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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	20-823
DRUG NAME:	Exelon ® (rivastigmine tartrate)
INDICATION:	Mild to moderate dementia associated with Parkinson's disease
APPLICANT:	Novartis
DATE OF RECEIPT:	Date of Document: 08/31/2005
REVIEW PRIORITY:	Standard
BIOMETRICS DIVISION:	Division of Biometrics I & VI
STATISTICAL REVIEWER:	Joanne Zhang, Ph.D. (HFD-705)
CONCURRENT REVIEWERS:	Kun Jin, Ph.D., team Leader Kooros Mahjoob, Ph.D., Deputy Director of Division of Biometrics I (HFD-710)
MEDICAL DIVISION:	Division of Neuropharm
CLINICAL TEAM:	Ranjit Mani, M.D., Clinical Reviewer
PROJECT MANAGER:	Melina Griffis, P.Ph.

Table of Contents

List of Tables

List of Figures

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Exelon ® (rivastigmine) was approved by the US Food and Drug Administration (FDA) on April 21, 2000 for the treatment of mild to moderate dementia of Alzheimer's type. The indication of this supplement NDA (the core study 2311 and its extension study 2311 E1) is the use of Exelon (3-12 mg/day) for the treatment of mild to moderate dementia associated with Parkinson's disease (PDD), for which no approved pharmacologic treatment is currently available. It is not totally unexpected a drug that is effective for Alzheimer's disease should work for PD related dementia as well. The core efficacy trial, study 2311, supported the efficacy of Exelon (3-12 mg/day) in the treatment of PDD. The extension of the core efficacy trial, 2311 E1 continuously demonstrated long-term effectiveness of Exelon in PDD patients.

1.2 Brief Overview of Clinical Studies

The submission of this sNDA consisted of one randomized controlled efficacy study 2311, one uncontrolled extension study 2311 E1 and one non-interventional study 2314.

Study 2311 was a 24-week, prospective, randomized, multi-center, double-blind, placebo-controlled study in patients with a clinical diagnosis of Parkinson's disease according to DSM-IV criteria. The study was designed to evaluate the efficacy, safety, and tolerability of Exelon at doses of 3 to 12 mg/day in this patient population. There were 68 centers in Europe and Canada from 12 countries. The 12 countries are Austria (1 center), Belgium (4 centers), France (9 centers), Germany (12 centers), Italy (11 centers), Netherlands (2 centers), Norway (1 center), Portugal (1 center), Spain (8 centers), Turkey (3 centers), United Kingdom (9 centers) and Canada (7 centers). A total of 541 patients with PDD were to be randomly assigned to treatment with either Exelon 3-12 mg/day or placebo in a 2:1 ratio of the drug and placebo.

There were 4 dose levels for Exelon, dose level 1 – Exelon 1.5 mg; dose level 2 – Exelon 3.0 mg; dose level 3 – Exelon 4.5 mg and dose level 4 - Exelon 6.0 mg. Exelon and placebo capsules were identical appearance. All patients were started on dose 1.5 mg or placebo, with increases to the next dose level after a minimum of 4 weeks. Dosage could be reduced to the next lower dose in case of tolerability problems and then increased again by one dose level. After finding the highest well-tolerated dose for each individual patient within the 16 week titration period, the highest well-tolerated dose for each individual patient was then to be maintained for the remaining 8 weeks, although dose adjustments were allowed at any time during this maintenance period. Throughout this report, Exelon 3-12mg/day refers to the above described flexible titration dosing scheme.

The primary endpoints were the "Alzheimer's Disease Assessment Scale-cognitive subscale" (ADAS-cog) and the "Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change" (ADCS-CGIC). The primary analysis for ADAS-cog was ANCOVA and the primary analysis for ADCS-CGIC was the nonparametric categorical analysis using country as blocking – Van Elteren test. The primary population proposed by the sponsor for comparing the treatment groups was the ITT+RDO population. This population was the intent to treat including patients

who discontinued study treatment early but continued to attend scheduled visits for efficacy evaluations (Retrieved Drop Out patients).

Following the completion of study 2311, all patients who participated in the core efficacy study 2311 were elected to continue in the extension study 2311 E1 for up to 24 weeks. Study 2311 E1 was an uncontrolled open-label study, where all patients received Exelon for up to 24 weeks. Regardless of whether they had been receiving placebo or Exelon in the core study, all patients who continued in the extension study, started a dose of 1.5 mg b.i.d. and were titrated to their maximum tolerated dose. No inferential statistics on efficacy evaluations were planned in this open-label study.

An additional uncontrolled study, study 2314, designed to show that the assessment scales used in study 2311 were valid and reliable in patients with PDD. In this study, patients did not receive study medication and efficacy was, therefore, not evaluated.

This reviewer will focus only on the efficacy core study 2311.

