CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:
22-083

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 22,083

Drug Name: Exelon Transdermal Patch

Indication(s): Mild to Moderate Alzheimer's

Applicant: Novartis

Date(s): Submission Date: Sep 8, 2006

Review Priority: Standard

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Tristan Massie, Ph.D.

Concurring Reviewers: Kun Jin, Ph.D., Team Leader

Kooros Mahjoob, Ph.D., Deputy Director, Division of Biometrics 1

Medical Division: Neurologic Drugs

Clinical Team: Ranjit Mani, M.D.

Russ Katz, M.D., Division Director

Project Manager: Melina Griffis

Keywords: Multiplicity; Conditional Testing

Table of Contents

Appears This Way
On Original

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The high dose patch group reached statistical significance compared to the placebo group on the change from baseline at week 24 in the co-primary Alzheimer's Disease Assessment Scale - Cognitive Subscale Total Score (ADAS-cog) but it did not quite reach significance on the co-primary Clinical Global Impression of Change (CGI). The decision rule was set up to require significance on the high dose patch vs. placebo comparison on both co-primary endpoints before testing the low dose patch vs. placebo. Therefore, technically the study failed. However, some other possible multiplicity adjustment methods, such as the Bonferroni method, had they been chosen would have yielded a positive study by means of significance of the low dose patch vs. placebo comparison on both the ADAS-cog and the CGI, even at the 0.025 level. Because this is a new formulation of a drug that is approved for this indication in it's original capsule formulation the evidence for the original oral formulation might also lend support to the efficacy of the new patch formulation in this close case.

1.2 Brief Overview of Clinical Studies

A 24-week, multicenter, randomized, double-blind, placebo- and active-controlled, parallel group evaluation of the efficacy, safety, and tolerability of the once-daily Exelon® patch formulation in patients with probable Alzheimer's Disease (AD) [Mini-Mental State Examination [MMSE] 10-20].

One adequate and well-controlled study (placebo and active-controlled study 2320) conducted in 1195 patients is used to support the efficacy claims in the target indication of mild to moderately severe AD. This study had a double-blind (DB), placebo- and active-controlled treatment phase of 24 weeks and was followed by an open-label (OL) extension study 2320E1 allowing a further 28 weeks of treatment. A total of 100 centers in 21 countries, including the U.S., randomized patients.

1.3 Statistical Issues and Findings

The sponsor chose a multiplicity adjustment method which required significance of the high dose patch vs. placebo on both primary endpoints before proceeding to the low dose patch vs. placebo comparison or to any secondary endpoints. Since the high dose vs. placebo comparison reached significance on the co-primary ADAS-cog but did not reach statistical significance at 0.05 on the co-primary CGI, technically, the study could be viewed as a failure. The sponsor reported a p-value for the high dose vs. placebo comparison on the CGI of 0.054 for the primary analysis. They reported some secondary analyses of the CGI for which the comparison between the high dose patch and placebo reached the nominal level of significance. Note that the sponsor excluded CGI scores from the primary analysis which were taken more than two days after the last dose as planned in the protocol. However, this reviewer found 16 other cases (8 placebo and 8 high dose patch) where some patient's week 24 CGI was not used in the sponsor's primary ITT(LOCF) analysis. The sponsor claimed that in cases where the last CGI was not used in the LOCF analysis it was because the week 24 CGI score was taken in the open label extension period, which followed the double blind phase and, therefore, in these cases they carried forward the week 16 score in their analysis. However, in all but 2 of these 16 cases the CGI was rated only

4

one day after the end of the double blind period, assuming that the rating was in fact taken in the open label phase. Furthermore, it is not clear from the data that they were from the open label period because the CGI score data set and the dosing record data set each have a variable to indicate if the record is from the extension and they disagree in all of these cases on whether or not the week 24 CGI score was assessed in the double blind period or the open label phase. If in these few cases we use the week 24 CGI scores that were taken one day after the end of the double blind phase in the primary analysis then the p-value for the Exelon 20 cm2 patch vs. placebo comparison increases to 0.109. Therefore, the high dose vs. placebo comparison on the co-primary CGI may not be as close to significance as reported by the sponsor. Note that at the beginning of the open label extension all continuing patients were started with the Exelon 10cm2 patch but their previous double blind treatment was not revealed, i.e., they might have been able to tell the size of the patch they had been on but it was not revealed whether it had contained placebo or Exelon. It is unknown in the 14 cases where the double blind week 24 CGI assessment was reportedly taken in the open label phase, but only on the first day of it, whether or not the assessment was made before the open label drug was administered. It is also unknown whether the drug could have an effect in only 1 day for placebo patients beginning the 10 cm2 patch or for Exelon 20 cm2 patients who had tolerated the Exelon 20 cm2 patch and were switched to the Exelon 10 cm2 patch at the start of the open label phase. This requires clinical judgment. At any rate, if there was considered to be an effect of the high dose patch on the CGI it was very small and the sample size was average or above average (290 to 300 randomized per group and 250 to 280 per group had sufficient data to be included in the ITT[LOCF] analysis). Although the effect size on the CGI was also small for the low dose patch, if a Bonferroni multiplicity adjustment had been chosen the study would have been positive on the basis of the **low dose patch because it's comparison with placebo** reached significance at the 0.025 level for both co-primary endpoints.

If the highest sensitivity analysis p-value was 0.054 and there were no other problems, e.g., if the 16 double blind Week 24 CGI scores possibly taken in the open label period were considered clinically irrelevant, then one might examine whether there were other supporting elements that could lead one to conclude that 0.054 was close enough to 0.050, given the supporting elements. In this case we can consider that the investigational drug is a new formulation of a drug for **which the original formulation is approved for mild to moderate Alzheimer's. Therefore, it may** be able to borrow strength from the original oral capsule formulation. In addition, the low dose patch achieved the nominal significance level on both co-primary endpoints and would have even after a Bonferroni adjustment to the significance level. Furthermore, the p-values for both the high dose patch and the low dose patch comparisons to placebo on the other co-primary endpoint, the change from baseline in ADAS-cog Total score, were more than an order of magnitude smaller than 0.05. Also, both the high dose and the low dose patch achieved nominal significance compared to placebo on two of the secondary endpoints, **the change in Alzheimer's** Disease Cooperative Study Activities of Daily Living (ADCS-ADL) Total score and the change in MMSE. Because of these supporting factors, in the present case 0.054 may be close enough to 0.05 at least for the first gate of the gatekeeping process to be passed to permit testing of the low dose patch.

It is not clear that the high dose patch adds anything over the low dose patch. In particular, while the high dose patch was numerically better than the low dose patch in terms of the change from

5

baseline in ADAS-cog Total score by 0.8, the low dose was numerically better for the CGI and there was essentially no difference between the high dose patch and the low dose patch on the change from baseline in ADCS-ADL Total score or the change from baseline in the MMSE score. Furthermore, the low dose achieved nominal significance levels compared to placebo on both co-primary endpoints while the high dose patch group only strictly achieved significance for one of the two.

2   INTRODUCTION

2.1   Overview

Rivastigmine (also called Exelon) is a carbamate-type slowly reversible (pseudo-irreversible), brain selective, dual inhibitor of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE). Rivastigmine patch is the once-a-day modified release formulation for rivastigmine using the transdermal delivery system technology developed by Lohmann Therapie-Systeme (LTS) AG, Andernach, Germany. It uses the ————————————————————————

b(4)

———— · because of a ————————————————————————
————————————————————————————————— Five different
sizes/strengths of rivastigmine patch, i.e. 5 cm2/9 mg, 7.5 cm2/13.5 mg, 10 cm2/18 mg, 15 cm2/27 mg and 20 cm2/36 mg were tested in the clinical development program. Novartis had previously developed an oral twice-a-day formulation for rivastigmine, the Exelon® capsule, which was approved by the FDA on April 21, 2000 ([NDA No. 20-823]) for the treatment of mild to **moderate dementia of the Alzheimer's type. Currently,** two formulations (immediate release) of Exelon® exist on the market, the capsule (for daily doses of 3 to 12 mg) and the bioequivalent oral solution (2 mg/mL).

One adequate and well-controlled study (placebo and active-controlled study 2320) conducted in 1195 patients is used to support the efficacy claims in the target indication of mild to moderately severe AD. This study had a double-blind (DB), placebo- and active-controlled treatment phase of 24 weeks and was followed by an open-label (OL) extension study 2320E1 allowing a further 28 weeks of treatment.

Industry Meeting Minutes Relevant to Statistics
End of Phase II meeting held on Oct 22, 2002:
The Division is prepared to accept the results of a single appropriately-designed trial of the Exelon® Transdermal System as evidence for its efficacy.

Pre-NDA meeting held on Nov 8, 2005:
FDA had comments on the proposed statistical analysis plan related to the testing of the four hypotheses and the missing item imputation strategy: FDA asked to remove the hypothesis of the ADCS-ADL from the hierarchical testing procedure from the proposed analysis plan for study 2320, because there would be too much overlap of the ADL outcome measure with the co-primary outcome measure ADCS-CGIC. FDA indicated that they are not interested in the non-inferiority hypothesis from the hierarchical testing procedure and proposed to remove this hypothesis from the testing strategy. Novartis explained that this testing would be required for EMEA (European regulatory authorities). FDA would accept to have two different analysis plans, one for EMEA, and one for FDA. FDA proposed an analysis plan with just two

hypotheses (20cm2 vs placebo, and 10cm2 vs placebo). Alternatively, they would accept an analysis plan with 1st hypothesis 20cm2 vs placebo, 2nd hypothesis 10cm2 vs placebo, 3rd hypothesis the non-inferiority, 4th hypothesis ADL testing, however, FDA would stop after the 2nd hypothesis and not review the 3rd and 4th hypotheses. The FDA statistician asked to use mean values to impute for missing individual scale items at baseline rather than performing a carry backward strategy as described in the analysis plan. Novartis should submit the changed analysis plan as a protocol amendment.

FDA did not agree that the results of the secondary outcome measure ADCS-ADL could be included in the label, because there would be too much overlap of the ADL outcome measure with the co-primary outcome measure ADCS-CGIC. FDA also pointed out that a change in the selection of the secondary parameter would be problematic at this late stage of the trial.

2.2   Data Sources

The raw data for study 2320 is located in the following directory.
\\Cdsesub1\n22083\N_000\2006-09-08\crt\datasets\2320\raw

The derived data for study 2320 is located in the following directory.
\\Cdsesub1\n22083\N_000\2006-09-08\crt\datasets\2320\derived

The directory containing the sponsor's study report is:
\\cdsesub1\n22083\N_000\2006-09-08\clinstat\controlled

3   STATISTICAL EVALUATION

3.1   Evaluation of Efficacy

3.1.1 Study 2320

The date the first patient was screened was November 27, 2003 and the date the last patient completed the study was January 11, 2006.

3.1.1.1 Study Design

Objectives

To confirm the efficacy of the Exelon® patch in patients with probable AD (MMSE 10-20) by testing the following hypotheses:
1. Exelon® target patch size of 20 cm2 is superior to placebo on change from baseline at Week 24 simultaneously on the ADAS-Cog and the ADCS-CGIC;
2. Exelon® target patch size of 20 cm2 is non-inferior to Exelon® capsule target dose of 12 mg on the change from baseline at Week 24 on the ADAS-Cog;
3. Exelon® target patch size of 10 cm2 is superior to placebo on change from baseline at Week 24 simultaneously on the ADAS-Cog and the ADCS-CGIC;

4. Exelon® 20 cm2 target patch size is superior to placebo on change from baseline at Week 24 on the ADCS-ADL.

All patients were to undergo a preliminary evaluation (Screening Visit) to assess eligibility. At the Baseline Visit, patients whose eligibility was confirmed were to be randomized into one of 4 treatment groups (placebo, Exelon 10 cm2(18 mg) patch, Exelon 20 cm2(36 mg) patch, or Exelon 12 mg capsule) in a 1:1:1:1 ratio. Starting on the day following the Randomization Visit, all patients were to be administered one patch in the morning, which was to be worn for 24 hours. All patients were also to take one capsule with breakfast and one capsule with the evening meal. All patients took both a patch and a capsule containing Exelon or placebo depending on the treatment assignment to blind the treatment assignment. Patients would then enter a 16-week Double-blind Titration Period followed by an 8-week Maintenance Period. Patients who completed the Double-blind Treatment Phase on study drug were to have the option to enter a 28-week Open-label Treatment Phase. Patients were to be asked to return for a follow up visit at the end of every 4 week period (28 days) until Visit 6 and then to return again after 8 weeks (56 days) for the final visit. Patients should be seen for all visits as per protocol or as close to it as possible. Patients who discontinued study drug were to return for all subsequent visits at which efficacy assessments were scheduled and were to have all required efficacy assessments. These patients are referred to as retrieved dropout (RDO) patients. Primary efficacy assessments were to be obtained at Visit 2 (Week 0), 6 (Week 16) and 7 (Week 24).

Table 1 Study Design

Phase	Pre-Randomization		Double-blind Treatment				
			Exelon® Patch, Capsule or Placebo				
Period	Screening	Baseline*	Titration Period				Maintenance**
Week	Wk -4 to -1	Wk 0	1-4	5-8	9-12	13-16	17-24
Visit	V1	V2	V3	V4	V5	V6	V7 or PD
Treatment	None		Group A: Exelon® patch titrated from 5 to 10 cm² patch size				10 cm² Exelon® patch size
			Group B: Exelon® patch titrated from 5 to 10, 15, and 20 cm² patch size				20 cm² Exelon® patch size
			Group C: Exelon® capsule titrated from 3 to 6, 9 and 12 mg/d				12 mg/d Exelon® capsule
			Group D: Placebo				Placebo

* Study medication will be started on the day after the baseline visit

** The maintenance dose is defined as the target patch size for the treatment group or the highest well-tolerated dose for each individual patient.

PD: Premature discontinuation

3.1.1.2 Efficacy Measures

Primary Efficacy Assessments

The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): The Alzheimer's Disease Assessment Scale (ADAS) is a performance-based test that measures specific cognitive and behavioral dysfunctions in patients with AD (Rosen et al. 1984). The cognitive subscale of the ADAS (ADAS-Cog) comprises 11 items that are summed to a total

score ranging from 0 to 70, with lower scores indicating less severe impairment. It was to be administered by a mental health professional (e.g., M.D., Ph.D., Pharm.D., or R.N.) who had a minimum of 2 years research experience.

Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (ADCS-CGIC): The scale provides a single global rating of change from baseline (Ferris et al. 1997). The original published version has been modified to facilitate the training of the raters and increase intra- and inter-rater reliability in this international study. Raters will receive group training prior to study start to enhance reliability. The rater was not to be involved in any other way with the patient's treatment or evaluation throughout the study. The rater was not to have access to any other safety or efficacy data, including all previous postbaseline ADCS-CGIC ratings. This is a 7 point scale ranging from 1 (Marked Improvement) to 7 (Marked Worsening).

Secondary Efficacy Assessments
The ADCS-ADL is a caregiver-based ADL scale composed of 23 items developed for use in dementia clinical studies (Galasko et al. 1997). It is designed to assess the patient's performance of both basic and instrumental activities of daily living such as those necessary for personal care, communicating and interacting with other people, maintaining a household, conducting hobbies and interests, as well as making judgments and decisions.
Other secondary endpoints include the Neuropsychiatric Inventory (NPI), the Mini-Mental State Examination (MMSE) , the Ten point clock test (TPCT), and the Trail Making test (Part A) (TMT).

3.1.1.3 Statistical Methods and Sample size
The following table lists all individual patient criteria which the sponsor required to be fulfilled to qualify a patient for the corresponding analysis population.

Table 1: Criteria that qualify a patient for the different patient populations

Criteria to be fulfilled to qualify for a population	Population			
	RND	Safety	ITT	ITT+RDO
Patient was randomized	X	X	X	X
Patient did take at least one dose of study drug		X	X	X
Patient has a safety measurement after baseline		X		
Patient has a valid baseline and post-baseline efficacy assessment on treatment(i.e. not more than 2 days after last known date of study drug) for either of the primary efficacy variables (ADAS-cog or ADCS-CGIC)			X	
Patient has a valid baseline and post-baseline assessment for either of the primary efficacy variables (ADAS-cog or ADCS-CGIC)				X

Handling of missing values/censoring/discontinuations

The handling of missing data followed the schemes used for the Exelon® capsule program. Imputation of missing values was done on two levels; on the level of individual scale items and on the level of (total) scores using the following order:
1. Efficacy assessments were allocated to the appropriate analysis weeks/visits.
2. Missing scale items were imputed.
3. For scales consisting of several items, the total score was calculated at each week/visit.
4. Patients were assigned to the analysis populations.
5. The population specific imputation scheme was applied to missing (total) scores at weeks/visits.

