

DIVISIONAL COMMENTS ON ORIGINAL PROTOCOL

Those in this Division who reviewed the protocol had the following comments

- The design of the study was not adequate for purposes of demonstrating an effect of the drug on preventing or slowing the progression of Alzheimer's disease, i.e., distinguishing such an effect on the pathogenesis of the disease from a symptomatic one. Dr Paul Leber, who was Division Director at that time, indicated that this Division would be happy to discuss maneuvers that might be applied to distinguish symptomatic from structural effects of the drug
- The sensitivity and specificity of the screening procedure for the detection of dementia needed to be specified; such a test would need to sacrifice specificity for sensitivity
- Patients with dementia would need to be systematically excluded from the study prior to enrollment; the assessment of whether dementia was present at study entry should not be made retrospectively
- Once the screening procedure showed the possibility of dementia, the entire Dementia Diagnostic Procedure would need to be initiated without waiting for Week 24
- The criteria for diagnosing different types of dementia should rely on generally-accepted definitions and criteria
- An analysis should be performed of the potential effects of medication on cognition
- Since the planned analysis involved making comparisons of the prevalence of several different kinds of dementia between the raloxifene and placebo groups, adjustments for multiple comparisons would have to be made.

AMENDMENTS TO ORIGINAL PROTOCOL

No amendments have apparently been made to the protocol, since it was initially reviewed by me on 1/27/98: the submission that was reviewed by me at that time did itself contain a protocol amendment that described the Dementia Diagnostic Evaluation for the first time.

NOTE

The MAPS Battery, according to the sponsor is a 7-item validated set of tests assessing 7 major areas of cognitive function including Attention-Vigilance, Fragmented Picture Identification, Prompted Recall, Fluencies, Delayed Recall With Confidence Ratings, Delayed Recognition And Primed Fragmented Pictures Recognition. The Buschke Selective Reminding Test was apparently added to this battery. The sponsor has stated that "a sample size of 40 to 60 patients per treatment arm is statistically sufficient to detect clinically relevant treatment effects over time". This battery will be performed at only 2 sites.

The Affective Rating Scale according to the sponsor, been determined to be reliable, valid and sensitive to drug effects. The scale is intended to be a measure of mood, is self-administered, and comprises 30 questions all of which require a "yes" or "no" response. The range of scores is 0-30 with a higher score believed to indicate increased depression.

RESULTS

APPEARS THIS WAY
ON ORIGINAL

Methods of Analysis

- No patients were diagnosed to have dementia during the course of the study until the Month 36 visit; thus the protocol-specified primary (prevalence of dementia due to Alzheimer's disease) and secondary (prevalence of dementia due to cerebrovascular disease and dementia due to all causes) outcome measures and plan of analysis could not be utilized.
- An ad-hoc plan of analysis thus appears to have been substituted.
- For the Cognitive and Neuropsychomotor Test Battery checked at baseline and Months 6, 12, 24, and 36
 - The analyses of actual and percentage change in scores for each test used "standard analyses" for continuous data
 - Statistical inferences were based on unranked data
- The number of patients who underwent testing with one or more components of the Cognitive and Neuropsychomotor Test Battery is as follows:
 - The Short Blessed Test, Word List Memory, Word Fluency, Trailmaking A, Trailmaking B, Word List Recall, Static Balance and Gait were performed at all except 2 US centers, involving about 6900 patients
 - The Word List Memory and Word List Recall tests were conducted in all countries where English, French or Spanish is the main language, involving about 440 patients
- The Affective Rating Scale was performed in all countries where English is the principal language, involving about 3500 patients.
- For the MAPS Battery checked, at selected centers only at baseline and Months 6, 12, 24, and 36 the following were the methods of analysis used: it is unclear from the submission.
 - For each test the analysis of actual score adjusted for baseline was based on ANCOVA
 - For each test the analysis of change score from baseline to endpoint was based on ANOVA
 - Statistical inferences were based on unranked data
- The number of actual and near falls for each patient between visits was recorded at baseline and at Months 3, 6, 12, 18, 24, 30 and 36 months. Definitions are provided for "actual" and "near" falls. The proportions of patients with greater than 1, 2, 3, 4, 5, or 6 actual and near falls, at baseline and endpoint, were compared using chi-square
- Treatment-by-substudy interaction analyses were performed on change from baseline to endpoint for each test in the Cognitive and Neuropsychomotor Battery using ANOVA
- Treatment-by-subgroup analyses were performed for baseline score, age and years of education. Each of these 3 variables was categorized into upper, middle and lower tertiles. For each of these variables, ANOVA on unranked data was performed to assess for a treatment by baseline score-, age- and years of education interaction.
- No adjustment of α was made to account for multiple comparisons
- The incidence of treatment-related adverse events related to cognition, affect and anxiety, was used to assess the cognitive and psychobehavioral safety of raloxifene. Frequency tables were prepared for these adverse events using COSTART terms. These were analyzed using chi-square or Fisher's exact test as appropriate.