1.3 Statistical Issues and Findings

The core efficacy study 2311 was a prospective, randomized, multi-center, double-blind, placebo-controlled, parallel group study in patients with PDD. Five hundred and forty one (541) patients from 12 countries, 68 centers were randomized to receive the drug Exelon or placebo (ratio 2:1). The objective of the study is to test if the drug, Exelon statistically performs better in terms of specified clinical endpoints. Two primary efficacy endpoints, the change from baseline of the total ADAS-Cog score and ADCS-CGIC at Week 16 and Week 24 were considered. The sponsor proposed to use least square means derived by ANCOVA model with the following explanatory variables, country, baseline and treatment to analyze ADAS-Cog. The main analysis for ADCS-CGIC was the nonparametric categorical analysis.

Statistical Issues

- The primary population for the analysis is recommended by the agency is normally the ITT+LOCF, the intent to treat population using LOCF methodology to impute the missing values. In this study, the primary population for comparing the treatment groups proposed by the sponsor was the ITT+RDO population. This population included patients who discontinued study treatment early but continued to attend scheduled visits for efficacy evaluations (RDO patients). There were 23 RDO patients and among them 19 from Exelon groups and 4 from placebo group. In the ITT+LOCF population, values more than 2 days after the last dose of study drug were not carried forward; therefore, sample size in the ITT+LOCF population is smaller than that in the ITT+RDO population. However, it has been noticed that patients excluded from the Exelon group in the LOCF population (41 patients) is almost 6 fold of the patients in the placebo group (7 patients). The sponsor should explain why more patients' assessments were performed two days after the last dose in the Exelon group than in the placebo group.

  In this review, ITT+LOCF and ITT+RDO mean ITT population using LOCF or RDO to impute missing values.

- It has been noticed that the standard deviations of the placebo group for Austria were substantially smaller than the rest of the groups, consistently for baseline, Week 16 and Week 24. The standard deviations for Austria and the average standard deviations for other counties (Austria was excluded) at each treatment group are listed in Table 1. For example, at Week 24, the standard deviation for the placebo group (4 patients) was only 2.1 compared with 16.8 in the Exelon group (5 patients) in Austria and 10.24 for the rest of Exelon group and 12.04 for the rest of placebo group. Figures 1, 2 and 3, the grouped bar with error plots, show the average total ADAS scores and the corresponding standard deviations for both Exelon and placebo against the 12 countries at baseline, Week 16 and Week 24. The numbers in parentheses are the sample sizes in each country for the placebo and Exelon, respectively. It can be seen clearly that the standard deviation of the placebo group in Austria is much smaller than the rest.

- In this study, the center specific sample sizes were quite variable, ranged from 1 to 32. The sponsor showed significant improvement of the patients in the Exelon group for the two primary endpoints at both Week 16 and Week 24 when combining all the centers together. Like any multi-center study, the evaluation of the consistency of a treatment effect across the centers should be considered. In this multi-center study, since some centers had no patient assigned to one of the treatment arms, this reviewer examined the treatment effect by countries instead of centers for the cognitive function scale. Figure 4 and Figure 5 display the total change of ADAS-Cog scale from baseline at both weeks 16 and 24 across all countries. As can be seen from these graphs, the magnitude of the treatment effects differs among countries and the direction of the treatment effects are not consistent as well. Austria and Portugal show the wrong trend of the direction.

  Four different models were considered. Two models with only the main effect with/without combining the small centers together and the two models with both the main effect and the interaction term of the treatment and country with/without combining the small centers together. Table 2 displays the two-tailed P values for the least mean square results with ADAS-Cog endpoint for the above mentioned four different models. Scenario 1 is what was reported by the sponsor. The explanatory variables considered in the model were the country and treatment. In scenario 2, another term, the interaction of country and treatment was added based on the model in scenario 1. In scenario 3, after combining 3 small countries, Austria (5 subjects in Exelon, 3 subjects in placebo), Norway (4 subjects in Exelon, 1 subject in placebo) and Portugal (6 subjects in Exelon, 3 subject in placebo), the same model as in scenario 1 was considered. In scenario 4, the interaction term was added based on the model considered in scenario 3.

  If allowing sample sizes vary across all the countries (without pooling the small countries together), the results for the treatment effect can be very different depending on if the country-by-treatment interaction term was included in the ANCOVA model (comparing scenarios 1 and 2). There is no consensus whether the interaction term should be included in the model. If the interaction term was left out from the model, each country receives the weight according to the sample size of the patients enrolled in that country; whereas for the interaction model, each country receives an equal weight. Therefore, it is

not surprise to observe a totally different result for the treatment effect based on the two different models if sample sizes are very different across countries. Even though only 9 patients enrolled in Portugal, since this center is treated as same important as others in the interaction model, due to the large reversed treatment effect, this center can change the final result. It needs to be noted that though in the original protocol, the sponsor only proposed to use the main effect model.

After combining the small countries together, the final conclusions for both the main effect model and interaction model are very similar (comparing scenarios 3 and 4) since the sample sizes in each country are relatively compatible now.