Missing scale items

For the scales consisting of several items (ADAS-Cog, ADCS-ADL, MMSE and NPI), the following imputation schemes for missing scale items was used:
• **Baseline: missing items were imputed using the** baseline mean (from all patients), rounded to the closest integer score for that individual item.
Post-Baseline: missing items were replaced with ratings from last non-missing previous visit. In cases where one or more specific scale items were missing at baseline and at all subsequent time points, no total score was calculated.
These imputation schemes were used if at least half of the items comprising the total score were present. If more than half of these items were missing, the corresponding total score was also missing. Imputed individual missing items were only used to calculate the corresponding total score but not stored in the data base.

Missing values/visit for (total) scores

Missing (total) values/visits were handled differently for the various populations. Therefore, the imputation scheme was always described with respect to the analysis population.
ITT(LOCF)
For the ITT population, the following imputation scheme for the total score was used:
• **If available, the scheduled assessment was used.**
• **If missing, the immediately preceding available** observation, scheduled or unscheduled, was utilized. Values more than 2 days after the last dose of study drug were not carried forward.
• **Evaluations assessed more than 2 days after** the last known date of study drug were not included in the analysis.
ITT+RDO(LOCF)
For the ITT+RDO population, the following imputation scheme for the total score was used:
• **If available, the scheduled assessment was used.**
• **If missing and the patient** returned for an efficacy assessment (retrieved dropout), the respective retrieved dropout assessment was used.
• **If missing and no retrieved dropout** assessment was available, then the immediately preceding available observation, scheduled or unscheduled, was utilized.

Key Hypotheses

There are four hypotheses which translate into six statistical null hypotheses:

1a) The change in the total score of ADAS-Cog from baseline to week 24 is equal between the Exelon patch 20 cm2 and the placebo group.

1b) The ADCS-CGIC score at week 24 is equal between the Exelon patch 20 cm2 and the placebo group.

2) The change in the total score of ADAS-Cog from baseline to week 24 differs between the Exelon patch 20 cm2 and the Exelon capsule group by a margin of -1.25.

3a) The change in the total score of ADAS-Cog from baseline to week 24 is equal between the Exelon patch 10 cm2 and the placebo group.

3b) The ADCS-CGIC score at week 24 is equal between the Exelon patch 10 cm2 and the placebo group.

4) The change in the total score of ADCS-ADL from baseline to week 24 is equal between the Exelon patch 20 cm2 and the placebo group.

These six statistical hypotheses are a priori ordered. In a first step, parts a) and b) were tested simultaneously and then parts 1 to 4 were tested sequentially in a confirmatory manner.

The statistical null hypotheses related to the secondary objectives followed the same scheme. However, the statistical testing was done in an exploratory fashion and the corresponding p-values are interpreted in a descriptive sense.

For the alternative testing strategy recommended by the FDA, hypotheses 2 and 4 are dropped while hypotheses 1a) and 1b) and 3a) and 3b) are retained (3a and 3b are renumbered to 2a and 2b).

Sample Size Calculations

For this study, the assumptions on delta (difference in means) and standard deviation (SD) for the change in ADAS-Cog and ADCS-CGIC from baseline are based on 24 week data from the Exelon® capsule studies which used ADAS-Cog and CIBIC-plus. The assumptions on change in ADCS-ADL from baseline are based on the 5 month data of galantamine, another cholinesterase inhibitor, (Tariot et al., 2000). All assumptions are summarized in Table 2.

In previous placebo-controlled Exelon® capsule trials in AD patients with a similar study design, a treatment difference to placebo in the ADAS-Cog change from baseline of approximately 2.5 points was observed in the ITT analysis. A non-inferiority margin of 1.25 points in the ADAS-Cog change from baseline has been chosen in order to preserve 50% of this effect.

Appears This Way
On Original

11

Table 2 Sponsor's Assumptions underlying Sample Size Calculations

	Superiority Exelon® patch 20 cm² over placebo		Non-inferiority Exelon® patch 20 cm² to capsules		Superiority Exelon® patch 10 cm² over placebo		Superiority Exelon® patch 20 cm² over placebo
Hypothesis	H1		H2		H3		H4
Efficacy variables	ADAS-Cog	ADCS-CGIC	ADAS-Cog		ADAS-Cog	ADAS-CGIC	ADCS-ADL
Delta	3.5	.4	1	0	2.5	.35	2.5
SD	7	1.2	7	7	7	1.2	10
Non-inferiority margin	.		1.25	1.25			

For illustrative purposes, Table 3 shows the number of patients (n) per treatment group required for 90% individual power, when using a two-sided t-test with a significant level of 5%, respectively, and the assumptions on delta and standard deviation of the primary efficacy variables as presented in Table 2. In order to reach an overall power of 80% for all of the first three hypotheses (which is defined as the product of the individual powers), the sample size was successively increased to n=260 patients per treatment group. As a result, a total sample size of 1040 patients with at least one post-baseline efficacy assessments on treatment was calculated. The power values for the individual comparisons of this sample size are also presented in Table 3.

Table 3 Sponsor's Calculated Sample Sizes Needed

	Superiority Exelon® patch 20 cm² over placebo		Non-inferiority Exelon® patch 20 cm² to capsules		Superiority Exelon® patch 10 cm² over placebo		Superiority Exelon® patch 20 cm² over placebo
Hypothesis	H1		H2		H3		H4
Efficacy variable	ADAS-Cog	ADCS-CGIC	ADAS-Cog		ADAS-Cog	ADCS-CGIC	ADCS-ADL
Delta			1	0			
n per group for 90% individual power	86	191	205	660	166	248	338
Individual power (%) with n=260 per group	99%	96%	95%	52%	98%	91%	81%

Pooling of countries

The statistical analyses of the efficacy variables were performed using country as "blocking" factor. Small countries were pooled to obtain a sufficient strata size; i.e. at least 24 patients per pooled country. The following pooling scheme of countries was decided prior to unblinding (based on all randomized patients) and was employed for analyses in which the country effect was examined.

• Scandinavia (SCA, N=28): Norway (NOR, N=21), Denmark (DNK, N=5), Finland (FIN, N=2).
• Latin America (LAM, N=53): Uruguay (URY, N=10), Venezuela (VEN, N=20), Guatemala (GTM, N=23)

Statistical methodology and assumptions

The primary analysis of cognitive function was to be based on the change from baseline of the total ADAS-Cog score. The treatment groups were to be compared using least square means derived by an analysis of covariance (ANCOVA) model with the following explanatory variables: treatment, country, and the baseline total ADAS-Cog score.

The primary analysis for the global clinical rating of change (ADCS-CGIC) was to be the treatment comparison based on a Cochran-Mantel-Haenzel test (CMH, van Elteren) using modified ridit scores with country as stratification variable.

The primary analysis of activities of daily living was to be based on the change from baseline of the total ADCS-ADL score. The statistical analysis was to be along the same lines as that for the ADAS-Cog.

The primary population for the confirmative testing of all four hypotheses was to be the ITT (LOCF) population as described above. For each superiority hypothesis, the corresponding confirmative statistical analysis was to be performed and the two-sided p-value for the difference between treatment groups was to be calculated.

For the non-inferiority hypothesis comparing the Exelon 20 cm2 patch and the 12 mg oral formulation groups on ADAS-Cog, the corresponding confirmative statistical analysis was to be performed and the lower bound of the two-sided 95%-confidence interval (95% -CI) for the difference between treatment groups was to be calculated. The non-inferiority margin for ADAS-Cog was to be 1.25 points.

3.1.1.4 Disposition of Patients

A total of 1464 patients were screened, of whom 1195 patients were randomized in 21 countries. The most common reasons for screening failures were unacceptable laboratory values (96 patients), did not meet diagnostic or severity criteria (52 patients), withdrawal of consent (52 patients) and unacceptable test procedure result (34 patients). Two patients ([0012-0014]/[0012-00016] and [0092-00018]/[0092-00020]) were randomized twice. First, both patients were randomized to target Exelon patch 10cm2, but were subsequently assigned to placebo in the second randomization. After the first randomization, patient [0012-00014] took no study drug at all and patient [0092-00018] was treated with Exelon patch 5cm2 for 4 days. Prior to database lock, it was decided to exclude the data from the first randomization ([0012-00014] and [0092-00018]) and present these in a separate listing.

Approximately 80% of patients in the Exelon patch and capsule treatment groups completed the study. In the placebo group, the proportion was somewhat higher at approximately 88% (Table 4). For Exelon-treated patients, discontinuations were primarily due to adverse events (AEs) and withdrawal of consent, with little apparent difference in rate for the capsule or patch groups. For placebo-treated patients, the proportion that discontinued due to AEs or withdrew their consent was lower than for those in the various Exelon groups. Only a small number of patients discontinued as a result of unsatisfactory therapeutic effect, although the frequency of such cases was lower in the Exelon patch groups than in the placebo or Exelon capsule group. Five of the randomized patients did not receive study medication.

Table 4 Patient Disposition for each treatment group- all patients

Disposition/Reason	Exelon 20 cm^2 n (%)	Exelon 10 cm^2 n (%)	Exelon capsule n (%)	Placebo n (%)	Total n (%)
Total number of patients					
Screened					1484
Randomized	303 (100.0)	293 (100.0)	297 (100.0)	302 (100.0)	1195 (100.0)
Exposed to study drug	303 (100.0)	291 (99.3)	294 (99.0)	302 (100.0)	1190 (99.6)
Completed	241 (79.5)	229 (78.2)	234 (78.8)	266 (88.1)	970 (81.2)
Discontinued	62 (20.5)	64 (21.8)	63 (21.2)	36 (11.9)	225 (18.8)
Adverse event(s)	26 (8.6)	28 (9.6)	24 (8.1)	15 (5.0)	93 (7.8)
Subject withdrew consent	19 (6.3)	21 (7.2)	17 (5.7)	6 (2.0)	63 (5.3)
Unsatisfactory therapeutic effect	4 (1.3)	3 (1.0)	8 (2.7)	6 (2.0)	21 (1.8)
Lost to follow-up	4 (1.3)	3 (1.0)	5 (1.7)	3 (1.0)	15 (1.3)
Death	5 (1.7)	4 (1.4)	2 (0.7)	3 (1.0)	14 (1.2)
Administrative problems	2 (0.7)	1 (0.3)	4 (1.3)	2 (0.7)	9 (0.8)
Protocol violation	2 (0.7)	3 (1.0)	2 (0.7)	1 (0.3)	8 (0.7)
Abnormal laboratory value(s)	0 (0.0)	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.1)
Subject's condition no longer required study drug	0 (0.0)	0 (0.0)	1 (0.3)	0 (0.0)	1 (0.1)

3.1.1.5 Patient Demographics

The treatment groups were well balanced and generally reflect the Exelon® patch target patient population (Table 5). The mean age was 73.6 years (range 50-90 years) with 87.1% of the patients ≥ 65 years old and 46.2% of patients > 75 years old. In the target 20 cm2 patch group, the proportion of patients who were > 75 years old was somewhat higher than in the other Exelon groups and the placebo group. Female patients comprised 66.6% of the population, reflecting the predominance of the disease in women. The racial composition of the population was 75% Caucasian, and approximately 9% Oriental and 15% "Other" (originating predominately from Latin American countries such as Chile, Mexico, Peru and Venezuela). The demographic composition of the treatment groups was similar and there were no differences which would be expected to have affected the interpretation of efficacy or safety.

Table 5 Baseline Demographics

Demographic variable	Exelon 20 cm² N = 303 n (%)	Exelon 10 cm² N = 291 n (%)	Exelon capsule N = 294 n (%)	Placebo N = 302 n (%)	Total N = 1190 n (%)
Age (years)					
Mean (SD)	74.2 (7.7)	73.6 (7.9)	72.8 (8.2)	73.9 (7.3)	73.6 (7.8)
Range	50-88	50-90	50-87	50-89	50-90
Age group – n (%)					
< 65 years	38 (12.5)	36 (12.4)	48 (16.3)	31 (10.3)	153 (12.9)
≥ 65 – ≤ 75 years	109 (36.0)	121 (41.6)	124 (42.2)	133 (44.0)	487 (40.9)
> 75 years	156 (51.5)	134 (46.0)	122 (41.5)	138 (45.7)	550 (46.2)
Sex – n (%)					
Male	103 (34.0)	93 (32.0)	101 (34.4)	101 (33.4)	398 (33.4)
Female	200 (66.0)	198 (68.0)	193 (65.6)	201 (66.6)	792 (66.6)
Race – n (%)					
Caucasian	227 (74.9)	220 (75.6)	219 (74.5)	227 (75.2)	893 (75.0)
Black	3 (1.0)	1 (0.3)	5 (1.7)	2 (0.7)	11 (0.9)
Oriental	27 (8.9)	25 (8.6)	29 (9.9)	27 (8.9)	108 (9.1)
Other	46 (15.2)	45 (15.5)	41 (13.9)	46 (15.2)	178 (15.0)

The background characteristics of the study population (Table 6), showed that approximately 25% of the patients had a relative with AD (most commonly the mother). Mean MMSE at baseline was 16.5 which is representative of a moderate disease state and is reflected in the high proportion (86.1%) of patients living with a caregiver or other individual. In addition, 4.4% of patients also met the criteria of probable dementia with Lewy bodies. Time since the first symptoms of AD were noticed by the patient or a caregiver ranged from 0 - 16.6 years, with a mean of 3.4 years across all treatment groups. The mean time since the first symptom of AD was diagnosed by a physician was 1.1 years in each of the treatment groups. The background characteristics of the treatment groups were similar and there were no differences which would be expected to have affected the interpretation of efficacy or safety.

Appears This Way
On Original

Table 6 Background Characteristics

Background characteristic	Exelon 20 cm² N = 303	Exelon 10 cm² N = 291	Exelon capsule N = 294	Placebo N = 302	Total N = 1190
Patient's relatives with AD disease – n (%)					
None	220 (72.6)	225 (77.3)	230 (78.2)	221 (73.2)	896 (75.3)
Mother	46 (15.2)	32 (11.0)	27 (9.2)	36 (11.9)	141 (11.8)
Father	9 (3.0)	11 (3.8)	11 (3.7)	16 (5.3)	47 (3.9)
Sibling	32 (10.6)	21 (7.2)	29 (9.9)	29 (9.6)	111 (9.3)
Other	14 (4.6)	15 (5.2)	12 (4.1)	19 (6.3)	60 (5.0)
Time since first symptom of AD was noticed by patient/caregiver (years)					
Mean (SD)	3.3 (2.5)	3.3 (2.2)	3.4 (2.3)	3.5 (2.4)	3.4 (2.3)
Range	0.0 - 16.6	0.3 - 15.7	0.1 - 16.0	0.2 - 15.0	0.0 - 16.6
Time since first symptom of AD was diagnosed by physician (years)					
Mean (SD)	1.1 (1.4)	1.1 (1.4)	1.1 (1.4)	1.1 (1.4)	1.1 (1.4)
Range	0.0 - 8.0	0.0 - 7.4	0.0 - 8.4	0.0 - 9.2	0.0 - 9.2
Patient who met criteria of probable dementia with Lewy bodies – n (%)					
No	289 (95.4)	276 (94.8)	282 (95.9)	291 (96.4)	1138 (95.6)
Yes	14 (4.6)	15 (5.2)	12 (4.1)	11 (3.6)	52 (4.4)
Patient's living situation – n (%)					
Living alone	30 (9.9)	43 (14.8)	35 (11.9)	27 (8.9)	135 (11.3)
Living with caregiver or other(s)	265 (87.5)	240 (82.5)	255 (86.7)	264 (87.4)	1024 (86.1)
Assisted living/group home	8 (2.6)	8 (2.7)	4 (1.4)	11 (3.6)	31 (2.6)
Years of formal education					
Mean (SD)	9.9 (4.4)	9.9 (4.3)	9.9 (4.4)	9.9 (4.3)	9.9 (4.3)
Range	0 - 20	0 - 19	0 - 18	0 - 20	0 - 20
MMSE at baseline					
Mean (SD) – n (%)	16.6 (2.9)	16.6 (3.1)	16.4 (3.1)	16.4 (3.0)	16.5 (3.0)
Range	10 - 24	6 - 24	9 - 26	10 - 20	6 - 26
< 15 – n (%)	75 (24.8)	68 (23.4)	89 (30.3)	88 (29.1)	320 (26.9)
≥ 15 – n (%)	228 (75.2)	222 (76.3)	205 (69.7)	213 (70.5)	868 (72.9)

3.1.1.6 Sponsor's Results

A total of 11.9% of the patients randomized were not included in the ITT population because they never received study drug (5 patients), or did not have a baseline or post-baseline assessment (on-treatment) for at least one of the primary efficacy variables (137 patients). The reason that a large number of patients were excluded from the ITT population was due to the stringent requirement to have a post-baseline measurement on-treatment and the fact that only two post baseline assessments were scheduled.