Efficacy Outcome

General

- The number of patients enrolled in each group was as follows:
Placebo: 2576 patients
Raloxifene 60 mg daily: 2557 patients
Raloxifene 120 mg daily: 2572 patients
- The number and percentage of patients in each treatment group who completed the initial 36 months of the study was as follows
Placebo: 1924 patients (75 %)
Raloxifene 60 mg daily: 1972 patients (77 %)
Raloxifene 120 mg daily: 2005 patients (78 %)
- Baseline demographic data for a few variables are indicated in the table below. The differences between groups were not statistically significant ($p < 0.05$)

Treatment Group	Mean Age (years)	% Caucasian	Mean Weight (kg)	Mean Years of Education
Raloxifene 60 mg	66.48	96.0	63.50	11.78
Raloxifene 120 mg	66.31	95.3	63.96	11.90
Placebo	66.60	95.7	63.64	11.82

- As noted earlier no patients developed dementia during the 36 months of the study, despite the estimate (in the protocol), based on the medical literature, that about 4.5 % of patients in the whole cohort would develop Alzheimer's disease over the first 3 years of the study

Cognitive and Neuropsychomotor Test Battery

Combined Substudies

- There were no statistically significant differences in mean test results between treatment groups at baseline except that a pairwise comparison of muscle strength between the raloxifene 120 mg and raloxifene 60 mg groups at baseline showed a trend towards statistical significance ($p = 0.057$)
- Except for the following, there were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint
On the Trailmaking A time to test completion, a statistically significant greater mean decrease and mean percentage decrease from baseline to endpoint was seen for the raloxifene 60 mg group vs placebo ($p < 0.05$). This result favored the raloxifene group

Substudy I

- There were no statistically significant differences in mean test results between treatment groups at baseline except that a pairwise comparison of word list fluency between the raloxifene 60 mg and placebo groups at baseline showed a trend towards statistical significance ($p = 0.056$)
- Except for the following, there were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint
When Short Blessed Test time was measured, the raloxifene 60 mg group showed a statistically significant greater percentage increase from baseline to endpoint in comparison with placebo ($p < 0.05$); this change favored placebo

Substudy 2

- There were no statistically significant differences in mean test results between treatment groups at baseline
- There were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint

Treatment-by-Substudy Interaction

A statistically significant interaction between treatment and substudy was seen for the Word List Fluency score (change from baseline to endpoint) for the pooled raloxifene group compared with placebo ($p < 0.05$)

Treatment-by-Subgroup Interaction

- For baseline scores there was no evidence of an interaction for any test
- For age, there was no evidence of an interaction for any test
- For years of education, there was no evidence of an interaction effect overall among the 3 treatment groups for any test. However there was a statistically significant interaction effect when comparing the pooled raloxifene groups with placebo for the following tests
 - For the Short Blessed Test time interaction effect ($p < 0.05$), the mean decrease in time for the pooled raloxifen group compared with placebo was largest in the lowest tertile (≤ 10 years of education)
 - For the Trailmaking A time interaction effect ($p < 0.05$), the mean decrease in time for the pooled raloxifen group compared with placebo was largest in the lowest tertile (≤ 10 years of education)