Table 1 Standard deviations of Austria and the average of other 11 countries (Source: Reviewer's Analysis for study 2311)

	Country	Exelon (SD)	Placebo (SD)
Baseline	Austria	13.1	5.0
	Mean of others	10.12	10.2
Week 16	Austria	16.8	2.3
	Mean of others	10.61	11.69
Week 24	Austria	16.8	2.1
	Mean of others	10.24	12.04

Figure 1 Raw average total ADAS-Cog scores in each country and the corresponding standard errors at baseline (Source: Reviewer's Analysis for study 2311)

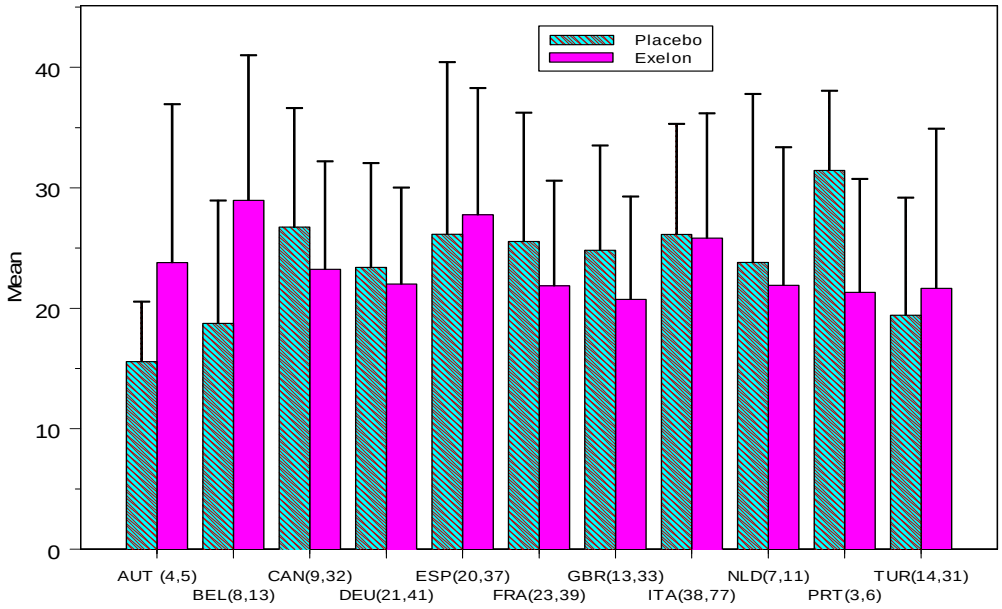

Figure 2 Raw average total ADAS-Cog scores in each country and the corresponding standard errors at Week 16 (Source: Reviewer's Analysis for study 2311)

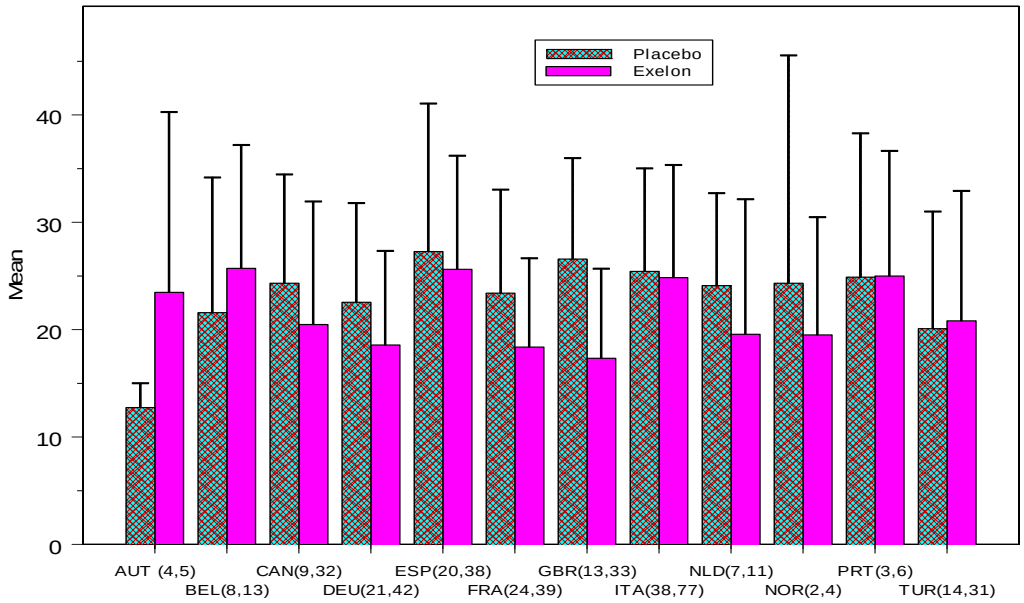

Figure 3 Raw average total ADAS-Cog scores in each country and the corresponding standard errors at Week 24 (Source: Reviewer's Analysis for study 2311)