Table 7 Number(%) of patients in analysis populations-all randomized population

Analysis population	Exelon 20 cm^2 N = 303 n (%)	Exelon 10 cm^2 N = 293 n (%)	Exelon capsule N = 297 n (%)	Placebo N = 302 n (%)	Total N = 1195 n (%)
All randomized (RND)	303 (100.0)	293 (100.0)	297 (100.0)	302 (100.0)	1195 (100.0)
Safety	303 (100.0)	291 (99.3)	294 (99.0)	302 (100.0)	1190 (99.6)
ITT	264 (87.1)	251 (85.7)	256 (86.2)	282 (93.4)	1053 (88.1)
ITT+RDO	279 (92.1)	267 (91.1)	275 (92.6)	289 (95.7)	1110 (92.9)

Primary efficacy results (4 objectives)

Superiority of the target Exelon 20 cm2 patch versus placebo at Week 24 was based on simultaneous testing of ADAS-Cog and ADCS-CGIC in the ITT (LOCF) population. For ADAS-Cog, the p-value was < 0.001, substantially lower than the pre-specified value of 0.05. The p-value for ADCS-CGIC marginally exceeded the predefined significance level of 0.05, by 0.004. However, supportive analyses for ADCS-CGIC at Week 24, yielded consistent and statistically significant results across all other prespecified efficacy population datasets (ITT, ITT+RDO and RND with their respective imputation schemes) as well as for the predefined proportional odds model (discussed below in the ADCS-CGIC results section). This was regarded by the sponsor as substantial evidence of effectiveness. Therefore, although the first objective was not achieved as planned, testing was continued for the remaining 3 hypotheses in the hierarchical scheme. The corresponding three objectives were achieved as follows:

• **Non-inferiority of the Exelon 20 cm2 patch over** capsules at Week 24 was established as the 95%-CI was below non-inferiority margin of 1.25.

• **For the target Exelon 10 cm2 patch, superiority versus placebo at Week 24 was demonstrated** by simultaneous testing of ADAS-Cog and ADCS-CGIC, with respective p-values of 0.005 and 0.010.

• **Superiority of the target Exelon 20 cm2 patch** versus placebo at Week 24 with regard to ADCS-ADL was achieved with a p-value of 0.017.

Appears This Way
On Original

17

Table 8 Summary of Primary Efficacy Results, ITT(LOCF) population

Objective	Variable		
	ADAS-Cog	ADCS-CGIC	ADCS-ADL
1 Superiority of Exelon 20 cm^2 target patch size over placebo at Week 24, based on ADAS-Cog and ADCS-CGIC (simultaneous testing)	$p < 0.001$	$p = 0.054$	-
2 Non-inferiority of Exelon 20 cm^2 target patch size compared to Exelon 12 mg/day target capsules at Week 24, based on ADAS-Cog	(-2.06, 0.17)*	-	-
3 Superiority of Exelon 10 cm^2 target patch size over placebo at Week 24, based on ADAS-Cog and ADCS-CGIC (simultaneous testing)	$p = 0.005$	$p = 0.010$	-
4 Superiority of Exelon 20 cm^2 target patch size over placebo at Week 24, based on ADCS-ADL	-	-	$p = 0.017$

* Non-inferiority established, as the 95%-confidence interval for the difference between treatment groups (a negative difference indicates greater efficacy of Exelon 20 cm^2 versus capsule) was entirely below the corresponding predefined non-inferiority margin of 1.25.

Table 9 shows the sponsor's results for the hypothesis ordering recommended by the FDA(2 objectives).

Table 9 Summary of Primary efficacy results, ITT (LOCF) population (FDA suggested hypothesis ordering)

Objective	Variable	
	ADAS-Cog	ADCS-CGIC
1 Superiority of Exelon 20 cm^2 target patch size over placebo at Week 24, based on ADAS-Cog and ADCS-CGIC (simultaneous testing)	$p < 0.001$	$p = 0.054$
2 Superiority of Exelon 10 cm^2 target patch size over placebo at Week 24, based on ADAS-Cog and ADCS-CGIC (simultaneous testing)	$p = 0.005$	$p = 0.010$

ADAS-Cog

In all three efficacy population datasets, ITT (LOCF), ITT+RDO (LOCF) and ITT (OC), patients treated with Exelon (20 cm2 patch, 10 cm2 patch and capsule) showed significantly better performance in cognition, assessed by the ADAS-Cog at the primary endpoint, Week 24, compared to placebo, which showed deterioration. For each population, the mean improvements in ADAS-Cog were higher for the Exelon 20 cm2 group than for the Exelon 10 cm2 and Exelon capsule groups suggesting a possible dose response between the two Exelon patch size groups. In addition, the improvement in the target Exelon 20 cm2 patch group at Week 16 achieved nominal statistical significance relative to placebo for all three population datasets.

Table 10 ADAS-cog change from baseline – ITT (LOCF) population

Visit			Exelon 20 cm² N = 284	Exelon 10 cm² N = 251	Exelon Capsule N = 258	Placebo N = 282
Week 16		n	257	248	253	280
	Baseline	Mean	27.5	27.0	27.9	28.5
	Post-baseline	Mean	26.1	26.1	27.4	28.5
	Change	Mean	-1.4	-0.8	-0.5	-0.0
		p-value	0.007*	0.090	0.274	
Week 24		n	252	248	253	281
	Baseline	Mean	27.4	27.0	27.9	28.6
	Post-baseline	Mean	25.8	26.4	27.3	29.5
	Change	Mean	-1.6	-0.6	-0.5	1.0
		p-value	<0.001*	0.005*	0.003*	

Only patients with a valid baseline and post-baseline score at Week 16 or Week 24 were included
Negative change score indicates improvement.
p-values are derived from two-way analyses of covariance and are based on comparison of each Exelon treatment group with placebo.
* p < 0.05

In all three efficacy population datasets, ITT (LOCF), ITT+RDO (LOCF) and ITT (OC), statistical non-inferiority of target Exelon 20 cm2 patch over target Exelon 12 mg/day capsule with respect to ADAS-Cog was established at week 24, the primary endpoint. In addition, although not specified in the confirmatory testing strategy, non-inferiority of the 20 cm2 patch at week 16(see Table 11) and non-inferiority of the 10 cm2 patch at Weeks 16 and 24 were also established.

Table 11 ADAS-cog mean treatment difference in change from baseline

Population / visit		Exelon 20 cm² versus Capsule		
		ITT (LOCF)	ITT+RDO (LOCF)	ITT (OC)
Week 16	LS-Mean	-0.83	-0.79	-0.83
	LS 95%-CI	-1.88	-1.80	-1.89
	UB 95%-CI	0.22*	0.22*	0.22*
Week 24	LS-Mean	-0.95	-0.95	-1.14
	LS 95% CI	-2.06	-2.05	-2.35
	UB 95% CI	0.17*	0.14*	0.08*

A negative LS-mean treatment difference indicates superiority of Exelon 20 cm² versus capsule
Mean and 95%-Confidence interval of LS mean between treatments are derived from two-way analyses of covariance
* upper boundary of 95%-Confidence interval (UB 95%-CI) for the difference between treatment groups is below the corresponding pre-defined non-inferiority margin 1.25

19

ADCS-CGIC

After 24 weeks of treatment, the Exelon patch and the Exelon capsule formulations yielded better results on the ADCS-CGIC scale than placebo, although as previously discussed, the p-value for the target Exelon 20 cm2 patch group marginally exceeded the prespecified significance level of 0.05 in the primary efficacy analysis. However, supportive analyses for ADCS-CGIC at week 24, yielded consistent and statistically significant results across the prespecified efficacy population datasets of ITT+RDO (LOCF) and ITT (OC), with respective p-values of 0.034 and 0.029. Furthermore, for the three additional datasets/imputation schemes, the statistical tests for the primary analysis and for the proportional odds model (predefined for three dataset) yielded similarly significant results with p-values of < 0.04. Overall, statistical significance (at the 5% level) for ADCS-CGIC in the target Exelon 20 cm2 patch group compared to placebo was achieved in 8 of the 9 pre-specified statistical tests at Week 24 which was regarded by the sponsor as substantial evidence of effectiveness.

Table 12 ADAS-CGIC categorical analysis - ITT (LOCF) population

Visit	Exelon 20 cm^2 N = 264	Exelon 10 cm^2 N = 261	Exelon capsule N = 268	Placebo N = 282
Week 16 – n (%)				
Markedly improved (1)	3 (1.2)	4 (1.6)	1 (0.4)	1 (0.4)
Moderately improved (2)	21 (8.2)	24 (9.7)	20 (8.0)	22 (8.1)
Minimally improved (3)	58 (22.7)	48 (19.4)	48 (19.3)	53 (19.4)
Unchanged (4)	109 (42.7)	104 (42.1)	111 (44.6)	116 (42.5)
Minimally worse (5)	40 (15.7)	53 (21.5)	43 (17.3)	53 (19.4)
Moderately worse (6)	20 (7.8)	14 (5.7)	23 (9.2)	25 (9.2)
Markedly worse (7)	4 (1.6)	0 (0.0)	3 (1.2)	3 (1.1)
n	255	247	249	273
mean	3.9	3.9	4.0	4.0
SD	1.13	1.08	1.13	1.10
p-value	0.177	0.195	0.034	
Week 24 – n (%)				
Markedly improved (1)	5 (1.9)	5 (2.0)	3 (1.2)	2 (0.7)
Moderately improved (2)	32 (12.3)	29 (11.7)	29 (11.5)	26 (9.4)
Minimally improved (3)	48 (18.5)	43 (17.3)	60 (23.7)	50 (18.0)
Unchanged (4)	94 (36.2)	105 (42.3)	95 (37.9)	91 (32.7)
Minimally worse (5)	50 (19.2)	41 (16.5)	30 (11.9)	65 (23.4)
Moderately worse (6)	27 (10.4)	22 (8.9)	30 (11.9)	36 (12.9)
Markedly worse (7)	4 (1.5)	3 (1.2)	5 (2.0)	8 (2.9)
n	260	248	253	278
mean	4.0	3.9	3.9	4.2
SD	1.27	1.20	1.25	1.26
p-value	0.034	0.010*	0.009*	

p-values are derived from CMH test (van Elteren test) blocking for country and are based on comparison of each Exelon treatment group with placebo.
* p < 0.05

Secondary Endpoints

For all three efficacy population datasets, the primary ITT (LOCF), and the secondary ITT+RDO (LOCF) and ITT (OC), patients in both the Exelon patch groups and the Exelon capsule group showed significantly better performance in the activities of daily living test (ADCS-ADL) at primary endpoint of Week 24 compared to placebo-treated patients (Table 13).

Table 13 ADCS-ADL change from baseline – ITT (LOCF) population

Visit			Exelon 20 cm^2 N = 264	Exelon 10 cm^2 N = 251	Exelon capsule N = 256	Placebo N = 282
Week 16		n	261	247	253	280
	Baseline	Mean	47.5	50.1	49.3	49.2
	Post-baseline	Mean	47.8	49.5	48.9	47.7
	Change	Mean	0.4	-0.6	-0.4	-1.6
		p-value	0.035*	0.226	0.143	
Week 24			263	247	254	281
	Baseline	Mean	47.6	50.1	49.3	49.2
	Post-baseline	Mean	47.6	49.9	48.8	46.9
	Change	Mean	-0.0	-0.1	-0.5	-2.3
		p-value	0.017*	0.013*	0.039*	

Only patients with a valid baseline and post-baseline score at Week 16 or Week 24 were included
Positive change score indicates improvement
p-values are derived from two-way analyses of covariance and are based on comparison of each Exelon treatment group with placebo.
* $p < 0.05$

NPI

NPI scores were analyzed for both the 12 item (NPI-12), 10 item (NPI-10) and caregiver distress scales. There were no significant differences versus placebo on the NPI-10 or NPI-12 for any of the Exelon treatment groups.

MMSE

For the ITT (LOCF) population dataset, the Exelon patch groups and the Exelon capsule group showed a greater increase in MMSE score from baseline than placebo at Weeks 16 and 24, indicating clinical improvement in cognitive symptoms of dementia (Table 14).

Appears This Way
On Original

Table 14 MMSE change from baseline – ITT (LOCF) population

Visit			Exelon 20 cm² N = 264	Exelon 10 cm² N = 251	Exelon capsule N = 256	Placebo N = 282
Week 16		n	259	250	256	281
	Baseline	Mean	16.6	16.7	16.4	16.4
	Post-baseline	Mean	17.8	17.7	17.0	16.6
	Change	Mean	1.1	1.0	0.6	0.2
		p-value	< 0.001*	0.007*	0.108	
Week 24		n	262	250	256	281
	Baseline	Mean	16.6	16.7	16.4	16.4
	Post-baseline	Mean	17.6	17.8	17.2	16.4
	Change	Mean	0.9	1.1	0.8	0.0
		p-value	0.002*	< 0.001*	0.002*	

Only patients with a valid baseline and post-baseline score at Week 16 or Week 24 were included
Positive change score indicates improvement.
p-values are derived from CMH test (van Elteren test) blocking for country and are based on
comparison of each Exelon treatment group with placebo.
* p < 0.05

3.1.1.7 Reviewer's Results

3.1.1.7.1 Primary Analysis of ADAS-Cog

The sponsor's primary analysis result for the change from baseline to week 24 in the ADAS-cog was verified. There was a small discrepancy between the reviewer and the sponsor in the number of patients included in the ITT(LOCF) analysis but the results were nevertheless essentially the same. The Exelon 20 cm2 patch group had a statistically significantly better LS mean change in the ADAS-cog than the placebo group at week 24 (or LOCF) as seen in Table 15. The comparisons of the Exelon 10 cm² group and the Exelon capsule group with placebo also reached nominal significance.

Appears This Way
On Original

Table 15 Change from Baseline to Week 24 in ADAS-cog in ITT(LOCF) Population (Reviewer's Results)

GROUP	N	BASELINE MEAN	LS MEAN (SE) CHANGE	DIFFERENCE FROM PLACEBO MEAN (SE)	P-VALUE FOR COMPARISON WITH PLACEBO
Placebo	281	28.6	1.1(0.4)	N/A	N/A
EXELON 20 cm²	261	27.4	-1.5(0.4)	2.6(0.6)	<0.001
EXELON 10 cm²	251	27.2	-0.7(0.4)	1.8(0.6)	0.002
EXELON Capsule	255	28.1	-0.8(0.4)	1.9(0.6)	0.001

3.1.1.7.2 Assessment of Sensitivity of Analysis of ADAS-cog to Missing Data

The analysis of the ADAS-cog was still significant for each drug group compared to placebo when restricted to the Observed Cases population (Table 16).

Table 16 Change from Baseline to Week 24 in ADAS-cog in Observed Cases Population (Reviewer's Results)

GROUP	N	BASELINE MEAN	LS MEAN (SE) CHANGE	DIFFERENCE FROM PLACEBO MEAN (SE)	P-VALUE FOR COMPARISON WITH PLACEBO
Placebo	259	28.5	1.2(0.4)	N/A	N/A
EXELON 20 cm²	225	27.4	-1.7(0.5)	2.9(0.6)	<0.001
EXELON 10 cm²	223	26.8	-0.5(0.5)	1.7(0.6)	0.004
EXELON Capsule	225	28.0	-1.0(0.4)	2.2(0.6)	<0.001

A mixed model for repeated measures was also investigated. This model assumed the most general structure for the correlation between scores over time within patient and fit a separate mean for each visit and each treatment group as well as each possible combination of visit and treatment group. The model also was adjusted for country effects and the baseline ADAS-cog score. The results of the mixed model analysis were very close to the ANCOVA results based on the observed case population.