Affective Rating Scale

Combined Substudies

- There were no statistically significant differences in mean test results between treatment groups at baseline
- There were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint

Substudy 1

- There were no statistically significant differences in mean test results between treatment groups at baseline
- Except for the following, there were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint

The mean percentage change from baseline comparing the raloxifene 120 mg group with placebo favored raloxifene with a trend toward statistical significance ($p = 0.058$)

Substudy 2

- There were no statistically significant differences in mean test results between treatment groups at baseline
- There were no statistically significant differences between treatment groups in mean change or mean percentage change from baseline to endpoint

MAPS Battery

- The number of patients who underwent testing with this battery is not stated; as indicated earlier, only 2 US centers used this battery
- There were no statistically significant differences in mean test results between treatment groups at baseline except for the following 2 items
 - For the Fragmented Picture test, mean absolute savings at baseline was significantly ($p < 0.05$) greater for the raloxifene 120 mg group versus the placebo group in a pairwise comparison (ANOVA; unranked data)
 - For the Buschke Test mean number correct at baseline was significantly ($p < 0.05$) higher for the raloxifene 120 mg group as compared with the raloxifene 60 mg group (ANOVA; ranked data)
- Except for the following, there were no statistically significant differences between treatment groups in mean at endpoint or mean change from baseline to endpoint
 - The mean change from baseline to endpoint, comparing the raloxifene 60 mg dose group to the placebo group (ANOVA; unranked data) on the absolute savings component of the Fragmented Pictures test favored placebo ($p < 0.05$)
 - The mean change from baseline to endpoint comparing the raloxifene 120 mg group with placebo (ANOVA; unranked data) on the consistency component of the Buschke test favored placebo ($p < 0.05$)
- No treatment-by-substudy interaction analysis was performed for this battery since patient numbers were limited (the precise number of patients who underwent testing with this battery has not been stated)

Falls and Near Falls

Combined and Individual Substudies

- There were no statistically significant differences in mean test results between treatment groups at baseline (overall comparisons)
- From baseline to endpoint there were no statistically significant differences in mean test results between treatment groups (overall comparisons)

Safety Outcome

- The incidence of cognition-, affect- and anxiety-related adverse events is summarized in the following table for all randomized patients

APPEARS THIS WAY
ON ORIGINAL

TESS GROUP	TESS EVENT	Placebo(1) N=2576		RLX060(2) N=2557		RLX120(3) N=2572	
		n	(%)	n	(%)	n	(%)
Cognition-Related Events	ANY EVENT	73	2.834	81	3.168	87	3.383
	AMNESIA	42	1.630	49	1.916	59	2.294
	CONFUSION	15	0.582	17	0.665	18	0.700
	DEMENTIA	7	0.272	6	0.235	7	0.272
	THINKING ABNORMAL	6	0.233	12	0.469	9	0.350
	STUPOR	6	0.233	3	0.117	3	0.117
	DELIRIUM	2	0.078	1	0.039	0	0.000
	MENTAL RETARDATION	0	0.000	0	0.000	2	0.078
Affect-Related Events	ANY EVENT	185	7.182	175	6.844	191	7.426
	DEPRESSION	158	6.134	148	5.788	166	6.454
	SOMNOLENCE	26	1.009	29	1.134	23	0.894
	INTENTIONAL INJURY	1	0.039	5	0.196	3	0.117
	PSYCHOTIC DEPRESSION	2	0.078	0	0.000	1	0.039
Anxiety-Related Events	ANY EVENT	175	6.793	176	6.883	191	7.426
	ANXIETY	144	5.590	139	5.436	159	6.182
	NERVOUSNESS	21	0.815	22	0.860	23	0.894
	EMOTIONAL LABILITY	6	0.233	14	0.548	6	0.233
	AGITATION	6	0.233	6	0.235	11	0.428
	HOSTILITY	0	0.000	1	0.039	2	0.078
	AGGRAVATION REACTION	0	0.000	0	0.000	1	0.039

N: Total number of patients in this group
 n: Number of patients with an event

Note that RLX060 and RLX120 stand for raloxifene 60 mg and raloxifene 120 mg, respectively. The placebo, raloxifene 60 mg and raloxifene 120 mg groups have been designated as Groups 1, 2 and 3, respectively

- No p-value for the overall, Group 1 vs Group 3, Group 2 vs Group 3 or pooled comparisons was at a statistically significant level ($\alpha = 0.05$).