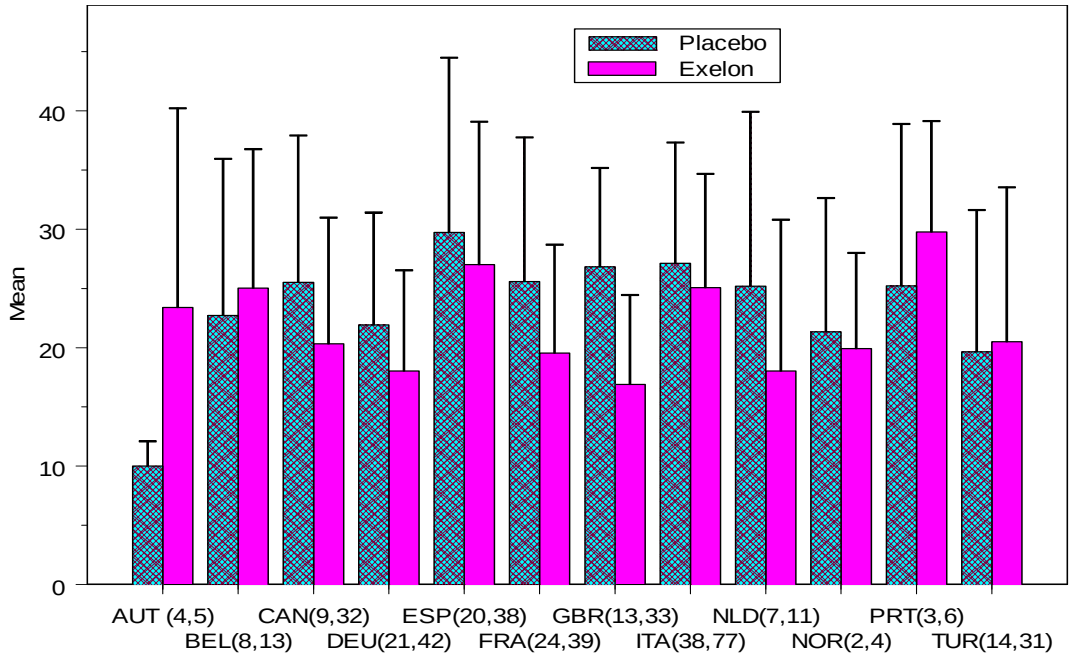

Figure 4 Total change from baseline for ADAS-Cog at Week 16 (Source: Reviewer's Analysis for study 2311)

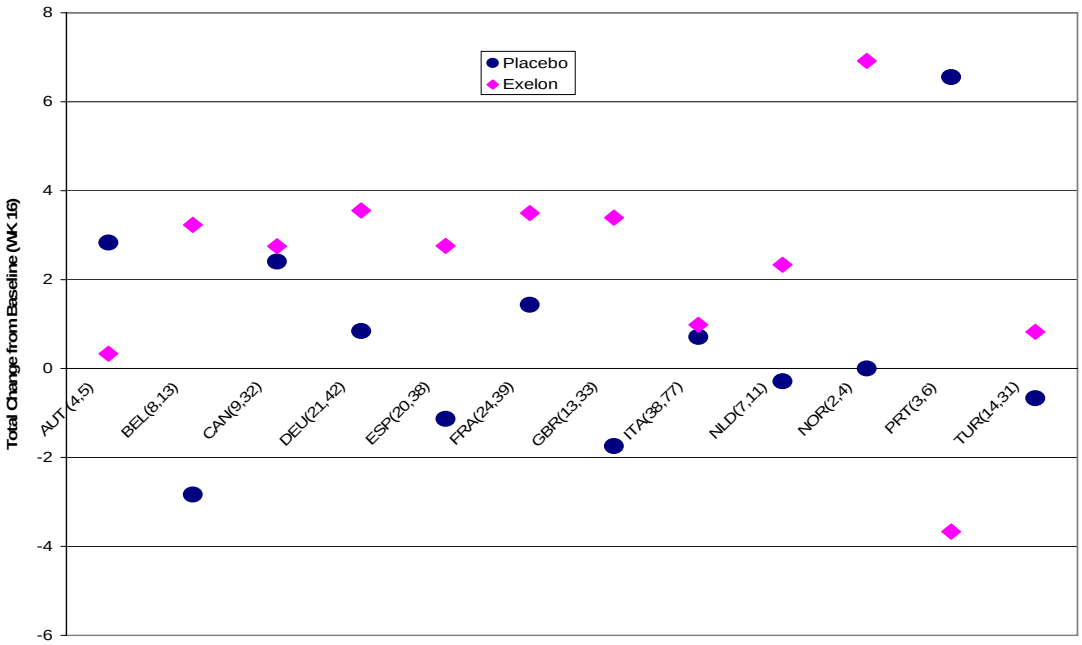

Figure 5 Total change from baseline for ADAS-Cog at Week 24 (Source: Reviewer's Analysis for study 2311)