Therefore, because of the agreement between the LOCF, Observed case, and mixed model analyses, there is no indication that missing data caused bias in the primary analysis result for the change in ADAS-cog at week 24.

The graph below shows the observed changes in ADAS-cog over time in the Exelon 20 cm2 patch and placebo group, as well as the LOCF means at week 24, and the week 24 means as estimated by a mixed model for repeated measures.

Appears This Way
On Original

Figure 1 Plot of Change from Baseline in ADAS-Cog over Time in Exelon 20 cm2 patch and Placebo groups
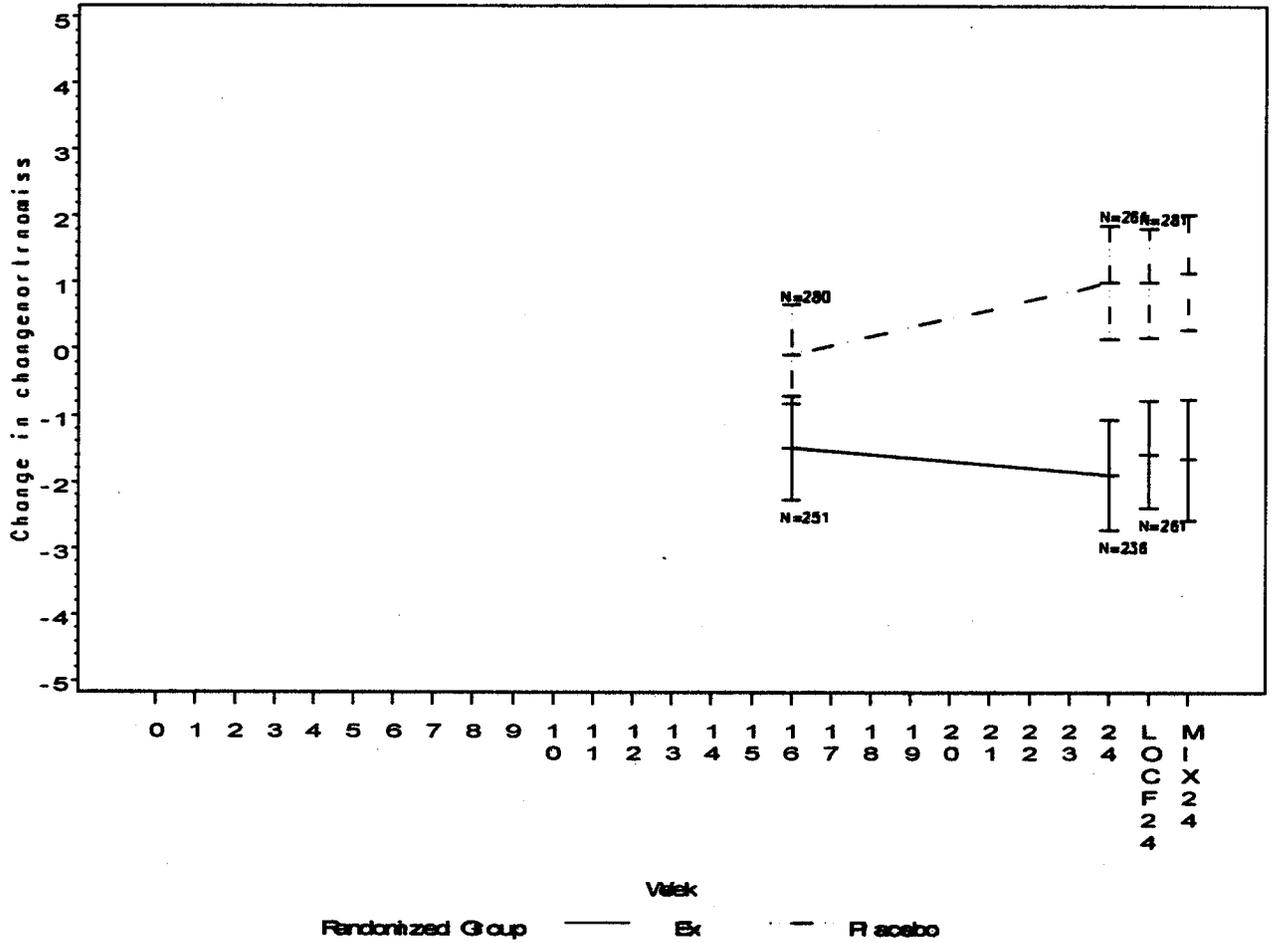
Appears This Way
On Original

More patients dropped out without having any post-baseline assessments in the drug groups (14%) than placebo (8%). Among the groups the number of patients that dropped out but had a week 16 ADAS-cog assessment ranged from 20(7%) to 27 (9%).

The analysis of the change from baseline in ADAS-cog which is the same as the LOCF analysis except it includes patients with no post-baseline ADAS-cog measures by assuming no change, is also significant at the nominal level favoring the Exelon 20 cm2 group over the placebo group. Therefore, we have some assurance that the dropouts with no post-baseline ADAS cog measures, although a considerable proportion of all randomized patients, would not have changed the result had they had a post-baseline ADAS-cog measured.

Table 17 ADAS-cog Analysis assuming no change for patients with no post baseline scores and LOCF for others

GROUP	N	LS MEAN (SE) CHANGE	DIFFERENCE FROM PLACEBO MEAN (SE)	P-VALUE FOR COMPARISON WITH PLACEBO
Placebo	302	1.0 (0.4)	N/A	N/A
EXELON 20 cm²	303	-1.3 (0.4)	2.4 (0.5)	<0.001
EXELON 10 cm²	293	-0.6 (0.4)	1.7 (0.5)	0.001
EXELON Capsule	296	-0.6 (0.4)	1.7 (0.5)	0.001

Appears This Way
On Original

3.1.1.7.3 Analysis of Co-Primary Endpoint CGI

Table 18 summarizes this reviewer's results for the primary analysis of the CGI at week 24 in the ITT(LOCF) population. This reviewer found a slightly larger p-value for the comparison of the high dose patch and placebo. The reason for this is explained below in section 3.1.1.7.5.

Table 18 Distributions of CGI at Week 24 in ITT(LOCF) Population (Reviewer's Results)

GROUP	PLACEBO	EXELON 20 CM^2	EXELON 10 CM^2	EXELON CAPSULE
CGI=1 N(%)	2 (0.7)	4 (1.5)	5 (2.0)	4 (1.6)
CGI=2 N(%)	24 (8.6)	34 (13.1)	30 (12.1)	29 (11.5)
CGI=3 N(%)	56 (20.1)	49 (18.8)	43 (17.3)	59 (23.3)
CGI=4 N(%)	90 (32.4)	91 (35.0)	107 (43.1)	95 (37.5)
CGI=5 N(%)	64 (23.0)	49 (18.8)	37 (14.9)	31 (12.3)
CGI=6 N(%)	34 (12.2)	29 (11.2)	23 (9.3)	30 (11.9)
CGI=7 N(%)	8 (2.9)	4 (1.5)	3 (1.2)	5 (2.0)
N Total	278	260	248	253
Mean CGI	4.17	3.96	3.90	3.91
CMH (ANOVA) p-value for comparison w/ placebo	N/A	0.096	0.018	0.018

3.1.1.7.4 Assessment of Sensitivity of Analysis of CGI to Missing Data

The results for the observed cases population were slightly more favorable for the Exelon 20 cm2 patch than the ITT(LOCF) population results and the p-value was just below the nominal level.

Table 19 Distributions of CGI at Week 24 in Observed Cases Population (Reviewer's Results)

GROUP:	PLACEBO	EXELON 20 CM^2	EXELON 10 CM^2	EXELON CAPSULE
CGI=1 N(%)	2 (0.8)	4 (1.7)	5 (2.2)	4 (1.7)
CGI=2 N(%)	24 (9.2)	33 (14.0)	27 (12.0)	28 (12.1)
CGI=3 N(%)	53 (20.2)	48 (20.3)	42 (18.7)	58 (25.1)
CGI=4 N(%)	86 (32.8)	80 (33.9)	94 (41.8)	84 (36.4)
CGI=5 N(%)	58 (22.1)	42 (17.8)	32 (14.2)	25 (10.8)
CGI=6 N(%)	32 (12.2)	27 (11.4)	22 (9.8)	28 (12.1)
CGI=7 N(%)	7 (2.7)	2 (0.8)	3 (1.3)	4 (1.7)
N Total	262	236	225	231
Mean CGI	4.14	3.90	3.88	3.86
CMH (ANOVA) p-value for comparison w/ placebo	N/A	0.046	0.024	0.009

The following analysis of the CGI at week 24 (shown in Table 20) which is the same as the LOCF analysis except it includes retrieved dropouts by using their retrieved score and patients with no post-baseline ADAS-cog measures by assuming no change, gives similar results to those previously described. Therefore, we have some assurance that the exclusion of randomized patients with no post-baseline CGI scores from the primary analysis did not cause bias.

Table 20 Distributions of CGI at Week 24 in All Randomized* (Reviewer's Results)

GROUP:	PLACEBO	EXELON 20 CM²	EXELON 10 CM²	EXELON CAPSULE
CGI=1 N(%)	2 (0.7)	5 (1.7)	6 (2.1)	4 (1.4)
CGI=2 N(%)	25 (8.3)	34 (11.2)	30 (10.3)	29 (9.9)
CGI=3 N(%)	56 (18.5)	53 (17.5)	43 (14.8)	62 (21.1)
CGI=4 N(%)	109 (36.1)	123 (40.6)	138 (47.4)	127 (43.2)
CGI=5 N(%)	64 (21.2)	51 (16.8)	43 (14.8)	36 (12.2)
CGI=6 N(%)	36 (11.9)	32 (10.6)	26 (8.9)	30 (10.2)
CGI=7 N(%)	10 (3.3)	5 (1.7)	5 (1.7)	6 (2.0)
N Total	302	303	291	294
Mean CGI	4.18	3.98	3.96	3.94
CMH (ANOVA) p-value for comparison w/ placebo	N/A	0.060	0.039	0.012

*using retrieved dropout scores and assuming no change for those with no post-baseline CGI scores

3.1.1.7.5 Discrepancies between Reviewer and Sponsor on Last CGI score used for Analysis

CGI assessments were scheduled for Visits 6 and 7 only. Visit 6 was to take place at the end of week 16 and visit 7 was to take place at the end of week 24. It appears that the last CGI in the double blind period was not always used in the primary LOCF analysis, apparently, because some patients were allowed to enter the open label period just before having their week 24 CGI assessment. In particular, in some cases the second to last CGI was used in the LOCF analysis. The sponsor replied to a question on this matter that all of these 16 cases (8 placebo and 8 Exelon 20 cm2) where the last CGI was not used in the analysis could be explained by the fact that the last CGI score was actually in the open label period as one could determine from the dosing information dataset, DAR.xpt. However, both the raw CGI dataset and the DAR dataset have a variable to indicate whether or not the record comes from the open label extension and these two variables disagree in all 16 cases in question. In particular, the DAR dataset indicates that the record is in the open label period but the CGI dataset indicates that the record is in the double blind period. If we accept that these week 24 assessments were made in the open label period we still must consider that for 14 of the 16 cases the assessment was made on the first day of the open label period. Note that the beginning of the open label extension all continuing patients were started with the Exelon 10cm2 patch but their previous double blind treatment was not revealed, i.e., they might have been able to tell the size of the patch they had been on but it was not revealed whether it had contained placebo or Exelon. It is unknown in the 14 cases where the double blind week 24 CGI assessment was reportedly taken in the open label phase,

but only on the first day of it, whether or not the assessment was made before the open label drug was administered. It is also unknown whether the drug could have an effect in only 1 day for placebo patients beginning the 10 cm2 patch or for Exelon 20 cm2 patients who had tolerated the Exelon 20 cm2 patch and were switched to the Exelon 10 cm2 patch at the start of the open label phase. This requires clinical judgment. Note that the only CGI assessment scheduled for the open label extension period was at Week 52, i.e., at the end of the 28 week open label extension period. If one is unwilling to accept the Week 24 ratings that occurred, for the most part (in 14 of the 16 cases), only 1 day after the end of the double blind period for the primary analysis then they can at least be used to perform a sensitivity analysis.

If one excludes all of these 16 patients last scores then the p-value of 0.054 reported by the sponsor results.
If the last score is used for patient 91-00005 even though it occurred at day 258 well beyond week 24 and their first assessment at day 203, then the p-value becomes 0.043.
However, all but two of these patients (in the placebo group) were on double blind treatment within one day of the Week 24 assessment so their week 24 assessments would be eligible for inclusion in the ITT(LOCF) analysis in that regard. The p-value that results from this sensitivity analysis in which all 16 of these last CGI scores are used is less favorable to the Exelon 20 cm2 patch (p=0.096). If one excludes **patient 91-00005's Week 24 CGI** score which occurred very late compared to the schedule, at day 258, in preference for the Visit 6 score (day 203) which also occurred late (but closer to the schedule) then the p-value is 0.120.
If we take the week 24 assessments from all of these 16 patients except for the two placebo patients whose week 24 CGI rating occurred more than one day into the open label period and therefore, their week 24 score would be excluded by the protocol requirement to only use scores within 2 days of the last dose, the p-value for the Exelon 20 cm2 vs. placebo comparison is 0.0842.
If we take the week 24 assessments from all of these 16 patients except for the two placebo patients just mentioned and **patient 91-00005's [since their earlier** assessment occurred after week 24 and closer to week 24 than the last CGI assessment (day 203 vs. day 258)] then the p-value increases to 0.109.

If we exclude all 8 of the placebo scores that occurred in the open label period but we retain the 8 Exelon 20 scores all of which occurred only 1 day into the open label period then the p-value is 0.075 (utilizing the day 203 score for patient 9100005 since it is closer to week 24 than day 258; the value is 0.060 if the day 258 score is used).

If we entirely exclude all 16 patients from the analysis then the p-value for the comparison of Exelon 20 cm2 and placebo on the CGI is 0.056.

Table 21 summarizes how the results depend on which CGI scores are used in the analysis for the 16 cases whose Week 24 CGI score may have actually been taken early in the Open Label extension period.

Table 21 Impact of Week 24 CGI scores possibly taken early in the Open Label extension

Method	PLACEBO		EXELON 20 CM2 PATCH		CMH TEST P-VALUE FOR EXELON 20 VS. PLACEBO
	N	Mean	N	Mean	
Sponsor's Primary: Use Week 16 score where Week 24 CGI scores possibly taken 1 day into OL	278	4.19	260	3.96	0.054
Using all Week 24 CGI scores possibly taken 1 day into OL	278	4.17	260	3.97	0.109
Using all Week 24 CGI scores possibly taken up to 11 days into OL	278	4.17	260	3.97	0.120
Use Week 16 score for Placebo only where Week 24 CGI scores possibly taken 1 day into OL	278	4.19	260	3.97	0.075
Entirely Excluding patients whose Week 24 CGI scores possibly taken in OL	272	4.19	252	3.96	0.055

The net result is that the outcome of the primary analysis of the CGI on the sponsor's primary analysis population may be somewhat worse than the sponsor reported. While the p-value of 0.054 they reported is relatively close to 0.050 this reviewer found that the true p-value may be closer to 0.10. The latter p-value raises doubts about the effect of the high dose patch on the CGI given the relatively large sample size. The sponsor's choice of primary population which excludes assessments that were made more than 2 days after the last dose of study treatment would seem to be the ideal population for demonstrating a treatment effect. Also, the relatively large sample size of the study would seem to be ideal for demonstrating an effect. Nevertheless, the evidence for the effect of the high dose patch on the CGI is relatively weak.

Figure 2 shows how the CGI scores changed over time for the 8 Placebo patients whose DB Week 24 CGI assessment was possibly taken early in OL period. The sponsor excluded these assessments even though 6 of them occurred only 1 day after the end of the double blind period (for the other 2 the assessments were less than 12 days after then end of the DB period) and they should have been taken before the patients were allowed to enter the open label period.