SPONSOR'S CONCLUSIONS

- While there were a few statistically significant findings, there were no consistent effects trends favoring either raloxifene or placebo, and no consistency in the results of tests assessing similar domains of cognitive function.
- Based on adverse event reporting, raloxifene does not adversely affect cognitive function or mood and is not associated with an increase in anxiety-related adverse events.
- Raloxifene does not appear to have any deleterious effect on cognition or neuropsychiatric function after 36 months of treatment.

COMMENTS

- The protocol-specified cognitive primary outcome measure was the prevalence of Alzheimer's disease. The secondary outcome measures

BEST POSSIBLE COPY

specified in the protocol were the prevalence of dementia due to cerebrovascular disease and the prevalence of dementia due to all causes. The protocol-specified analysis plan was based on these outcome measures. However, no patient developed dementia during the course of the study and, thus, the protocol-designated analysis plan was not put into effect

- Instead, the Cognitive and Neuropsychomotor Test Battery, the Affective Rating Scale, the MAPS Battery, and the incidence of falls and near falls have been subjected to an ad hoc (and post hoc) method of analysis, not specified in the protocol.
- Patients with pre-existing dementia were not excluded from the study.
- No adjustment of α was made to account for multiple comparisons
- The relevance of the few statistically significant ($p < 0.05$) comparisons noted among the many (comparisons) made is questionable, as the sponsor has also suggested; the lack of a consistent trend favoring either raloxifene or placebo also raises questions about the true meaning of these results
- It is unclear if the sample size for this study is adequate to demonstrate a statistically significant difference (at even a $p < 0.05$ level) between treatment groups for the many comparisons for which a statistically significant difference was not demonstrated; an even larger sample might have been needed if α was adjusted downwards to account for multiple comparisons. **Thus the study may be lacking in power to support the sponsor's conclusions regarding the effect of the above doses of raloxifene on cognition and affect.**
- Treatment-emergent adverse events related to cognition, affect and anxiety were only slightly more frequent in the raloxifene groups than in those treated with placebo; these p-values for overall, pooled and individual-raloxifene-group-versus-placebo comparisons were not statistically significant at an α of 0.05.

APPEARS THIS WAY
ON ORIGINAL

4. Review of Study # H3S-MC-GGGN

Note that:

- This study is ongoing; the submitted study report contains the 24-month data.
- All cognitive function data were analyzed at Memory Assessment Clinics, Gaithersburg, Maryland. In this context, "cognitive function" apparently includes true cognitive function, as well as affect. 2 separate study sub-reports from Memory Assessment Clinics describe the results, and conclusions drawn, from these data

OUTLINE OF ORIGINAL PROTOCOL

The items in the protocol pertinent to this review are very limited. Hence, only a summary of the protocol will be provided below.

Objectives

The primary objective of the study was to assess the efficacy and safety of raloxifene versus placebo in the treatment of post-menopausal osteoporosis. In the assessment of efficacy, measures of generalized and lumbar spine-proximal femur bone mineral density were to be used. Secondary objectives included the assessing the effect of raloxifene in comparison with placebo on radial bone mineral density, biochemical markers of bone metabolism, fracture rates, serum lipids, endometrial thickness and other uterine changes, and cognitive function

Design

Randomized, double-blind, placebo-controlled, parallel-arm, 3-group study.

The 3 treatment groups were to consist of:

Raloxifene 120 mg daily
Raloxifene 60 mg daily
Placebo

The study was to consist of 4 phases:

1. A screening/washout phase lasting 2 days to 6 months for purposes of discontinuing concurrent estrogen, calcitonin, calcium and Vitamin D therapy; concurrent estrogen therapy would mandate a 6 month washout
2. The main double-blind treatment phase, lasting 1 year, at the start of which patients would be randomized to one of the above 3 treatment groups
3. An optional double-blind extension phase, again lasting 1 year, during which patients who received raloxifene in the immediately preceding phase would continue to receive the same dose, whereas those who received placebo would be randomly assigned to either of the raloxifene arms
4. An indefinite open-label extension phase during which all patients would receive raloxifene 60 mg daily

Sample Size

A total of at least 138 patients were to be randomized to the 3 treatment groups.