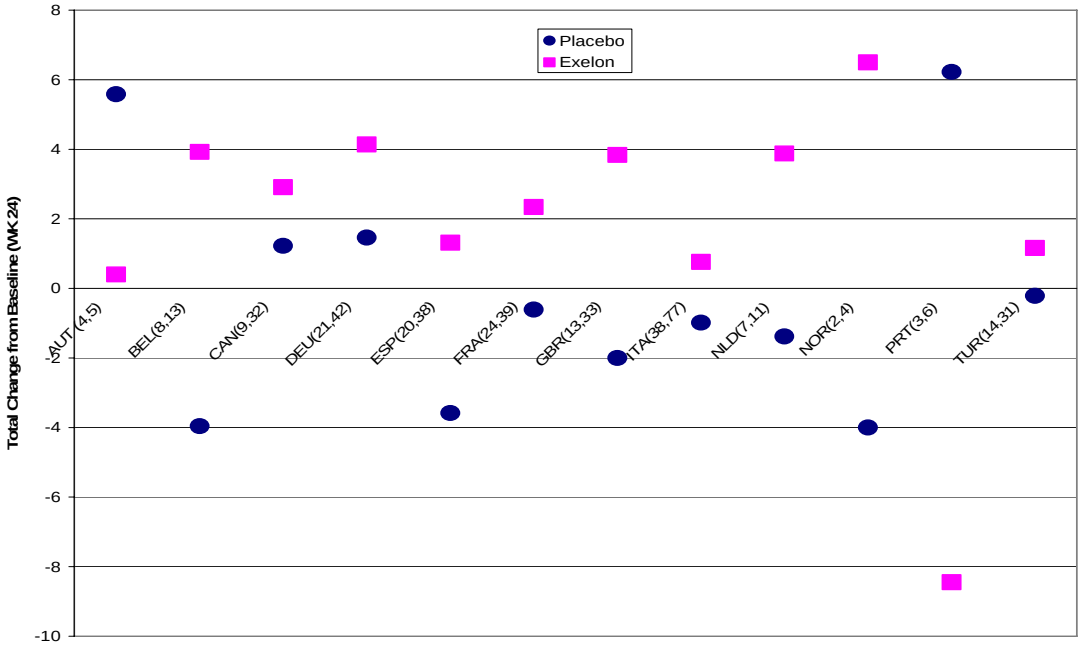

Table 2 P values for testing Exelon and placebo effect in a multi-center trial with/without interaction and combining small centers (Source: Reviewer's Analysis for study 2311)

	P values
Main effects [1]	
Week 16	0.0016
Week 24	<.0001
Interaction [2]	
Week 16	0.2019
Week 24	0.1121
Main effects (combining) [3]	
Week 16	0.0015
Week 24	<.0001
Interaction (combining) [4]	
Week 16	0.0058
Week 24	0.001

1: Scenario 1; 2: Scenario 2; 3: Scenario 3; 4: Scenario 4.

2 INTRODUCTION

Exelon ® (rivastigmine) was approved for treatment of mild to moderately severe Alzheimer's disease (AD) in 2000. The current core efficacy study 2311 aimed to evaluate the safety and efficacy of Exelon (3-12 mg/day) for 24 weeks in patients with Parkinson's Disease Dementia (PDD). The sponsor also conducted an uncontrolled open-label extension study, where all the PDD patients received Exelon for up to 24 weeks. In addition, another uncontrolled study, where all patients diagnosed with PDD dementia did not receive Exelon, was designed to validate various assessment scales used in the core efficacy study for the PDD patients. In this review, only the core efficacy study 2311 is relevant to the efficacy evaluation.

2.1 Overview

According to the sponsor's report, dementia occurs in approximately 20-60% of individuals with Parkinson's disease (PD), and is more likely to be present in elderly patients or those with more severe or advanced disease. Dementia in patients with PD is characterized by a clinical syndrome of mental slowing, executive dysfunction, retrieval type memory deficit and attentional impairment that may lead to a pronounced decline in the level of cognitive functioning, activities of daily living and behavior. Deficits in similar symptom domains of dementia are also observed in patients with AD. Exelon ® (rivastigmine) is a brain-selective, dual inhibitor of both acetylcholinesterase and butyrylcholinesterase that has been approved for the treatment of mild to moderately severe Alzheimer's disease. The present study aimed to study the efficacy and safety of Exelon (3-12 mg/day) in patients with PDD. It is a clinical judgment though how different AD and PDD are and whether practitioners can differentiate these differences.

The efficacy of Exelon in the treatment of PDD was evaluated in study 2311. This study was a 24-week prospective, randomized, multi-center, double-blind, placebo-controlled, two treatment

arm parallel group study. Patients enrolled were of either sex aged 50 years or older with the onset of dementia symptoms according to DSM IV criteria, occurring at least 2 years after the first diagnosis of idiopathic PD according to UK Parkinson's Disease Society Brain Bank clinical diagnostic criteria, with an MMSE score of 10 to 24. The dose of the drug was 3-12 mg/day. The overall duration of treatment was 24 weeks and consisted of a 16 week titration phase with titration steps at 4 week intervals and an 8-week maintenance phase. The primary efficacy endpoints included the change from baseline in ADAS-Cog total scores and ADCS-CGIC scale. The evaluation was performed at Week 16 and Week 24.

2.2 Data Sources

All documents reviewed for this NDA submission are in electronic form. The path to CDER Electronic Document Room for documents of this NDA is listed below:
\\CDSESUB1\N20823\S_016\2005-08-31

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 STUDY OBJECTIVES

The primary objective of Study 2311 was to evaluate the efficacy of Exelon (3-12 mg/day for 24 weeks) compared with placebo in patients with PDD based on ADAS-Cog (Alzheimer's Disease Assessment Scale-cognitive subscale) and the clinical global rating of change, ADCS-CGIC (Alzheimer's Disease Cooperative Study – Clinician's Global Impression of Change).

