E106-96B and S9303-P9804 (see Table 6.). It suggests a possible placebo effect.

5. Study S9303-E113-98B

The study period was between August 31, 1998 to August 9, 1999. Two study amendments were made to this protocol on August 28, 1998 and January 15, 1999. Almost all of the amended items were related to study conduct and exclusion/inclusion criteria, etc., except that the second amendment expanded the description of statistical methodology in greater details.

5.1   Study Design

The design of this study is similar to study S9303-E114 except that this study only included two treatment arms. This was an 8-week, double-blind, placebo-controlled, parallel-group, multicenter study which randomized patients to either selegiline 20 mg /20cm^2 or placebo 20cm^2 in a 1:1 ratio. This study included a 1-week, single-blind, placebo run-in period prior to randomization and a baseline visit (visit 3) and 5 clinical visits (Visits 4-6) in the 8 weeks treatment period. This study included 13 US centers

5.2   Objectives

The primary objective of this study was to assess the safety and efficacy of STS.

5.3   Efficacy Variables

Similar to previous studies, the total score of Items 1 to 17 in the Hamilton Depression Scale (HAMD) was the primary efficacy assessment.

The secondary assessment include total score of HAMD (items 1-28), Clinical Global Impression of Severity (CGI-s), Clinical Global Impression of Change (CGI-c), Montgomery-Asberg Depression Rating Scale (MADRS) and MED-D for assessment of sexual function.

5.4   Sample Size

A sample size of 125 in each treatment group was planned to provide 80% power to detect a 2.5 unit treatment difference in change from baseline of $HAMD_{1-17}$. The calculation used 2-sided t-test with 0.05 significance level, assuming standard deviation of 6.8 in the mean change from baseline of $HAMD_{1-17}$.

5.5   Population and Statistical Analysis

Similar to all the previous studies, the protocol specified that ITT population (defined as all randomized patients who took at least one dose of study medication and had at least one on-treatment measurement of the primary efficacy variable, $HAM-D_{1-17}$) would be evaluated, while the study report used different ITT population (defined as all randomized patients who were dispensed double-blind therapy) for the primary efficacy analysis. Only protocol specified ITT population would be presented in this review.

Similar to previous studies, an ANCOVA model included treatment, center and baseline measurement was used to compare the change from baseline of $HAMD_{1-17}$ between treatment groups.

The sponsor also indicated that treatment by center interaction would be evaluated. Since no evidence of statistical significance was found ($p>0.1$), the interaction term was not included in the final model.

5.6   Sponsor's Results

A total of 297 patients were randomized to treatment. Two hundred-twenty patients completed the study (71.4% and 76.4% for STS and placebo groups, respectively). The most frequent reason for discontinuation is lost to follow-up. Among those premature discontinued patients, 14 patients were found to discontinue for reasons relating to lack of efficacy based on investigator assessment.

Treatments appeared to be compatible in the evaluation of patient demography and baseline characteristics. The majority of patients were female (>60%) and Caucasian (>80%). The majority of patients were classified as moderately ill (>70%) based on the baseline CGI score.

There were no significant treatment differences in vital signs, medical history or physical examination results except that placebo patients had higher percentage of patients with psychiatric illness history (66.7%) than did the STS patients (53.7%).

Overall, 64.2% STS patients and 54.1% placebo patients used at least one medication within 90 days prior to the start of the study. The most commonly used medication was nervous system agents (acetaminophen was the most commonly used within the class).

In addition, 80.3% STS patients and 71.9% placebo patients had at least one concomitant medication during the course of the study. Similar to the prior medication, the most commonly used concomitant medications was nervous system agent (acetaminophen was the most commonly used within the class).

Table 5.A.1   Patient Disposition by Treatment Group

	Selegiline N (%)	Placebo N (%)	Total N
Number of patients:			
Randomized [a]	147	150	297
Efficacy Analysis			
ITT ♣	147 (100)	150 (100)	297
Modified ITT ♦	137 (93.2)	146 (97.3)	283
Evaluable population	105 (71.4)	115 (76.7)	220
Who had w0 visit	147 (100)	150 (100)	177
Who had w1 visit	134 (91.2)	143 (95.3)	171
Who had w2 visit	134 (91.2)	137 (91.3)	171
Who had w4 visit	118 (80.3)	128 (85.3)	153
Who had w6 visit	110 (74.8)	121 (80.7)	160
Who had w8 visit	104 (70.7)	112 (74.7)	153
Who completed study	105 (71.4)	115 (76.7)	220
Who discontinued	42 (28.6)	35 (23.3)	77
Reason for discontinued			
Adverse Event			
Noncompliance	11 (7.5)	5 (3.3)	16
Lost to follow-up	1 (0.7)	2 (1.3)	3
Withdrew consent	15 (10.2)	14 (9.3)	29
Protocol violation	5 (3.4)	11 (7.3)	16
Pregnancy	1 (0.7)	1 (0.7)	2
Other	1 (0.7)	0 (0.0)	1
	8 (5.4)	2 (1.3)	10