Selection Criteria

Post-menopausal women, aged 45 to 70 years, with osteoporosis (defined by specific clinical and laboratory criteria) were to be included.

Patients with certain bone disorders other than osteoporosis, potential contraindications to estrogen therapy, post-menopausal symptoms warranting conventional estrogen therapy, alcohol or drug abuse and a variety of specified renal, hepatic, endocrine, and gastrointestinal abnormalities or diagnoses were to be excluded. Also to be excluded were patients taking Vitamin D, anticonvulsants, phosphate-binding acids, sodium fluoride, biphosphonates, estrogens and other specified hormonal agents and investigational drugs (washout periods were specified for several of these agents). Neither the inclusion nor exclusion criteria listed any aspect of cognitive functioning; however patients considered to be poor medical or psychiatric risks were to be excluded from the study

Schedule for Cognitive and Affective Assessments

The MAC battery (see below) and the Affective Rating Scale were scheduled to be checked at baseline and at Months 1, 6, 12, 18 and 24.

Cognitive and Affective Outcome Measures

According to the original protocol, the following outcome measures were to be used:

- Computerized Memory Assessment Clinics (MAC) psychometric battery
- Affective Rating Scale

However, in the study report, a third outcome measure has been added to this list. This consists of:

- Walter Reed Performance Assessment Battery

Further details of these outcome measures are provided below

The MAC battery used in this study is reported to measure statistically-independent aspects of memory and learning. According to the study report, these measures have been previously shown to be reliable and valid and have also been used in numerous drug studies. The reportedly independent aspects of learning and memory that were assessed and the variables measured in each instance are listed in the following table.

Learning/Memory Function	Variables Measured
Name-Face Association	Initial Learning Total Acquisition Delayed Recall
First-Last Names Association (Associative Verbal Learning and Memory)	Initial Learning Total Acquisition Delayed Recall
Facial Recognition (Delayed Non-Matching to Sample Paradigm)	Number Before First Error Total Correct
Numeric Recall (Telephone Dialing)	Seven Digit Recall Before Interference Ten Digit Recall Before Interference Seven Digit Recall After Interference Ten Digit Recall After Interference

According to the study report, the Total Acquisition Score on the Name-Face Association test was designated as the primary outcome measure prior to the study. However a primary cognitive outcome measure is not explicitly stipulated in the protocol

The Walter Reed Performance Assessment Battery is a computerized test battery which provides measures of attention, information processing speed and efficiency, and reaction time. The reliability and validity of the measures used in this battery are unclear and do not appear to have been reported in the citations provided by the sponsor. The functions that are assessed with this battery and the variables derived from them are in the table below. **As noted earlier this battery was not designated as an outcome measure in the protocol, but has been described in detail in the study report**

Function	Variables Measured
Two-Letter Search	Effective Speed Throughput
Six-Letter Search	Effective Speed Throughput
Four-Choice Serial Reaction Time	Effective Speed Throughput

APPEARS THIS WAY
ON ORIGINAL

The 2 batteries described above together measured a total of 18 cognitive variables.

The Affective Rating Scale of Yesavage was designated as a secondary outcome measure in the original protocol. This scale has according, to the sponsor, been determined to be reliable, valid and sensitive to drug effects. The scale is intended to be a measure of mood, is self-administered, and comprises 30 questions all of which require a "yes" or "no" response. The range of scores is 0-30 with a higher score believed to indicate increased depression.

Analysis Plan for Cognitive and Affective Outcome Measures

- The original protocol does not specify a plan of analysis for the cognitive and affective outcome measures listed
- The sample size estimate is based solely on measures of bone mineral density

AMENDMENTS TO ORIGINAL PROTOCOL

There is no indication in this submission that there were any formal amendments to the protocol.