The secondary objectives included

- To evaluate the effects of Exelon on attention, executive functioning, activities of daily living, behavior and health economic parameters.
- To explore potential differences in efficacy of Exelon depending on preexisting attentional deficits.
- To explore the potential genetic factors related to PDD.
- To explore the potential biomarkers related to PDD.
- To evaluate the safety and tolerability of Exelon.

3.1.2 STUDY DESIGN

The core study 2311 was a 24-week, prospective, randomized, multi-center, double-blind, placebo-controlled, parallel group study in patients with a diagnosis of Parkinson's disease dementia according to the DSM-IV criteria (Code 294.1). The study was to be conducted in 68 centers in Europe and Canada. A total of 541 patients with PDD were to be randomly assigned to treatment with either Exelon 3-12 mg/day, or placebo in an assignment ratio of 2:1, i.e. 362 patients on Exelon and 179 patients on placebo.

After completion of the double-blind treatment phase, patients had the option to receive open-label treatment with Exelon for up to 6 months. This open-label extension study were to evaluate the safety and tolerability of Exelon for up to 24 weeks of exposure to the treatment in patients with PDD who completed a 24 week double-blind placebo-controlled core study, and to provide access or continued access to Exelon.

This reviewer will focus on the core study 2311 only.

3.1.3 EFFICACY MEASURES

3.1.3.1 Primary Efficacy Endpoints

There were two primary efficacy variables, a cognitive measure (Alzheimer's disease Assessment Scale-cognitive subscale, ADAS-cog) and a global measure (The Alzheimer's Disease Cooperative Study – Clinician's Global Impression of Change, ADCS-CGIC).

3.1.3.2 Secondary Efficacy Endpoints

Secondary efficacy parameters included:

- Alzheimer's Disease Cooperative Study – Activities of Daily Living (ADCS-ADL) for the assessment of activities of daily living
- Cognitive Drug Research (CDR) Computerized Assessment System tests for the assessment of attention
- D-KEFS Verbal Fluency Test, D-KEFS Color-Word Interference Test, D-KEFS Card Sorting Test and Symbol Digit Modalities Test for the assessment of executive functioning.
- Mini Mental State Examination (MMSE) score.
- NPI Caregiver Distress Scale (NPI-D) for the assessment of caregiver distress.
- Health Economic parameters, including caregiver burden, patient and caregiver resource utilization.

3.1.4 STATISTICAL ANALYSIS METHODS

The statistical efficacy tests were performed on several analysis data sets including Intent to Treat with Retrieved Dropouts (ITT+RDO), Last Observation Carried Forward (LOCF) and Observed Cases (OC). The proposed primary population for comparing the treatment groups was the ITT+RDO population. Analysis of covariance, ANCOVA, on the mean change from baseline was performed for the primary endpoint, ADAS-cog. A nonparametric categorical analysis, Van Elteren test was performed for the second primary endpoint, ADCS-CGIC in the presence of country as the blocking. All statistical tests were two-sided at the 5% significance level.

Primary Efficacy Analysis

Change from Baseline to Weeks 16 and 24 in Total ADAS-cog Score

The primary efficacy analysis of the change in total ADAS-cog score from baseline was based on a general linear model for analysis of covariance (ANCOVA) with factors for treatment group, countries and with baseline score of ADAS-cog as a covariate.

Global Clinical Rating of Change (ADCS-CGIC) at Week 24

The primary efficacy analysis of ADCS-CGIC was the treatment comparison based on a nonparametric test (Van Elteren test) with country as stratification variable.

3.1.5 STUDY RESULTS

3.1.5.1 Analysis Populations

The primary population used for the treatment comparison is the Intent To Treat with Retrieved Dropouts (ITT+RDO). This population includes all randomized patients who received at least one dose of study medication and had at least a pre-baseline assessment and a post-baseline assessment for one of the primary efficacy variables, either under treatment or not. This population included patients who discontinued study treatment early and continued to attend scheduled visits for efficacy evaluations.

Additional analyses based on populations ITT-Last observation carried forward (LOCF) and Observed Cases (OC) are considered supportive to the main analysis.

3.1.5.2 Analysis Populations

Patient disposition and main reasons for discontinuation are summarized in Table 3. Of the 541 patients randomized, 362 were in the Exelon group and 179 were in the placebo group. A total of 410 patients (76%) completed the study. The percentage of patients who discontinued was higher in the Exelon group (27.3%) compared to placebo (17.9%). This difference was mainly because of the adverse events (17.1% on Exelon and 7.8 % on placebo) and withdrawals of consent by the patients (5.8 % on Exelon and 1.1 % on placebo).