Figure 2 Placebo patients whose DB Week 24 CGI assessment was possibly taken early in OL period
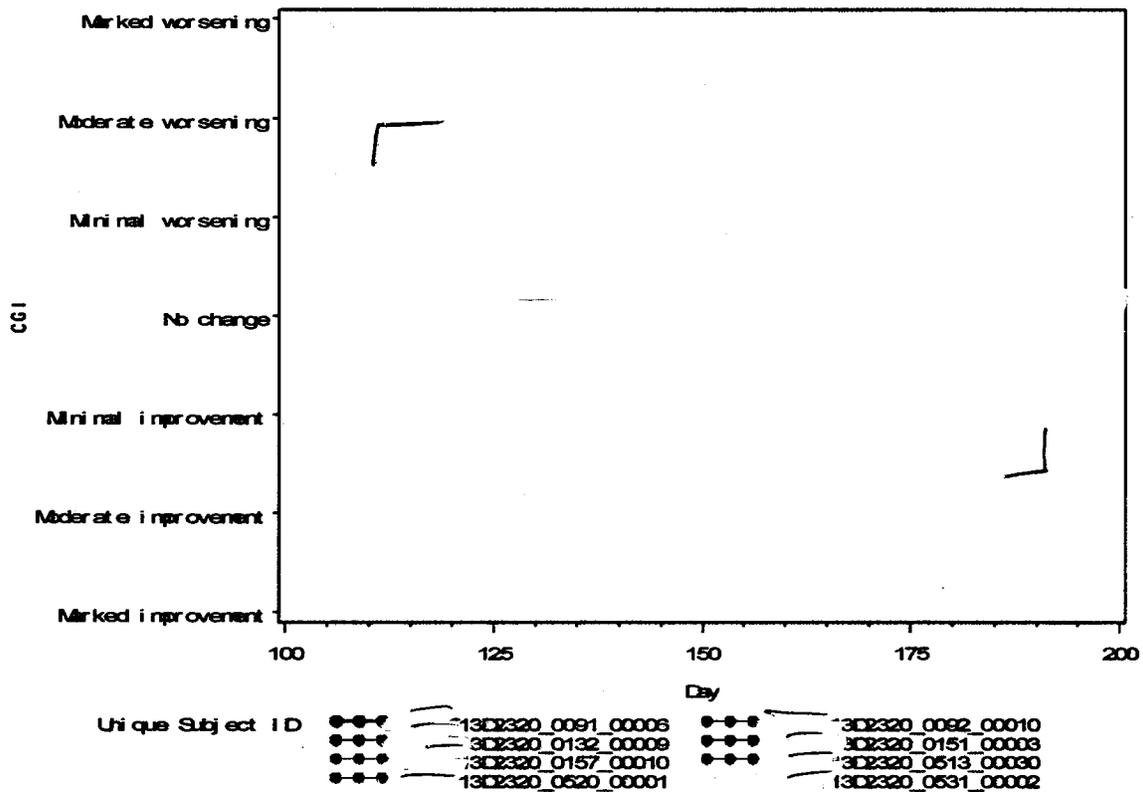
b(4)

b(4)

Note: the dashed line indicates the day the patient entered the open label extension period

Appears This Way
On Original

Figure 3 shows how the CGI scores changed over time for the 8 Exelon 20 cm2 group patients whose DB Week 24 CGI assessment was possibly taken early in OL period. The sponsor excluded these assessments even though all of them occurred only 1 day after the end of the double blind period and they should have been taken before the patients were allowed to enter the open label period. Note that one patient's assessments were much later than scheduled and the earlier assessment was closer to week 24 than the later one for this patient.

Figure 3 Exelon 20 cm2 patients whose DB Week 24 CGI assessment was possibly taken early in OL period
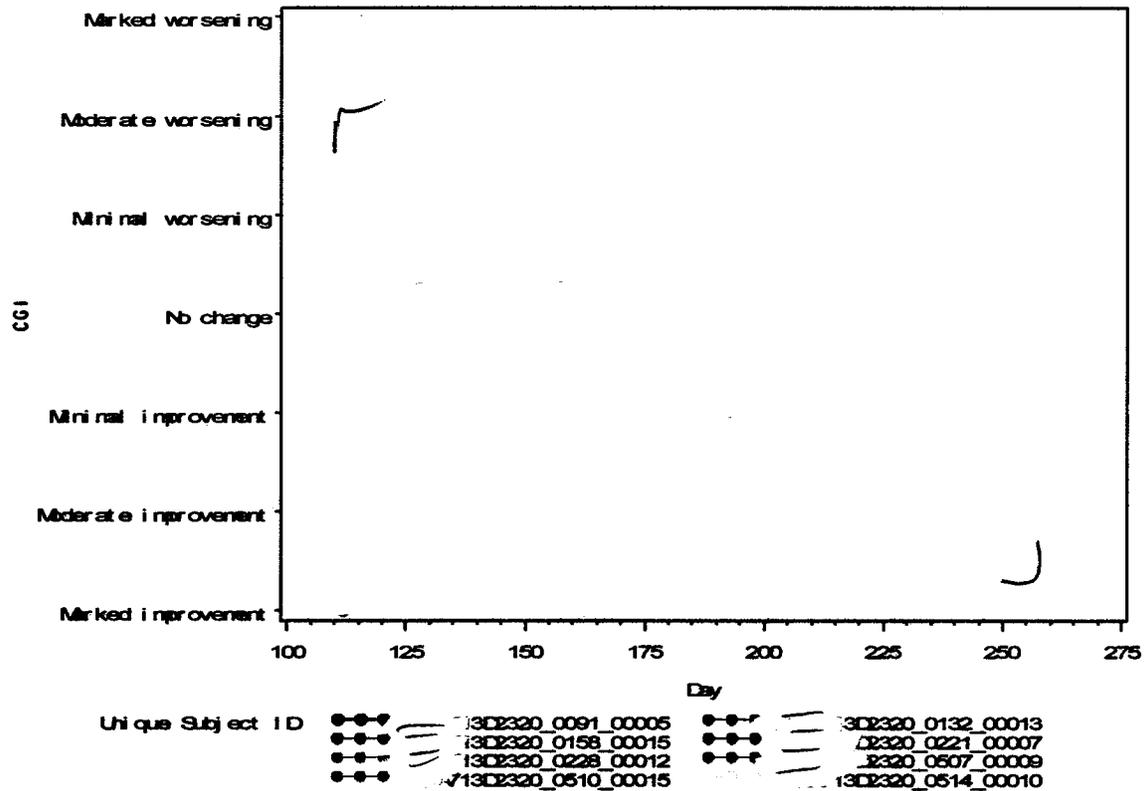
b(4)

b(4)

Note: the dashed line indicates the day the patient entered the open label extension period

Table 22 summarizes the cases where the sponsor and the reviewer used different CGI scores in the primary LOCF analysis and the timing of the assessments relative to the end of the DB and the start of the open label periods.

Table 22 Cases where CGI score indicated as Week 24 value not used by sponsor in their ITT(LOCF) Population - Exelon 20 cm2 and Placebo groups

Treatment Group Description	Unique Subject ID	Sponsor's Last CGI (GLBC)	Reviewer's Last CGI (LCGICHGX AFTERLDOSP2)	Time of Last Visit or Time of Second to Last Visit	Time of Week 24 Visit	DB* End day	Week 24 CGI Day - DB End Day	First OL Day
Exelon 20 cm2	13D2320_0091_00005			258	258	258	0	259
	713D2320_0132_00013			112	173	172	1	173
	13D2320_0158_00015			114	169	168	1	169
	13D2320_0221_00007			115	171	170	1	171
	713D2320_0228_00012			118	168	167	1	168
	713D2320_0507_00009			112	167	166	1	167
	713D2320_0510_00015			109	164	163	1	164
	713D2320_0514_00010			110	168	167	1	168
	Mean Score	3.88	4.00					
Placebo	13D2320_0091_00006			119	190	181	9	182
	713D2320_0092_00010			120	177	176	1	177
	13D2320_0132_00009			111	172	171	1	172
	713D2320_0151_00003			113	172	171	1	172
	13D2320_0157_00010			126	181	180	1	181
	13D2320_0513_00030			118	167	166	1	167
	13D2320_0520_00001			123	176	175	1	176
	13D2320_0531_00002			129	196	185	11* "crf" has fewer dosing records than dataset	186
	Mean Score	4.25	3.38					

b(

* as determined from DAR.xpt dosing administration dataset using the EXT1N variable which indicates whether or not the dosing record belongs to the open label extension

There were 7 Exelon 10 cm2 patients and 4 Exelon Capsule group patients that had the same issue but the conclusions for these groups do not depend on which scores are chosen for the analysis for these patients. Four out of the 7 Exelon 10 cm2 patients had their Week 24 CGI assessments only one day into the open label period. If these CGI values are used instead of carrying the week 16 assessment forward as the sponsor did, then the p-value for the Exelon 10 cm2 vs. placebo comparison is 0.013, which is **very close to the sponsor's result.**

In fact, if instead of starting the testing with a test of the high dose vs. placebo at 0.05 and only testing the low dose if the high dose was significant at 0.05 the sponsor had chosen a Bonferroni multiplicity adjustment then the low dose patch would have achieved statistical significance on both the ADAS-cog and the CGI at week 24 (at the 0.025 level).

It is interesting to note though that none of the three drug groups were nominally significantly better than placebo in terms of the CGI at week 16, the earlier time point at which the CGI was assessed. **From the sponsor's report we can determine** that the lowest p-value for comparing a drug group with placebo at week 16 was 0.177. This reviewer found that even if the drug groups were combined for the week 16 analysis of the CGI there would be no difference from placebo apparent at week 16, despite there being 1024 patients in the analysis.

3.1.1.7.6 Observed Estimate of Treatment Effect by Investigator

There were 100 individual centers that randomized at least 1 patient. The mean difference in change from baseline in ADAS-cog between the Exelon 20 cm2 patch group and placebo favored Exelon 20 cm2 in 59 of the 80 centers that had at least one patient per group with post baseline efficacy data. The primary analysis of the change from baseline in the ADAS-cog at week 24 was an ANCOVA adjusted for countries rather than individual investigators. Mean differences favored Exelon 20 cm2 over placebo in 16 of 20 countries that had at least one patient per group with post baseline efficacy data. The results did not appear to be driven by any individual center or country. Although they are not shown, similar conclusions were drawn about the observed treatment effects for particular investigators and countries for the Exelon 10 cm2 vs. placebo comparison.

Figure 4 Exelon 20 cm2 vs. Placebo Mean Difference in ADAS-cog change by Investigator
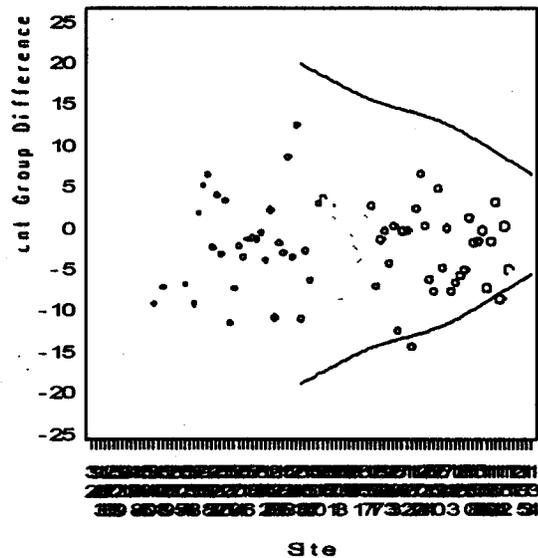
The primary analysis of the CGI was a country stratified Van Elteren test (Cochran Mantel Haenszel test with modified ridit scores). Although the primary analysis of the CGI was nonparametric and not based on the mean scores it is convenient to investigate center effects based on mean differences. The mean difference between the Exelon 20 cm2 patch group and placebo numerically favored Exelon 20 cm2 in 42 of the 81 centers that had at least one patient per group with post baseline efficacy data. Mean differences favored Exelon 20 cm2 over placebo in 10 of the 20 countries that had at least one patient per group with post baseline efficacy data. The results did not appear to be driven by any individual center or country. Although they are not shown, similar conclusions were drawn about the observed treatment effects for particular investigators and countries for the Exelon 10 cm2 vs. placebo comparison.

Figure 5 Exelon 20 cm2 vs. Placebo Mean Difference in CGI by Investigator
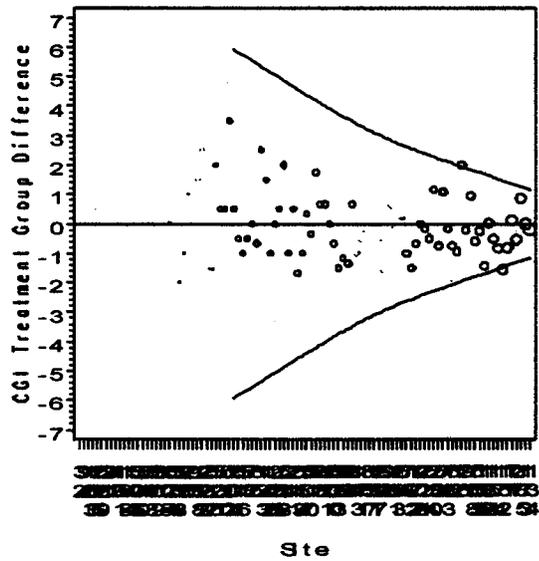
3.2 Evaluation of Safety

Safety is not evaluated in this review. Please see the clinical review(s) for the evaluation of safety.

Appears This Way
On Original

35

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

This section contains this reviewer's summary statistics for gender, race, and age subgroups. The studies were not adequately powered to estimate treatment differences in subgroups precisely and reported p-values should be interpreted cautiously as they have not been adjusted for multiple comparisons.

<u>Gender</u>

Sixty six percent (66%) in the sponsor's ITT(LOCF) population were female. There are no clear gender differences in efficacy in terms of the change in the ADAS-cog. The smaller sample size in males may explain the lack of nominal significance there. Furthermore, a test for interaction between treatment and gender is not significant.

Table 23 Gender Subgroups: Mean Change from Baseline to Week 24 or LOCF in ITT(LOCF) Population

	MALE			FEMALE			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	93	0.4 (7.0)	.	188	1.3 (6.8)	.	281	1.0 (6.9)
Ex 20 cm2 patch	94	-1.9 (7.7)	0.022	167	-1.4 (5.8)	<0.001	261	-1.6 (6.6)
Ex 10 cm2 patch	80	-1.0 (6.8)	0.207	171	-0.7 (6.5)	0.003	251	-0.8 (6.6)
Ex Capsule	92	-0.4 (7.0)	0.52	163	-1.0 (6.3)	<0.001	255	-0.8 (6.6)

Treatment by Gender Interaction test p-value= 0.4553

In terms of the CGI the treatment group difference between Exelon 20 cm2 and placebo was numerically larger in males than females.

If there was in fact a larger effect in males than females, the larger subgroup, it would help to explain the modest overall effect on the CGI. However, as seen above, there was no apparent gender difference in effects on the ADAS-cog which calls into question the apparent gender difference on the CGI.

Table 24 Gender Subgroups: Mean CGI at Week 24 or LOCF in ITT(LOCF) Population

CGI	MALE			FEMALE			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	93	4.2 (1.2)	.	185	4.1 (1.3)	.	278	4.1 (1.3)
Ex 20 cm2 patch	94	3.8 (1.2)	0.101	166	4.1 (1.3)	0.662	260	4.0 (1.3)
Ex 10 cm2 patch	79	3.8 (1.3)	0.219	169	3.9 (1.2)	0.054	248	3.9 (1.2)
Ex Capsule	91	4.0 (1.4)	0.433	162	3.9 (1.2)	0.018	253	3.9 (1.3)

Treatment by Gender Interaction test p-value=   0.128

Age

Ages ranged from 50 to 90. The mean age was 73 and the median age was 75.
There was no compelling evidence that the treatment effects between any of the Exelon groups
and placebo on the ADAS-cog varied significantly with age.

Table 25 Age Subgroups: Mean Change from Baseline to Week 24 or LOCF in ITT(LOCF) Population

	AGE ≤ 69			69 < AGE ≤ 75			75 < AGE ≤ 79			AGE > 79			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	70	0.4 (6.2)	.	81	1.0 (7.0)	.	65	1.2 (6.5)	.	65	1.5 (7.9)	.	281	1.0 (6.9)
Ex 20 cm2 patch	65	-2.4 (6.2)	0.008	69	-1.2 (6.5)	0.052	66	-2.3 (6.0)	0.002	61	-0.3 (7.2)	0.153	261	-1.6 (6.5)
Ex 10 cm2 patch	66	-0.2 (5.7)	0.657	73	-2.2 (7.2)	0.001	58	0.2 (6.1)	0.332	54	-0.5 (7.2)	0.138	251	-0.8 (6.6)
Ex Capsule	75	-0.9 (7.4)	0.237	75	-0.9 (6.3)	0.049	46	-0.5 (6.7)	0.15	59	-0.8 (5.6)	0.058	255	-0.8 (6.6)

Treatment by Age subgroup Interaction test p-value=0.35

The apparently larger treatment effects on the CGI in the 75-79 age group could be attributable to sampling variability and the smaller sample size in the subgroups.