RESULTS

Outline of Methods of Analysis

(as stated in the study report only)

Initial Double-Blind Phase (Months 0 through 12)

- This analysis was carried out on the 143 patients initially randomized to the study
- 2 cohorts were subjected to the analysis, which was conducted in parallel.
 - The first cohort comprised the Completer Analysis consisting of all patients who completed the initial double-blind phase
 - The second cohort comprised the Intention-To-Treat Analysis consisting of all those who were randomized to the double-blind phase; the last-observation-carried-forward imputation scheme was used for those who did not complete the study through Month 12.

APPEARS THIS WAY
ON ORIGINAL

- In both cohorts, overall ANCOVA (see details of model below) comparisons were made comparing all 3 treatment groups on each variable, both cognitive and affective
- The groups compared have been as follows: raloxifene 60 mg, raloxifene 120 mg and placebo. The 3 groups were compared together.
- These groups were compared at the following post-baseline timepoints: Months 1, 6 and 12; the primary emphasis in assessing efficacy was made on comparisons at Month 12
- A total of 57 comparisons have been made of the 3 groups together.
- In addition separate comparisons were made between each of the raloxifene groups and placebo
- Comparisons were made between the groups using ANCOVA: in the model used, age and baseline score were the covariates, and treatment the only main effect.
- No adjustment of α was made to account for multiple comparisons
- Prior to conducting efficacy analyses, data were examined to determine if patients from the 3 treatment groups and 2 study sites were comparable with respect to age.

Double-Blind Extension Phase (Months 12 through 24)

- The analysis was confined to the 121 patients who entered the optional double-blind extension phase
- 2 separate groups were analyzed
 - The 2 Year Group consisted of those who had received raloxifene, in a dose of 60 mg or 120 mg, during the initial 12 month mandatory double-blind phase and continued on the same dose of medication during the double-blind extension phase
 - The Placebo Crossover Group consisted of those who had received placebo in the initial 12 month mandatory double-blind phase, and were then randomized to receive raloxifene in a dose of either 60 mg or 120 mg during the double-blind extension phase
- 2 cohorts were subjected to the analysis, which was conducted in parallel, in each of the 2-Year Group and the Placebo Crossover Groups
 - The first cohort comprised the Completer Analysis consisting of all patients who completed the double-blind extension phase
 - The second cohort comprised the Intention-To-Treat Analysis consisting of all those who entered the double-blind extension phase; the last-observation-carried-forward imputation scheme was used for those who did not complete the study. The authors of the study report state that data from all cognitive and affective assessments made during the initial double-blind phase were also used for analysis of this cohort but were used only in the 2-Year Group
- All comparisons were made between the 60 mg and 120 mg (of raloxifene) groups
- For the 2-Year Group comparisons were made at 5 timepoints during the first 24 months of the study: Months 1, 6, 12, 18 and 24
- For the Placebo Crossover Group comparisons were made at Months 18 and 24
- Comparisons were made between the 2 raloxifene groups using ANCOVA: in the model used, age and baseline score were the covariates, and treatment the only main effect.

- No adjustment of α has been made to account for multiple comparisons

Both Phases

The incidence of treatment-related adverse events, deaths, serious adverse events and adverse events leading to study discontinuation was compared between treatment groups

APPEARS THIS WAY
ON ORIGINAL

Efficacy Outcome

Initial Double-Blind Phase (Months 0 through 12)

- 143 patients were randomized to the treatment groups and 125 patients completed the initial 12 months of the study; their distribution among the treatment groups is outlined in the following table

Group	Raloxifene 60 mg	Raloxifene 120 mg	Placebo	Total
Randomized	48	47	48	143
Completed	42	40	43	125

- The mean age at baseline in each treatment group in the intention-to-treat population summarized below; the differences are statistically significant ($p = 0.03$)

Treatment Group	Mean Age
Raloxifene 60 mg	69.4 years
Raloxifene 120 mg	66.68 years
Placebo	67.73 years

APPEARS THIS WAY
ON ORIGINAL

- Comparisons that were considered statistically significant ($p < 0.05$) are illustrated in the table below