Note : a : All randomized patients : received study drug and was referred as the
        "safety population".
    ♣: ITT : the sponsor's ITT stated in the report
    ♦ : modified ITT : the protocol specified ITT

29

Table 5.A.2 Patient Demography and Baseline Characteristics

	Selegiline (n=147)	Placebo (n=150)	p-value
Age			0.293
N	147	150	
Mean ± STD	41.04 ± 11.41	39.79 ± 10.65	
Median	40	40.0	
Range	18.0-65.0	18.0-64.0	
Sex			0.843
Female	91 (61.9)	91 (60.7)	
Male	56 (38.1)	59 (39.3)	
Race			0.400
Asian	3　(2.0)	0　(0.0)	
Black	10 (6.8)	7　(4.7)	
Caucasian	118 (80.3)	127 (84.7)	
Hispanic	12 (8.2)	13　(8.7)	
Other	4　(2.7)	3　(2.0)	
Weight (lb)			0.796
N	147	149	
Mean ± STD	178.10 ± 46.58	176.65 ± 41.42	
Median	171.0	175.0	
Range	95.5 – 330.0	107.0-331.0	
Height (in)			0.099
N	147	150	
Mean ± STD	67.11 ± 3.84	66.38 ± 3.74	
Median	66.5	66.0	
Range	60.0-77.0	59.0-76.0	
HAMD 1-17 baseline			0.529
N	147	150	
Mean ± STD	23.00 ± 2.96	22.8 ± 2.92	
Median	23.0	22.0	
Range	17.0-32.0	17.0-33.0	
HAMD 1-28 baseline			0.506
N	147	150	
Mean ± STD	30.01 ± 4.89	29.62 ± 4.77	
Median	29.0	29.0	
Range	21.0-44.0	21.0-46.0	
MDRS baseline			0.558
N	147	150	
Mean ± STD	27.26 ± 5.78	26.96 ± 5.65	
Median	28.0	27	
Range	13.0-42.0	15.0-44.0	
CGI baseline			0.070
Mildly ill	1　(0.7)	1　(0.7)	
Moderately ill	104 (70.7)	119(79.3)	
Markedly ill	37　(25.2)	26(17.3)	
Severely ill	5　(3.4)	4 (2.7)	

In the sponsor's primary efficacy analysis, there was no treatment significant difference found based on the protocol specified ITT population at week 8 LOCF endpoint (p=0.117; mean change from baseline in $HAMD_{1-17}$ scores were –6.64 and –7.81 for STS and placebo groups, respectively). All the treatment comparisons were found to be in favor of placebo during the course of the study based on the mean change from baseline in $HAMD_{1-17}$ scores.

There was not any trend in favor of STS based on the secondary efficacy variables ($HAMD_{1-28}$, MADRS, CGI-s, CGI-c) at the LOCF, week 8 endpoint using the protocol specified ITT population.

Table 5.A.3   **Efficacy Analysis (protocol defined ITT-LOCF) : Mean Scores and Mean Changes From Baseline to Endpoint**

	Selegiline (n=137)	Placebo (n=146)	p-value #
$HAM-D_{1-17}$			
N	137	146	0.117
Mean ± STD	16.31 ± 7.65	14.93 ± 7.17	
Chg Mean ± STD	-6.64 ± 6.88	-7.81 ± 6.98	
$HAM-D_{1-28}$			
N	137	146	0.215
Mean ± STD	21.00 ± 10.16	19.43 ± 9.21	
Chg Mean ± STD	-9.02 ± 8.88	-10.16 ± 9.20	
MADRS			
N	137	146	0.188
Mean ± STD	20.04 ± 10.79	18.52 ± 9.72	
Chg Mean ± STD	-7.22 ± 9.41	-8.43 ± 9.54	
CGI-s (severity of illness)			
Normal, not ill	10 (7.3)	14 (9.6)	0.911
Borderline ill	25 (18.2)	21 (14.4)	
Mildly ill	23 (16.8)	26 (17.8)	
Moderately ill	59 (43.1)	72 (49.3)	
Markedly ill	17 (12.4)	12 (8.2)	
Severely ill	3 (2.2)	1 (0.7)	
CGI-c (change in severity of illness)			
Very much improved	20 (14.6)	19 (13.0)	0.494
Much improved	24 (17.5)	29 (19.9)	
Minimally improved	38 (27.7)	42 (28.8)	
Unchanged	41 (29.9)	52 (35.6)	
Minimally worse	9 (6.6)	4 (2.7)	
Much worse	5 (3.6)	0 (0.0)	

Note : # : The p-values for the continuous variables were based on the ANCOVA model = treatment+center+baseline score.