Table 26 Age Subgroups: Mean CGI score at Week 24 or LOCF in ITT(LOCF) Population

CGI	AGE ≤ 69			69 < AGE ≤ 75			75 < AGE ≤ 79			AGE > 79			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	69	4.0 (1.2)	.	79	4.0 (1.3)	.	65	4.4 (1.2)	.	65	4.2 (1.2)	.	278	4.1 (1.2)
Ex 20 cm2 patch	66	3.8 (1.3)	0.835	67	3.9 (1.2)	0.378	65	3.9 (1.3)	0.021	62	4.2 (1.3)	0.748	260	3.9 (1.3)
Ex 10 cm2 patch	66	4.0 (1.3)	0.668	72	3.8 (1.1)	0.127	57	3.8 (1.2)	0.001	53	4.0 (1.3)	0.652	248	3.9 (1.2)
Ex Capsule	75	3.8 (1.2)	0.318	75	4.0 (1.3)	0.291	45	3.8 (1.2)	0.013	58	4.1 (1.4)	0.246	253	3.9 (1.3)

Treatment by Age subgroup Interaction test p-value=0.66

Race

In the sponsor's ITT(LOCF) population 76% were Caucasian, 9% were Oriental, and 15% were classified as Other races.

As can be seen in Table 27 there were no clear differences in efficacy between Caucasians and Oriental ethnicities.

In the other race category treatment differences on the ADAS-cog were numerically worse for the Exelon 20 cm2 than placebo and for Exelon capsule than placebo. However, this could be attributable to the small sample size in the other category and the attendant increased variability in the means.

ADAS-Cog ITT(LOCF)

Table 27 Race Subgroups: Mean Change from Baseline to Week 24 or LOCF in ADAS-cog in ITT(LOCF) Population

	CAUCASIAN			ORIENTAL			OTHER			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	208	1.5 (6.9)	.	46	-0.5 (7.8)	.	27	0.1 (3.9)	.	281	1.0 (6.8)
Ex 20 cm2 patch	200	-1.2 (6.1)	0	38	-5.1 (8.0)	0.002	23	1.0 (4.9)	0.658	261	-1.6 (6.3)
Ex 10 cm2 patch	196	-0.4 (6.8)	0.004	33	-2.7 (6.3)	0.139	22	-1.0 (5.1)	0.538	251	-0.8 (6.6)
Ex Capsule	191	-0.7 (6.6)	0.001	39	-2.5 (6.7)	0.115	25	0.6 (5.1)	0.795	255	-0.8 (6.5)

Treatment by Race Interaction test p-value=   0.41

As measured by the CGI, treatment effects compared to placebo did not reach nominal significance level except for the Exelon 10 cm2 patch and the Exelon capsule in Caucasians (see Table 28).

These findings for the comparisons of Exelon 10 cm2 patch and Exelon capsule with placebo could be attributable to the small sample sizes for Oriental and Other ethnicities. The lack of nominally significant effects between Exelon 20 cm2 patch and placebo in the race subgroups is not surprising since overall the effect between Exelon 20 cm2 patch and placebo was not significant. Overall, there is no compelling evidence that treatment effects depend on ethnicity.

Table 28 Race Subgroups: Mean CGI at Week 24 or LOCF in ITT(LOCF) Population by Treatment Group

TREAT	CAUCASIAN			ORIENTAL			OTHER			ALL	
	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	205	4.2 (1.3)	.	27	4.0 (0.8)	.	46	3.9 (1.3)	.	278	4.1 (1.3)
Ex 20 cm2 patch	200	4.1 (1.2)	0.209	23	4.0 (1.0)	0.989	37	3.4 (1.6)	0.104	260	4.0 (1.3)
Ex 10 cm2 patch	193	3.9 (1.2)	0.022	22	3.8 (1.3)	0.326	33	3.9 (1.2)	0.99	248	3.9 (1.2)
Ex Capsule	189	4.0 (1.2)	0.033	25	4.1 (1.3)	0.964	39	3.6 (1.4)	0.251	253	3.9 (1.2)

Treatment by Race Subgroup Interaction test p-value=0.53

4.2 Other Special/Subgroup Populations

There was a slight indication that there were larger treatment effects on the CGI and the change in ADAS-cog in the MMSE < 15 subgroup but a test for interaction between treatment group and MMSE subgroup was not statistically significant.

Table 29 MMSE subgroups: Mean Change from Baseline to Week 24 or LOCF in ADAS-cog in ITT(LOCF) Population

TREAT	MMSE$\geq$15			MMSE<15			ALL	
	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	196	0.1 (6.1)	.	85	3.1 (8.0)	.	281	1.0 (6.8)
Ex 20 cm2 patch	199	-1.9 (5.7)	0.003	62	-0.5 (8.6)	<0.001	261	-1.6 (6.5)
Ex 10 cm2 patch	196	-1.0 (6.3)	0.078	55	0.1 (7.7)	0.011	251	-0.8 (6.7)
Ex Capsule	180	-1.0 (6.3)	0.098	75	-0.4 (7.1)	<0.001	255	-0.8 (6.6)

Treatment by MMSE subgroup Interaction test p-value=0.17

Table 30 MMSE subgroups: Mean CGI at Week 24 or LOCF in ADAS-cog in ITT(LOCF) Population

	MMSE≥15			MMSE<15			ALL	
TREAT	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)	P-value (vs. Placebo)	N	MEAN (SD)
Placebo	196	3.9 (1.2)	.	82	4.7 (1.1)	.	278	4.1 (1.2)
Ex 20 cm2 patch	197	3.9 (1.2)	0.842	63	4.3 (1.4)	0.024	260	4.0 (1.3)
Ex 10 cm2 patch	194	3.8 (1.2)	0.502	54	4.2 (1.3)	0.034	248	3.9 (1.2)
Ex Capsule	178	3.8 (1.2)	0.38	75	4.2 (1.4)	0.001	253	3.9 (1.3)

Treatment by MMSE subgroup Interaction test p-value=0.10

5   SUMMARY AND CONCLUSIONS

5.1   Statistical Issues and Collective Evidence

The sponsor chose a multiplicity adjustment method which required significance of the high dose patch vs. placebo on both primary endpoints before proceeding to the low dose patch vs. placebo comparison or to any secondary endpoints. Since the high dose vs. placebo comparison reached significance on the co-primary ADAS-cog but did not reach statistical significance at 0.05 on the co-primary CGI, technically, the study could be viewed as a failure. The sponsor reported a p-value for the high dose vs. placebo comparison on the CGI of 0.054 for the primary analysis. They reported some secondary analyses of the CGI for which the comparison between the high dose patch and placebo reached the nominal level of significance. Note that the sponsor excluded CGI scores from the primary analysis which were taken more than two days after the last dose as planned in the protocol. However, this reviewer found 16 other cases (8 placebo and 8 high dose patch) where some patient's week 24 CGI was not used in the sponsor's primary ITT(LOCF) analysis. The sponsor claimed that in cases where the last CGI was not used in the LOCF analysis it was because the week 24 CGI score was taken in the open label extension period, which followed the double blind phase and, therefore, in these cases they carried forward the week 16 score in their analysis. However, in all but 2 of these 16 cases the CGI was rated only one day after the end of the double blind period, assuming that the rating was in fact taken in the open label phase. Furthermore, it is not clear from the data that they were from the open label period because the CGI score data set and the dosing record data set each have a variable to indicate if the record is from the extension and they disagree in all of these cases on whether or not the week 24 CGI score was assessed in the double blind period or the open label phase. If in these few cases we use the week 24 CGI scores that were taken one day after the end of the double blind phase in the primary analysis then the p-value for the Exelon 20 cm2 patch vs. placebo comparison increases to 0.109. Therefore, the high dose vs. placebo comparison on the

40

co-primary CGI may not be as close to significance as reported by the sponsor. Note that at the beginning of the open label extension all continuing patients were started with the Exelon 10cm2 patch but their previous double blind treatment was not revealed, i.e., they might have been able to tell the size of the patch they had been on but it was not revealed whether it had contained placebo or Exelon. It is unknown in the 14 cases where the double blind week 24 CGI assessment was reportedly taken in the open label phase, but only on the first day of it, whether or not the assessment was made before the open label drug was administered. It is also unknown whether the drug could have an effect in only 1 day for placebo patients beginning the 10 cm2 patch or for Exelon 20 cm2 patients who had tolerated the Exelon 20 cm2 patch and were switched to the Exelon 10 cm2 patch at the start of the open label phase. This requires clinical judgment. At any rate, if there was considered to be an effect of the high dose patch on the CGI it was very small and the sample size was average or above average (290 to 300 randomized per group and 250 to 280 per group had sufficient data to be included in the ITT(LOCF) analysis). Although the effect size on the CGI was also small for the low dose patch, if a Bonferroni multiplicity adjustment had been chosen the study would have been positive on the basis of the **low dose patch because it's comparison with placebo** reached significance at the 0.025 level for both co-primary endpoints.

If the highest sensitivity analysis p-value was 0.054 and there were no other problems, e.g., if the 16 double blind Week 24 CGI scores possibly taken in the open label period were considered clinically irrelevant, then one might examine whether there were other supporting elements that could lead one to conclude that 0.054 was close enough to 0.050, given the supporting elements. In this case we can consider that the investigational drug is a new formulation of a drug for **which the original formulation is approved for mild to moderate Alzheimer's. Therefore, it may** be able to borrow strength from the original oral capsule formulation. In addition, the low dose patch achieved the nominal significance level on both co-primary endpoints and would have even after a Bonferroni adjustment to the significance level. Furthermore, the p-values for both the high dose patch and the low dose patch comparisons to placebo on the other co-primary endpoint, the change from baseline in ADAS-cog Total score, were more than an order of magnitude smaller than 0.05. Also, both the high dose and the low dose patch achieved nominal significance compared to placebo on two of the secondary endpoints, the change in ADCS-ADL Total score and the change in MMSE. Because of these supporting factors, in the present case 0.054 may be close enough to 0.05 at least for the first gate of the gatekeeping process to be passed to permit testing of the low dose patch.

It is not clear that the high dose patch adds anything over the low dose patch. In particular, while the high dose patch was numerically better than the low dose patch in terms of the change from baseline in ADAS-cog Total score by 0.8, the low dose was numerically better for the CGI and there was essentially no difference between the high dose patch and the low dose patch on the change from baseline in ADCS-ADL Total score or the change from baseline in the MMSE score. Furthermore, the low dose achieved nominal significance levels compared to placebo on both co-primary endpoints while the high dose patch group only strictly achieved significance for one of the two.

5.2 Conclusions and Recommendations

The high dose patch reached statistical significance compared to placebo on the ADAS-cog but it did not quite reach significance on the co-primary CGI. The decision rule was set up to require significance on the high dose patch vs. placebo comparison before testing the low dose patch vs. placebo. Therefore, technically the study failed. However, some other possible multiplicity adjustment methods, such as the Bonferroni method, had they been chosen would have yielded a positive study by means of significance of the low dose patch vs. placebo comparison on both the ADAS-cog and the CGI. Because this is a new formulation of a drug that is approved the evidence for the original oral formulation might also lend some support to the efficacy of the new patch formulation in this close case.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/
- - - - - - - - - - - - - - - - - - - -
Tristan Massie
6/18/2007 10:57:39 AM
BIOMETRICS

Kooros Mahjoob
6/18/2007 11:09:19 AM
BIOMETRICS

Appears This Way
On Original

U.S. Department of Health and Human Services
Food and Drug Administration

Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDIES

NDA:	22-083
Drug Name:	Exelon® Transdermal Delivery System Patch
Applicant:	Novartis
Biometrics Division:	Biometrics Division VI
Statistical Reviewer:	Ling Chen, Ph.D.
Concurring Reviewers:	Karl Lin, Ph.D.
Medical Division:	
Pharmacologist:	David Hawver, Ph.D.
Regulatory Manager	Melina Griffis
Keywords:	NDA review, carcinogenicity, Dose response

Appears This Way
On Original

Table of Contents

Appears This Way
On Original

1. Background

Since NDA 022-083 is for a new formulation of an approved product, in this submission the sponsor included reports of only one carcinogenicity study in mice.

The study was to evaluate the carcinogenicity potential of the test substance, SDZ ENA 713 base, following the daily dermal administration to mice. The test substance is a selective acetylcholine esterase brain inhibitor. The duration of the study was 98-99 weeks.

2. Mouse Study

Two separate experiments, one in males and one in females were conducted. In each of these two experiments there were three treated and two control groups. The dose-levels were specified by the Sponsor based on a 13-week dose finding study with dose-levels up to 1.6 mg/kg/day. Due to exaggerated cholinergic effects and mortality, the dose-levels of 0.25, 0.5 and 0.75 mg/kg/day were chosen. Control groups are untreated and treated with a vehicle.

The dermal route was chosen since it is the expected route of clinical use. Two hundred and fifty ⎯⎯CD-1®(ICR) BR stain mice of each sex were randomly allocated to treated groups and control groups in equal size of 50 animals. In this review the test treatment groups will be **b(4)** termed as the Low, Medium, and High dose groups, respectively. The control received the vehicle, ethyl alcohol absolute – USP, batch supplied by ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯ ⎯⎯⎯⎯. Each animal was given the test substance formalations or the vehicle once a day, at approximately the same time, seven days a week over a period of 98-99 weeks (682 to 690 days specifying days for terminal sacrifice).

Each animal was checked at least twice a day, including during weekends and public holidays, for mortality or signs of morbidity. All animals showing signs of poor clinical conditions were humanely killed. On completion of the treatment period (week 98 or 99), after at least 14 hours fasting, all surviving animals were killed by carbon dioxide asphyxiation and exsanguination. The survival rates before final sacrifice exceeded 50% in all treatment groups.

A complete macroscopic examination was performed on all animals, including any that died during the study or were killed prematurely. All macroscopic observations were recorded individually.

The body weight of each animal was recorded once before allocation of the animals to groups, on the first day of treatment, and then once a week until the end of the study.

2.1 Sponsor's analysis

2.1.1 Survival analysis

Sponsor did not perform any formal statistical analysis of the mortality data. As summary statistics the sponsor reported that for both sexes the survival rates after weeks 52, 78 or 99 were similar or were slightly higher in vehicle control group 2. See Table 1 below.

Table 1: Survival rates (expressed in %)

Deinogest (mg/kg/day) Weeks	Male					Female				
	CD 1 0	CD 2 0	LD 0.25	MD 0.5	HD 0.75	CD 1 0	CD 2 0	LD 0.25	MD 0.5	HD 0.75
0 - 52	96	98	90	96	96	92	96	92	96	96
53-78	86	90	84	88	78	78	82	84	80	82
79-99	48	58	66	50	52	50	70	54	58	68

Note: CD 1 denotes untreated while CD 2 denotes Vehicle.

The sponsor made the following statement regarding the mortality data:

For test-treated male and female groups, the survival rates after weeks 52, 78 or 99 as well as the mean duration of treatment were similar to that of control groups; the few differences recorded between the groups were slight, neither dose-related, nor statistically significant and did not show a similar trend in two sexes. Consequently, the survival rates of the animals were not affected by the treatment with the test substance; the slight differences noted similarly in both control and test treated groups were considered to be of spontaneous occurrence.

2.1.2 Tumor data analysis

Sponsor submitted a tumor data set of the study. However, no statistical test for positive linear dose-tumor trends was reported. In conclusion Sponsor stated that the test substance did not show a carcinogenic potential or any effect on the incidence of spontaneously occurring tumors at any dose-level.

2.2   Reviewer's analysis

To verify the results of sponsor's analyses this reviewer independently performed survival and tumor data analyses. Data used in this reviewer's analyses were provided by the sponsor electronically. The link to the data set is

\\Cdsesub1\n22083\N_000\2006-09-08\crt\datasets\tumor data\15151\ TUMOR.xpt.