Variable	Comparison	Cohort	Timepoint	P-value**	Group Favored*
Name-Face Association Initial Learning	3 groups	Completer	Month 6	< 0.05	Both drug groups
Numeric Recall-Seven Digit Recall Before Interference	3 groups	Completer	Month 1	0.01	Both drug groups
Name-Face Association Initial Learning	3 groups	Intention-to-treat	Month 6	< 0.05	Raloxifene 120 mg
First-Last Names Association Delayed Recall	3 groups	Intention-to-treat	Month 1	< 0.05	Raloxifene 120 mg
Name-Face Association Initial Learning	Raloxifene 60 mg vs placebo	Completer	Month 6	< 0.05	Raloxifene 60 mg
Name-Face Association Initial Learning	Raloxifene 60 mg vs placebo	Intention-to-treat	Month 6	< 0.05	Raloxifene 60 mg
Numeric Recall-Seven Digit Recall Before Interference	Raloxifene 60 mg vs placebo	Intention-to-treat	Month 1	< 0.05	Raloxifene 60 mg
Name-Face Association Initial Learning	Raloxifene 120 mg vs placebo	Completer	Month 6 Month 12	< 0.05	Raloxifene 120 mg
Numeric Recall-Seven Digit Recall Before Interference	Raloxifene 120 mg vs placebo	Completer	Month 1	< 0.05	Raloxifene 120 mg
Numeric Recall-Ten Digit Recall After Interference	Raloxifene 120 mg vs placebo	Completer	Month 6	< 0.05	Placebo
Numeric Recall-Seven Digit Recall Before Interference	Raloxifene 120 mg vs placebo	Intention-to-treat	Month 1	0.01	Raloxifene 120 mg

* i.e., having a better performance
 ** all p-values designated as < 0.05 were > 0.01

- No "significant" differences other than the above were noted; no differences were noted at any Month 12 comparison among all 3 treatment groups

Double-Blind Extension Phase (Months 12 through 24)

- 121 patients were enrolled in the study; 81 patients were in the 2 Year Treatment Group and 40 patients were in the placebo cross-over group
- In the 2 Year Treatment Group 74 patients completed the 12 months of the double-blind extension phase of study: their distribution among the treatment groups is outlined in the following table

Group	Raloxifene 60 mg	Raloxifene 120 mg	Total
Enrolled	43	38	81
Completed	39	35	74

- In the 2 Year Treatment Group the mean age at enrolment in each treatment group in the intention-to-treat population summarized below; the differences are statistically significant ($p = 0.03$)

Treatment Group	Mean Age
Raloxifene 60 mg	69.49 years
Raloxifene 120 mg	67.11 years

- In the Placebo Crossover Group all 40 patients completed the 12 months of the double-blind extension phase of study; their distribution among the treatment groups is outlined in the following table

Group	Raloxifene 60 mg	Raloxifene 120 mg	Total
Enrolled	18	22	40

- In the Placebo Crossover Group the mean age at enrolment in each treatment group is summarized below; the difference was not statistically significant

Treatment Group	Mean Age
Raloxifene 60 mg	68.22
Raloxifene 120 mg	67.64

- Comparisons that were considered statistically significant ($p < 0.05$) are illustrated in the table below; note that in 4/5 instances the statistically significant comparisons were during the initial double-blind phase

Variable	Group	Cohort	Timepoint	P-value**	Treatment Group Favored*
First-Last Names Association Delayed Recall	2 Year	Completer	Month 1	0.05	Raloxifene 120 mg
First-Last Names Association Delayed Recall	2 Year	Intention-to-treat	Month 1	< 0.05	Raloxifene 120 mg
Facial Recognition-Total Correct	2 Year	Intention-to-treat	Month 1	< 0.05	Raloxifene 120 mg
Numeric Recall- Ten Digit Recall After Interference	2 Year	Intention-to-treat	Month 6	< 0.05	Raloxifene 60 mg
Numeric Recall- Ten Digit Recall After Interference	Placebo Crossover	Completer	Month 24	< 0.05	Raloxifene 60 mg

* i.e., having a better performance

** all p-values designated as < 0.05 were > 0.01