5.7　Reviewer's Evaluation and Comments

This reviewer confirmed the sponsor's finding of no difference between treatment groups (p=0.1275, -6.67 and -7.81 for the STS and placebo group, respectively, based on the mean change from baseline in $HAMD_{1-17}$ scores). The placebo group was found to have more improvement based on change from baseline in $HAMD_{1-17}$ score.

Similar to the result of study S9303-E114-98B, this reviewer found that this study also had higher placebo response.

Since there were not any significant results found in the sponsor's primary and secondary efficacy endpoints, no further evaluation was performed.

6. Summary

This reviewer had confirmed that among four pivotal studies, study E106 was positive; studies E114 and E113 were failed studies, while the reviewer can not confirm a positive result for study P9804 (Table 6). The analytic issues related to study P9804 was that the sponsor's results were not based on the protocol specified ITT population and the ANCOVA model adjusted for age was not pre-specified. By using the protocol specified ITT population and fitting model without adjusting for age, the sponsor did not achieve statistical significance based on the primary endpoint analysis (treatment difference in mean change from baseline of $HAMD_{1-17}$). Therefore, this reviewer concluded that the sponsor did not provide sufficient evidence to prove the efficacy of Selegiline.

Table 6.　　Summary of the mean change from baseline of $HAMD_{1-17}$ to endpoint, based on ANCOVA model (treatment+center+baseline covariate) using protocol specified ITT population

Study	Treatment	Mean change from baseline of $HAMD_{1-17}$	p-value
S9303-E106-96B N=176	STS 20 mg Placebo	-8.74 ± 7.53 -6.16 ± 6.68	0.0147
S9303-P9804 N=289	STS 20 mg Placebo	-8.08 ± 7.27 -6.71 ± 6.82	0.0884
S9303-E114-98B N=435	STS 10 mg STS 20 mg Placebo	-9.02 ± 7.51 -9.18 ± 7.39 -8.06 ± 7.78	Overall :0.312
S9303-E113-98B N=283	STS 20 mg Placebo	-6.67 ± 6.91 -7.81 ± 6.98	0.1275

Yuan-Li Shen, Dr. PH
                                        Mathematical Statistician

Concur :

<u>Dr. Jin</u>   <u>Dr. Chi</u>

CC:
NDA: 21-336
HFD-120/Dr. Katz
HFD-120/Dr. Laughren
HFD-120/Dr. Dubitsky
HFD-120/Dr. Bates
HFD-710/Dr. Chi
HFD-710/Dr. Jin
HFD-710/Dr. Shen
Note : This document was saved in c:\nda2001\selegiline_trns dermal_depression\
       selegiline_trns dermal_depression.stat.doc

Appendix I   Reason of the discrepancy in the primary efficacy results for study S9303-P9804

The results from the reviewer's analysis with baseline values carried over and based on ANCOVA model adjusted for age were a little different from the sponsor's results. The reason is that the treatment effect over time was analyzed by the actual length of time the patient received drug (not by visit number; see the sponsor's fax response with regard to this reviewer's request of clarification on by-visit and by-week analysis dated October 31,2001). The sponsor indicated that visit number may not correspond to actual time on drug and the actual time span is more accurate to reflect the time on drug. Based on the analysis based on actual week, the records at the actual week 8 did not match with visit 8 records.

For example, the Patient 0212's last HAM-D_{1-17} measure (i.e. at Visit 8) was not used in the sponsor's analysis since the actual week 8 measure was not the same as the visit 8 measure. Instead, the actual week 8 HAM-D_{1-17} value (rather than the Visit 8 measure) for patient 0212 was used in the analysis. The actual week 8 HAM-D_{1-17} value was from the second-to-the-last measure during the treatment period for patient 0212. However, this reviewer used the last endpoint measure for the analysis since there were no rule set in the protocol with regard to endpoint time window. So, this reviewer obtained p-value of 0.0519 versus sponsor's p-value of 0.046, based on the model including age in the model and using the sponsor's ITT population. Note that this reviewer obtained the means of the change from baseline HAM-D_{1-17} score equal to −7.87 (standard deviation=7.29) for STS group and −6.36 (standard deviation=6.81) for placebo group. The mean and the standard deviation for the STS group obtained from this reviewer was the same as the sponsor's result, but the results were slightly different for placebo group due to the discrepancy of the patient 0212 records.

/s/
- - - - - - - - - - - - - - - - - - - -
Yuan-li Shen
1/18/02 03:01:40 PM
BIOMETRICS

Kun Jin
1/25/02 11:03:34 AM
BIOMETRICS

George Chi
1/25/02 11:10:39 AM
BIOMETRICS