2.2.1   Survival analysis

The summaries of the intercurrent mortality data are given in Tables 1A and 1B for males and females, respectively. Since the termination sacrifices were done in week 98 or 99, the time intervals 0-52, 53-78 and 79-97 weeks were chosen. Note that the time to terminal sacrifice of male mouse Q12052 in low dose treatment group was 583 days. This record is different from the time to terminal sacrifice of week 98-99 in the report. This reviewer treated the death of mouse Q12052 as natural death or moribund sacrifice in her analysis.

From Tables 1A and 1B, it can be seen that number of deaths in treated groups within each time period are similar to those in the untreated and vehicle groups, except the numbers in male mice with high dose and in female mice with low dose during weeks 53-78. For male mice with high dose, the number of deaths was 9 during weeks 53-78, comparing to 5, 4, 3 and 4 in untreated, vehicle, low dose and medium dose respectively. For female mice with low dose, 4 mice died during weeks 53-78 which is approximately 50% lower than the numbers of death in the other groups.

For male mice the survival rates before terminal sacrifices (weeks 98-99) are 54%, 58%, 68%, 58% and 56% for untreated, vehicle, low dose, medium dose and high dose respectively. For female mice, those rates are 54%, 70%, 56%, 60% and 68% for untreated, vehicle, low dose, medium dose and high dose respectively. It is easy to see that the survival rates of female mice are higher than those of male mice in the end of the study, especially in the vehicle control and the high dose group. The differences in the survival rates are 12%.

The survival rates were also estimated using the Kaplan-Meier product limit method. The Kaplan-Meier curves for survival rate are given in Figures 1A and 1B for males and females, respectively. It can be noticed that for male mice, the survival curve of low dose group is lower than those of other groups during middle period of the study (35 to 65 weeks), and very few deaths occurred in the period from 55 to 90 weeks.

The homogeneity of survival distributions of three treatment groups and untreated control (or vehicle control) was tested separately for males and females using the Cox test (Cox, 1972) and the Kruskal-Wallis test (Gehan, 1965; Thomas, *al*, 1977). Results of the tests are given in

NDA22-083 CIT/Study No. 1514 TCS/SDZ ENA 713 base

Tables 2A and 2B (or Tables 3A and 3B) for males and females, respectively. The tests showed no statistically significant differences in survivals across treatment groups in female mice. However, the Cox tests showed statistically significant results in male mice (with untreated control: p=0.0149, and with vehicle control: p=0.0265).

It can be seen from Figure 1A that for male mice, the survival curve of low dose group is lower than those of the other groups in middle period of the study (approximately 35 days to 65 days), and very few deaths occurred in the period from 55 days to 90 days. These differences between low dose and other groups may be the reason for the significant results of the Cox test for homogeneity.

The homogeneity tests were performed on treatment groups and control group excluding the low dose group for male mice. Results from both the Cox test and the Kruskal-Wallis test were very insignificant with p-values 0.8618 and 0.8993 respectively.

2.2.2 Tumor data analysis

The positive dose response analysis was performed using the Peto test (1980). The actual dose levels of treatment groups were used as the weights for the trend analysis. The tumor rates and the p-values of the tumor types tested for dose response relationship are listed in Table 3A with untreated control (or Table 4A with vehicle control), and Table 3B (or Table 4B) for males and females, respectively. The p-values reflect one-sided tests for increases in tumors **with dose. Per Pharm/Tox reviewer Dr. Hawver's request, some combinations were done for** some tumor types within organs or across organs in the trend analysis (see negative tumor codes in these tables).

Adjustment for multiplicity for the trend testing was done using a significance level of α=0.05 for rare tumors and α=0.01 for common tumors. A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. Adjustment for multiple pairwise comparisons was done using a significance level α=0.05 for rare tumors and α=0.01 for common tumors. Since there is only one species in this study, the adjustment designed for two species is not needed.

The trend test was performed separately for different controls. Notice that significant results of any tumor type observed with either control in the trend test would be considered as having a significant result for this tumor type. Therefore, adjustment for the significance level is not needed for performing trend tests separately for the two controls.

It can be seen from the Tables 3A and 3B, and Tables 4A and 4B that all p-values exceed 0.1 regardless methods of statistical test (either exact or asymptotic). No significant results were found.

2.2.3 Reviewer's findings

The reviewer's analyses did not show statistically significant dose response in the incidence in any of the tested tumor types in either sex with respect to vehicle control or untreated control. Also none of the pairwise comparisons of each treated group with either vehicle control or untreated control was considered to be statistically significant in either sex.

2. Summary

In this submission the sponsor included reports of one animal carcinogenicity study in mice. The study was intended to assess the carcinogenicity potential of the test substance, SDZ ENA 713 base, following daily dermal administration in mice. The test substance is a selective acetylcholine esterase brain inhibitor. The duration of the study was 98-99 weeks.

In this review, the phrase "dose response" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor rate as dose increases.

This study was conducted in CD-1® strain mice. The dermal route was used. The study had 5 treatment groups namely, untreated control, vehicle control, 0.25, 0.5, and 0.75 mg/kg/day. Tests showed no statistically significant differences in survivals across treatment groups in female mice. But the Cox test showed statistically significant differences in survivals across treatments in male mice. It appears that the statistically significant difference was due to the lower mortality in low dose group. Tests showed no statistically significant dose response in the incidence of any tested tumor types (individual tumor types and combinations of tumor types suggested by the reviewing pharmacologist) with respect to the untreated control or the vehicle control. In pairwise comparisons none of the tested tumor types showed statistically significant increased incidence in the high dose group compared to the untreated control or the vehicle control in the tested tumor types.

Although the study results supported the Sponsor's conclusion that the test substance did not show a carcinogenic potential or any effect on the incidence of spontaneously occurring tumors at any dose-level, it is important that the pharm/tox reviewer evaluates the appropriateness of the doses used in the study to see if the high dose is close to MTD and presented enough tumor challenge to the tested animals.

Ling Chen, Ph.D.
Mathematical Statistician

Concur: Karl Lin, Ph.D.
Team leader, DB6/OB

Appendix

Table 1A: Analysis of Mortality Data for Male Mice
by Treatment and Time

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
Untreated 0 mg/kg/day	0-52	50	2	48	96.0	4.0
	53-78	48	5	43	86.0	14.0
	79-97	43	16	27	54.0	46.0
	FINALKILL 98-99	27	27 (3)	0		
Vehicle 0 mg/kg/day	0-52	50	1	49	98.0	2.0
	53-78	49	4	45	90.0	10.0
	79-97	45	16	29	58.0	42.0
	FINALKILL 98-99	29	29 (0)	0		
Low Dose 0.25mg/kg/day	0-52	50	5	45	90.0	10.0
	53-78	45	3	42	84.0	16.0
	79-97	42	8	34	68.0	32.0
	FINALKILL 98-99	34	34 (1)	0		
Median Dose 0.5 mg/kg/day	0-52	50	2	48	96.0	4.0
	53-78	48	4	44	88.0	12.0
	79-97	44	15	29	58.0	42.0
	FINALKILL 98-99	29	29 (4)	0		
High Dose 0.75 mg/kg/day	0-52	50	2	48	96.0	4.0
	53-78	48	9	39	78.0	22.0
	79-97	39	11	28	56.0	44.0
	FINALKILL 98-99	28	28 (2)	0		

Note: The number in the parentheses of column "No. Died" is the number of animals died due to natural death or moribund in terminal sacrifice weeks 98 and 99.

Table 2A: Intercurrent Mortality Comparison Male Mice
(Untreated control)

	Method			
Test	**Cox**		**Kruskal-Wallis**	
	Statistics	**P-Value**	**Statistics**	**P-Value**
Dose-Mortality Trend	0.3737	0.5410	0.4179	0.5180
Homogeneity	10.4798	0.0149	7.4056	0.0600

Table 3A: Intercurrent Mortality Comparison Male Mice
(Vehicle control)

	Method			
Test	**Cox**		**Kruskal-Wallis**	
	Statistics	**P-Value**	**Statistics**	**P-Value**
Dose-Mortality Trend	2.0881	0.1485	1.7403	0.1871
Homogeneity	9.2214	0.0265	6.6167	0.0852

Appears This Way
On Original

Table 4A: Report on Test for Positive Linear Dose-Tumor Trends in Male Mice (Untreated Control)

Organ Code	Organ Name	Tumor Code	Tumor Name	P-Value (Exact Method)	P-Value (Asymptotic Method)
0000	WHOLE BODY	-95	All HEMANGIOMA/HEMANGIOSARCOMA	0.4910	0.4930
0900	LUNGS	-91	ALVEOLAR/BRONCHIOLAR ADENOMA /MULT	0.3102	0.3099
0900	LUNGS	-92	ALVEOLAR/BRONCHIOLAR CARCINOMA/MULT	0.6814	0.6820
0900	LUNGS	-93	ALVEOLAR/BRONCHIOLAR CARCINOMA/ADENOMA	0.4893	0.4893
0900	LUNGS	090001	ALVEOLAR/BRONCHIOLAR ADENOMA	0.3026	0.3023
0900	LUNGS	090004	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.5667	0.5670
0900	LUNGS	090025	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.8455	0.8512
0900	LUNGS	090030	ALVEOLAR/BRONCHIOLAR ADENOMA,	0.5660	0.5684
1506	FORESTOMACH	150606	SQUAMOUS CELL CARCINOMA	0.4914	0.5032
1800	LIVER	-94	ALVEOLAR/BRONCHIOLAR CARCINOMA/MULT	0.8836	0.8841
1800	LIVER	180006	HEPATOCELLULAR CARCINOMA	0.3900	0.3865
1800	LIVER	180020	HEPATOCELLULAR ADENOMA, MULTIP	0.9804	0.9775
1800	LIVER	180021	HEPATOCELLULAR ADENOMA	0.6546	0.6560
1800	LIVER	180037	HEPATOCELLULAR CARCINOMA, MULT	0.5031	0.4967
1800	LIVER	180039	MULTIPLE HEMANGIOMA	0.2373	0.1768
1800	LIVER	590107	HEMANGIOSARCOMA	0.4902	0.4941
1900	GALL BLADDER	540001	ADENOMA	0.4985	0.5000
2000	PANCREAS	200019	ISLET CELL ADENOMA	0.7712	0.8191
2500	TESTES	250001	INTERSTITIAL CELL ADENOMA	0.8954	0.8963
2500	TESTES	250007	HEMANGIOMA	0.2373	0.1768
4200	THYROID	420006	FOLLICULAR CELL ADENOMA	1.0000	0.9665
4400	ADRENAL	440004	CORTICAL ADENOMA	0.4286	0.2857
4400	ADRENAL	440013	SUBCAPSULAR CELL ADENOMA	0.6215	0.6238
4500	SYSTEMIC	450003	HISTIOCYTIC SARCOMA	0.1488	0.1434
4500	SYSTEMIC	450006	MALIGNANT LYMPHOMA	0.7570	0.7587
4600	SPLEEN	590107	HEMANGIOSARCOMA	0.7980	0.8137
5000	THYMUS	500013	FOLLICULAR CELL CARCINOMA/ECTO	0.7719	0.8196
5400	HARDERIAN GLAND	540001	ADENOMA	0.8449	0.8471
5700	SKIN	250007	HEMANGIOMA	0.5200	0.4723
5901	FEMUR	590107	HEMANGIOSARCOMA	0.5200	0.4723
5902	STERNUM	590107	HEMANGIOSARCOMA	0.5200	0.4723

Appears This Way
On Original

Table 5A: Report on Test for Positive Linear Dose-Tumor Trends in Male Mice
(Vehicle control)

Organ Code	Organ Name	Tumor Code	Tumor Name	P-Value (Exact Method)	P-Value (Asymptotic Method)
0000	WHOLE BODY	-95	All HEMANGIOMA/HEMANGIOSARCOMA	0.4785	0.4806
0900	LUNGS	-91	ALVEOLAR/BRONCHIOLAR ADENOMA /MULT	0.1944	0.1938
0900	LUNGS	-92	ALVEOLAR/BRONCHIOLAR CARCINOMA/MULT	0.3788	0.3781
0900	LUNGS	-93	ALVEOLAR/BRONCHIOLAR CARCINOMA/ADENOMA	0.2389	0.2385
0900	LUNGS	090001	ALVEOLAR/BRONCHIOLAR ADENOMA	0.2713	0.2705
0900	LUNGS	090004	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.2598	0.2587
0900	LUNGS	090025	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.8390	0.8448
0900	LUNGS	090030	ALVEOLAR/BRONCHIOLAR ADENOMA,	0.3137	0.3102
1506	FORESTOMACH	150606	SQUAMOUS CELL CARCINOMA	0.4831	0.4938
1800	LIVER	-94	ALVEOLAR/BRONCHIOLAR CARCINOMA/MULT	0.7817	0.7830
1800	LIVER	180006	HEPATOCELLULAR CARCINOMA	0.5104	0.5117
1800	LIVER	180020	HEPATOCELLULAR ADENOMA, MULTIP	0.8050	0.8237
1800	LIVER	180021	HEPATOCELLULAR ADENOMA	0.7268	0.7284
1800	LIVER	180037	HEPATOCELLULAR CARCINOMA, MULT	0.8083	0.8245
1800	LIVER	180039	MULTIPLE HEMANGIOMA	0.2333	0.1727
1800	LIVER	250007	HEMANGIOMA	1.0000	0.9524
1800	LIVER	590107	HEMANGIOSARCOMA	0.4870	0.4907
1900	GALL BLADDER	540001	ADENOMA	0.9818	0.9798
2000	PANCREAS	200019	ISLET CELL ADENOMA	0.7583	0.8113
2500	TESTES	250001	INTERSTITIAL CELL ADENOMA	0.7859	0.7887
2500	TESTES	250007	HEMANGIOMA	0.2333	0.1727
4100	PITUITARY	410009	ADENOMA/PARS INTERMEDIA	1.0000	0.9630
4400	ADRENAL	440004	CORTICAL ADENOMA	0.4500	0.3054
4400	ADRENAL	440013	SUBCAPSULAR CELL ADENOMA	0.2333	0.1727
4500	SYSTEMIC	450003	HISTIOCYTIC SARCOMA	0.1443	0.1391
4500	SYSTEMIC	450006	MALIGNANT LYMPHOMA	0.4894	0.4878
4600	SPLEEN	590107	HEMANGIOSARCOMA	0.7933	0.8087
5000	THYMUS	500013	FOLLICULAR CELL CARCINOMA/ECTO	0.7586	0.8116
5400	HARDERIAN GLAND	540001	ADENOMA	0.9027	0.9038
5700	SKIN	250007	HEMANGIOMA	0.5200	0.4723
5700	SKIN	570020	OSTEOSARCOMA	1.0000	0.9626
5901	FEMUR	590107	HEMANGIOSARCOMA	0.5200	0.4723
5902	STERNUM	590107	HEMANGIOSARCOMA	0.5200	0.4723

Appears This Way
On Original

Table 1B: Analysis of Mortality Data for Female Mice
by Treatment and Time

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
Untreated **0 mg/kg/day**	0-52	50	4	46	92.0	8.0
	53-78	46	7	39	78.0	22.0
	79-97	39	12	27	54.0	46.0
	FINALKILL 98-99	27	27 (2)	0		
Vehicle **0 mg/kg/day**	0-52	50	2	48	96.0	4.0
	53-78	48	7	41	82.0	18.0
	79-97	41	6	35	70.0	30.0
	FINALKILL 98-99	35	35 (0)	0		
Low Dose **0.25mg/kg/day**	0-52	50	4	46	92.0	8.0
	53-78	46	4	42	84.0	16.0
	79-97	42	14	28	56.0	44.0
	FINALKILL 98-99	28	28 (1)	0		
Median Dose **0.5 mg/kg/day**	0-52	50	2	48	96.0	4.0
	53-78	48	8	40	80.0	20.0
	79-97	40	10	30	60.0	40.0
	FINALKILL 98-99	30	30 (1)	0		
High Dose **0.75 mg/kg/day**	0-52	50	2	48	96.0	4.0
	53-78	48	7	41	82.0	18.0
	79-97	41	7	34	68.0	32.0
	FINALKILL 98-99	34	34 (0)	0		

Note: The number in the parentheses of column "No. Died" is the number of animals died due to natural death or moribund in terminal sacrifice weeks 98 and 99.