Table 3 Summary of Patient Disposition - All Patients Randomized (Source: Table 7-1 from 2311 study report)

	Exelon		Placebo		Total	
Number (%) of patients						
Screened					650	
Randomized	362	(100)	179	(100)	541	(100)
Exposed	362	(100)	179	(100)	541	(100)
Completed	263	(72.7)	147	(82.1)	410	(75.8)
Discontinued	99	(27.3)	32	(17.9)	131	(24.2)
Main reason for discontinuation	n	(%)	n	(%)	n	(%)
Adverse event(s)	62	(17.1)	14	(7.8)	76	(14.0)
Subject withdrew consent	21	(5.8)	2	(1.1)	23	(4.3)
Death	4	(1.1)	7	(3.9)	11	(2.0)
Protocol violation(s)	5	(1.4)	2	(1.1)	7	(1.3)
Unsatisfactory therapeutic effect	2	(0.6)	4	(2.2)	6	(1.1)
Lost to follow-up	4	(1.1)	1	(0.6)	5	(0.9)
Administrative reasons	0	(0.0)	2	(1.1)	2	(0.4)
Abnormal test procedure result(s)	1	(0.3)	0	(0.0)	1	(0.2)

3.1.5.3 Demographic Characteristics and Baseline Comparability

The patient demographic values at baseline are summarized in Table 4. Baseline demographic characteristics for age, gender and race were comparable in both treatment groups. The majority of patients were Caucasians.

Duration of PD, duration of PDD, and time interval between diagnosis of PD and initial symptoms of PDD were reported in Table 5. In the total population, the durations of PD reported by patients/caregivers and diagnosed by physicians were about 10 and 9 years, respectively. The durations of PDD reported by patients/caregivers and diagnosed by physicians were about 2.2 and 1.2 years. The mean duration between diagnosis of PD and first symptoms of PDD was 6.8 years. The distribution of PD severity as measured by Hoehn and Yahr as well as the average MMSE scores in both treatment groups were also reported in the table.

Table 4 Demographic Summary by Treatment Group (Source: Table 7-4 from 2311 study report)

		Exelon	Placebo	Total
		N = 362	N = 179	N = 541
Age (years)	Mean ± SD	72.8 ± 6.7	72.4 ± 6.4	72.7 ± 6.6
	Median	73.5	73.0	73.0
	Range	50 - 91	53 - 88	50 - 91
Age group – n (%)	< 65 years	49 (13.5)	19 (10.6)	68 (12.6)
	≥ 65 years	313 (86.5)	160 (89.4)	473 (87.4)
Gender – n(%)	Male	234 (64.6)	117 (65.4)	351 (64.9)
	Female	128 (35.4)	62 (34.6)	190 (35.1)
Race – n(%)	Caucasian	360 (99.4)	179 (100)	539 (99.6)
	Other	2 (0.6)	0	2 (0.4)

Table 5 Background Characteristics by Treatment Group (source: Table 7-5 from 2311 study report)

		Exelon	Placebo	Total
		N = 362	N = 179	N = 541
Time since first symptom of idiopathic PD was noticed by patient/ caregiver (years)	n	360	179	539
	Mean ± SD	9.8 ± 5.9	10.5 ± 6.3	10.0 ± 6.0
	Median (min-max)	8.8 (2.2 - 33)	9.8 (2.1 - 34.9)	9.0 (2.1 - 34.9)
Time since idiopathic PD was first diagnosed by physician (years)	n	362	179	541
	Mean ± SD	8.7 ± 5.7	9.4 ± 5.9	9.0 ± 5.8
	Median (min-max)	7.0 (0.1 - 32)	7.9 (2.0 - 34.8)	7.6 (0.1 -34.8)
Time since first symptom of dementia was noticed by patient / caregiver (years)	n	360	178	538
	Mean ± SD	2.1 ± 1.7	2.3 ± 1.9	2.2 ± 1.7
	Median (min-max)	1.8 (0 – 9.6)	1.9 (0.1 – 15.6)	1.8 (0 – 15.6)
Time since PDD was first diagnosed by physician (years)	n	362	179	541
	Mean ± SD	1.1 ± 1.3	1.4 ± 1.8	1.2 ± 1.5
	Median (min-max)	0.6 (0 – 8.0)	0.7 (0 – 13.6)	0.7 (0 – 13.6)
Time between diagnosis of PD and first symptoms of dementia (years)	n	360	178	538
	Mean ± SD	6.6 ± 5.2	7.2 ± 5.2	6.8 ± 5.2
	Median (min-max)	4.8 (-0.4 – 27.9)	5.9 (1.5 – 30.5)	5 (-0.4 – 30.5)
Modified Hoehn and Yahr staging (UPDRS Part V)	0	1 (0.3)	0	1 (0.2)
	1	7 (1.9)	4 (2.2)	11 (2.0)
	1.5	20 (5.5)	9 (5.0)	29 (5.4)
	2	65 (18.0)	31 (17.3)	96 (17.7)
	2.5	89 (24.6)	41 (22.9)	130 (24.0)
	3	114 (31.5)	63 (35.2)	177 (32.7)
	4	51 (14.1)	28 (15.6)	79 (14.6)
	5	15 (4.1)	2 (1.1)	17 (3.1)
Number of years of education	n	362	179	541
	Mean ± SD	8.8 ± 4.1	9.2 ± 3.9	9.0 ± 4.1
	Median (range)	8.0 (0-23)	9.0 (0-21)	8.0 (0-23)
MMSE score at baseline	Mean ± SD	19.4 ± 3.8	19.2 ± 4.1	19.3 ± 3.9
	Median	20.0	20.0	20.0
	Min-max	3 - 30	8 - 27	3 - 30