Appears This Way
On Original

**Table 2B: Intercurrent Mortality Comparison Female Mice
(Untreated control)**

	Method			
Test	**Cox**		**Kruskal-Wallis**	
	Statistics	**P-Value**	**Statistics**	**P-Value**
Dose-Mortality Trend	1.9890	0.1584	1.7502	0.1859
Homogeneity	2.0583	0.5604	1.8648	0.6009

**Table 3B: Intercurrent Mortality Comparison Female Mice
(Vehicle control)**

	Method			
Test	**Cox**		**Kruskal-Wallis**	
	Statistics	**P-Value**	**Statistics**	**P-Value**
Dose-Mortality Trend	0.0090	0.9243	0.0160	0.8995
Homogeneity	1.9507	0.5827	1.3114	0.7264

Appears This Way
On Original

Table 4B: Report on Test for Positive Linear Dose-Tumor Trends in Female Mice

(Untreated Control)

Organ Code	Organ Name	Tumor Code	Tumor Name	P-Value (Exact Method)	P-Value (Asymptotic Method)
0000	ORAL CAVITY/FORESTOMACH	-100	SQUAMOUS CELL CARCINOMA/PAPILLOMA	0.1540	0.1447
0001	UTERUS/CERVIX	-101	ENDOMETRIAL STROMAL POLYP/MULT	0.7843	0.7851
0002	UTERUS/CERVIX	-102	LEIOMYOMA/COMBINE	0.9164	0.9176
0003	WHOLE BODY	-103	HEMANGIOSARCOMA/HEMANGIOMA	0.2148	0.2101
0100	BRAIN	010008	MENINGIOMA	1.0000	0.9669
0900	LUNGS	-91	ALVEOLAR/BRONCHIOLAR ADENOMA/MULT	0.8571	0.8580
0900	LUNGS	-92	ALVEOLAR/BRONCHIOLAR CARCINOMA/MULT"	0.8305	0.8319
0900	LUNGS	-93	ALVEOLAR/BRONCHIOLAR CARCINOMA/ADENOMA	0.8790	0.8795
0900	LUNGS	090001	ALVEOLAR/BRONCHIOLAR ADENOMA	0.7521	0.7532
0900	LUNGS	090004	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.5764	0.5768
0900	LUNGS	090025	ALVEOLAR/BRONCHIOLAR CARCINOMA	1.0000	0.9892
0900	LUNGS	090030	ALVEOLAR/BRONCHIOLAR ADENOMA,	0.8749	0.8791
1000	ORAL CAVITY	570021	SQUAMOUS CELL CARCINOMA	0.6667	0.4298
1506	FORESTOMACH	570022	SQUAMOUS CELL PAPILLOMA	0.5424	0.5359
1800	LIVER	180021	HEPATOCELLULAR ADENOMA	0.5974	0.5968
1800	LIVER	180038	MULTIPLE HEMANGIOSARCOMA	0.2556	0.1914
1800	LIVER	180039	MULTIPLE HEMANGIOMA	0.7209	0.7728
1900	GALL BLADDER	540001	ADENOMA	0.2857	0.2110
3200	OVARIES	320012	GRANULOSA CELL TUMOR	0.7593	0.8186
3200	OVARIES	320016	ADENOMA, TUBULOSTROMAL	0.5378	0.5342
3200	OVARIES	320020	LUTEOMA	0.8423	0.8574
3400	UTERUS	-94	ENDOMETRIAL STROMAL POLYP/MULT	0.8868	0.8875
3400	UTERUS	-95	ENDOMETRIAL STROMAL POLYP/SARCOMA	0.8763	0.8770
3400	UTERUS	340003	ENDOMETRIAL STROMAL POLYP	0.8108	0.8120
3400	UTERUS	340011	HEMANGIOSARCOMA	0.3549	0.3432
3400	UTERUS	340013	ENDOMETRIAL STROMAL SARCOMA	0.4001	0.3948
3400	UTERUS	340017	LEIOMYOSARCOMA	0.6731	0.6692
3400	UTERUS	340019	ENDOMETRIAL STROMAL POLYP, MUL	0.8735	0.8781
3400	UTERUS	340020	LEIOMYOMA	1.0000	0.9901
3400	UTERUS	340025	ADENOCARCINOMA	0.7731	0.8351
3600	CERVIX	-96	ENDOMETRIAL STROMAL POLYP/SARCOMA	0.2337	0.2305
3600	CERVIX	340003	ENDOMETRIAL STROMAL POLYP	0.3375	0.3335
3600	CERVIX	340013	ENDOMETRIAL STROMAL SARCOMA	0.3678	0.3551
3600	CERVIX	340017	LEIOMYOSARCOMA	0.4865	0.4766
3600	CERVIX	340020	LEIOMYOMA	0.4348	0.4315
4100	PITUITARY	-97	ADENOMA/PARS DISTALIS/COMBINE	0.2857	0.2110
4100	PITUITARY	410006	ADENOMA/PARS DISTALIS	0.9992	0.9977
4100	PITUITARY	410009	ADENOMA/PARS INTERMEDIA	0.2857	0.2110
4100	PITUITARY	410012	MULTIPLE ADENOMA/PARS DISTALIS	0.5378	0.5342
4400	ADRENAL	440018	PHEOCHROMOCYTOMA, BENIGN	0.5378	0.5342
4500	SYSTEMIC	450003	HISTIOCYTIC SARCOMA	0.9264	0.9275
4500	SYSTEMIC	450006	MALIGNANT LYMPHOMA	0.2769	0.2751
4600	SPLEEN	340011	HEMANGIOSARCOMA	0.5378	0.5342

5400	HARDERIAN GLAND	540001	ADENOMA	0.7758	0.7771
5600	MAMMARY GL	-98	ADENOCARCINOMA/MULT	0.8421	0.8473
5600	MAMMARY GL	340025	ADENOCARCINOMA	1.0000	0.9861
5600	MAMMARY GL	560007	MULTIPLE ADENOCARCINOMA	0.2545	0.1889
5700	SKIN	-99	SQUAMOUS CELL CARCINOMA/PAPILLOMA	0.5378	0.5342
5700	SKIN	570018	MALIGNANT FIBROUS HISTIOCYTOMA	0.6630	0.6609
5700	SKIN	570019	SARCOMA, NOT OTHERWISE SPECIFI	0.7786	0.8304
5700	SKIN	570020	OSTEOSARCOMA	0.9398	0.9453
5700	SKIN	570022	SQUAMOUS CELL PAPILLOMA	0.5378	0.5342

Appears This Way
On Original

Table 5B: Report on Test for Positive Linear Dose-Tumor Trends in Female Mice

(Vehicle Control)

Organ Code	Organ Name	Tumor Code	Tumor Name	P-Value (Exact Method)	P-Value (Asymptotic Method)
0000	ORAL CAVITY/FORESTOMACH	-100	SQUAMOUS CELL CARCINOMA/PAPILLOMA	0.1677	0.1579
0001	UTERUS/CERVIX	-101	ENDOMETRIAL STROMAL POLYP/MULT	0.8835	0.8839
0002	UTERUS/CERVIX	-102	LEIOMYOMA/COMBINE	0.6257	0.6283
0003	WHOLE BODY	-103	HEMANGIOSARCOMA/HEMANGIOMA	0.4249	0.4235
0900	LUNGS	-91	ALVEOLAR/BRONCHIOLAR ADENOMA/M	0.9220	0.9221
0900	LUNGS	-92	ALVEOLAR/BRONCHIOLAR ADENOMA/MULT	0.8799	0.8809
0900	LUNGS	-93	ALVEOLAR/BRONCHIOLAR CARCINOMA/ADENOMA	0.9711	0.9708
0900	LUNGS	090001	ALVEOLAR/BRONCHIOLAR ADENOMA	0.8649	0.8658
0900	LUNGS	090004	ALVEOLAR/BRONCHIOLAR CARCINOMA	0.8799	0.8809
0900	LUNGS	090030	ALVEOLAR/BRONCHIOLAR ADENOMA,	0.8637	0.8682
1000	ORAL CAVITY	570021	SQUAMOUS CELL CARCINOMA	0.5000	0.2819
1506	FORESTOMACH	570022	SQUAMOUS CELL PAPILLOMA	0.5079	0.5000
1800	LIVER	180021	HEPATOCELLULAR ADENOMA	0.9420	0.9418
1800	LIVER	180038	MULTIPLE HEMANGIOSARCOMA	0.2514	0.1870
1800	LIVER	180039	MULTIPLE HEMANGIOMA	0.8378	0.8439
1900	GALL BLADDER	540001	ADENOMA	0.2677	0.1926
3200	OVARIES	320012	GRANULOSA CELL TUMOR	0.7546	0.8158
3200	OVARIES	320016	ADENOMA, TUBULOSTROMAL	0.5039	0.4986
3200	OVARIES	320020	LUTEOMA	0.9304	0.9337
3400	UTERUS	-94	ENDOMETRIAL STROMAL POLYP/MULT	0.9107	0.9109
3400	UTERUS	-95	ENDOMETRIAL STROMAL POLYP/SARCOMA	0.9339	0.9339
3400	UTERUS	340003	ENDOMETRIAL STROMAL POLYP	0.9210	0.9212
3400	UTERUS	340011	HEMANGIOSARCOMA	0.5838	0.5858
3400	UTERUS	340013	ENDOMETRIAL STROMAL SARCOMA	0.5713	0.5725
3400	UTERUS	340017	LEIOMYOSARCOMA	0.2677	0.1926
3400	UTERUS	340019	ENDOMETRIAL STROMAL POLYP, MUL	0.6247	0.6187
3400	UTERUS	340020	LEIOMYOMA	1.0000	0.9579
3400	UTERUS	340025	ADENOCARCINOMA	0.7244	0.8055
3600	CERVIX	-96	ENDOMETRIAL STROMAL POLYP/SARCOMA	0.7044	0.7067
3600	CERVIX	340003	ENDOMETRIAL STROMAL POLYP	0.7434	0.7474
3600	CERVIX	340013	ENDOMETRIAL STROMAL SARCOMA	0.5794	0.5815
3600	CERVIX	340017	LEIOMYOSARCOMA	0.5143	0.5050
3600	CERVIX	340020	LEIOMYOMA	0.4069	0.4043
4100	PITUITARY	-97	ADENOMA/PARS DISTALIS/COMBINE	0.2698	0.1948
4100	PITUITARY	410006	ADENOMA/PARS DISTALIS	0.9288	0.9393
4100	PITUITARY	410009	ADENOMA/PARS INTERMEDIA	0.2698	0.1948
4100	PITUITARY	410012	MULTIPLE ADENOMA/PARS DISTALIS	0.5079	0.5027
4400	ADRENAL	440018	PHEOCHROMOCYTOMA, BENIGN	0.5039	0.4986
4500	SYSTEMIC	450003	HISTIOCYTIC SARCOMA	0.9059	0.9073
4500	SYSTEMIC	450006	MALIGNANT LYMPHOMA	0.6382	0.6389
4600	SPLEEN	340011	HEMANGIOSARCOMA	0.5039	0.4986

5400	HARDERIAN GLAND	540001	ADENOMA	0.5458	0.5463
5600	MAMMARY GL	-98	ADENOCARCINOMA/MULT	0.2500	0.1841
5600	MAMMARY GL	560007	MULTIPLE ADENOCARCINOMA	0.2500	0.1841
5700	SKIN	-99	SQUAMOUS CELL CARCINOMA/PAPILLOMA	0.8031	0.8202
5700	SKIN	570018	MALIGNANT FIBROUS HISTIOCYTOMA	0.8372	0.8421
5700	SKIN	570019	SARCOMA, NOT OTHERWISE SPECIFI	0.7445	0.8098
5700	SKIN	570020	OSTEOSARCOMA	0.7461	0.8133
5700	SKIN	570021	SQUAMOUS CELL CARCINOMA	1.0000	0.9579
5700	SKIN	570022	SQUAMOUS CELL PAPILLOMA	0.5039	0.4986

Appears This Way
On Original

Figure 1A: Kaplan-Meier Survival Functions for Male Mice
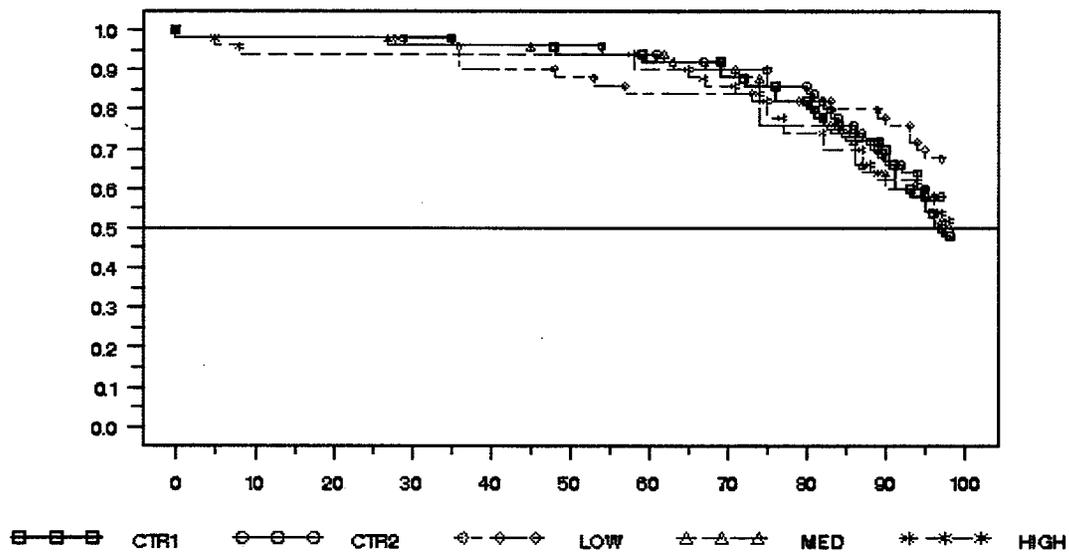
Figure 1B: Kaplan-Meier Survival Functions for Female Mice
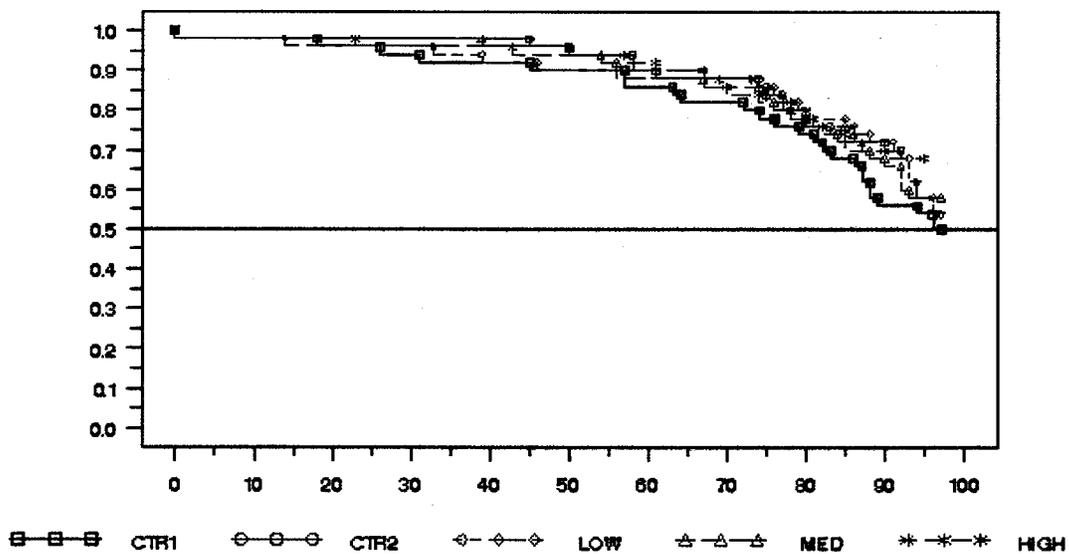
References

Cox D. R. "Regression models and life tables", *Journal of the Royal Statistical Society*, B, 34, 187-220, 1972.

Gehan "A generalized Wilcoxon test for comparing arbitrarily singly censored samples", *Biometrika*, 52, 203-223, 1965.

Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf, "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, Annex to supplement, World Health Organization, Geneva, 311-426, 1980.

Thomas, Donald G., Breslow Norman, and Gart, John J. "Trend and Homogeneity analyses of Propertions and life Tale Data", *Computers and Biomedical Research*, 10, 373-381, 1977.

U.S. Department of Health and Human services, "Guidance for Industry: Statistical Aspects of the Design, Analysis and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals", Center for Drug Evaluation and Research (CDER), Food and Drug Administration, 2001.

Appears This Way
On Original

Appears This Way
On Original