3.1.5.4 Protocol Violations

The type of protocol violations is listed in Table 6. Nine patients had MMSE scores outside the range of 10-24 permitted by the protocol. The duration between date of diagnosis of PD and initial symptoms of PDD was less than 2 years in 16 patients. The most frequent type of protocol violation in all patients was either new introduction or increase in dose of ongoing

dopaminergic or psychotropic medication. Forty patients discontinued the trial prematurely because of no primary assessment scales after the baseline evaluation. The percentage of patients with protocol violations was slightly higher in the Exelon group.

Table 6 Protocol Violations (Source: Table 7-4 from 2311 study report)

	Exelon	Placebo	Total
Total number of patients	362	179	541
Number (%) of patients with:			
At least one protocol violation	82 (22.7)	33 (18.4)	115 (21.3)
MMSE score < 10 or > 24	6 (1.7)	3 (1.7)	9 (1.7)
Date diagnosis PD> Date of first symptoms of PDD -2 years	13 (3.6)	3 (1.7)	16 (3.0)
Increased dose or newly introduced psychotropic/dopaminergic medication	39 (10.8)	18 (10.1)	57 (10.5)
No valid assessment of both primary variables	27 (7.5)	13 (7.3)	40 (7.4)

MMSE scores at baseline visit are reported.

3.1.5.5 Efficacy Results Reported by Sponsor

Primary Efficacy Results

ADAS-Cog

The results for the primary efficacy endpoint ADAS-Cog at week 16 and week 24 in both the primary analysis population (ITT+RDO) and the additional analysis populations (LOCF and OC) are listed in Table 7. The treatment groups were compared using least square means derived by ANCOVA with the following explanatory variables: treatment, country, and baseline total ADAS-Cog score. The treatment group difference for the change from baseline was statistically significantly in favor of Exelon in all three analysis populations, both at week 16 and at week 24.

Table 7 ADAS-Cog Change from Baseline (Source: Table 9-1 from 2311 study report)

	Exelon		Placebo				
	n	mean ± SD	n	mean ± SD	LS means difference	p-value	95% CI (Exelon – placebo)
ITT+RDO baseline	329	23.8 ± 10.2	161	24.3 ± 10.5			
Change at week 16	329	2.3 ± 7.3	161	0.3 ± 6.8	2.06	0.002 *	0.78 3.34
Change at week 24	329	2.1 ± 8.2	161	-0.7± 7.5	2.88	<0.001 *	1.44 4.31
LOCF baseline	287	24.0 ± 10.3	154	24.5 ± 10.6			
Change at week 16	287	2.8 ± 7.4	154	0.3 ± 6.7	2.74	<0.001 *	1.42 4.06
Change at week 24	287	2.5 ± 8.4	154	-0.8 ± 7.5	3.54	<0.001 *	2.05 5.04
OC baseline wk 16	284	23.9 ± 10.3	150	24.5 ± 10.6			
Change at week 16	284	2.8 ± 7.4	150	0.3 ± 6.8	2.78	<0.001 *	1.43 4.12
OC baseline wk 24	256	23.7 ± 10.4	139	23.4 ± 9.8			
Change at week 24	256	2.9 ± 8.3	139	-1.0 ± 7.6	3.80	<0.001 *	2.22 5.37

Higher change scores indicate greater improvement.

* $p < 0.05$. p-value based on two-way analysis of covariance model using treatment and country as factors and baseline ADAS-cog as a covariate; 95% confidence interval calculated for the difference between Least Squares Means (LSMEANS).

ADCS-CGIC

The endpoint ADCS-CGIC ratings were grouped into seven categories: (1) Markedly improved, scored as 1; (2) Moderately improved, scored as 2; (3) Minimally improved, scored as 3; (4) Unchanged, scored as 4; (5) Minimally worse, scored as 5; (6) Moderately worse, scored as 6 and (7) Markedly worse, scored as 7. The results for this primary efficacy endpoint at Week 24 are listed in Table 8. The treatment comparison for the mean scores in the two treatment groups was based on categorical analysis with country as a stratification variable. The difference of the ADCS-CGIC ratings at Week 24 was statistically significant different between two groups in favor of Exelon. This reviewer also performed the same analysis for Week 16. The improvement of ADCS-CGIC ratings due to Exelon at Week 16 was also statistically significant.

Table 8 ADCS-CGIC Ratings at Week 24 (Source: Table 9-3 from 2311 study report)

	ITT+RDO		LOCF		OC	
	Exelon	Placebo	Exelon	Placebo	Exelon	Placebo
N	329	165	289	158	252	145
Mean ± SD at week 24	3.8 ± 1.4	4.3 ± 1.5	3.7 ± 1.4	4.3 ± 1.5	3.7 ± 1.4	4.2 ± 1.5
Change	Exelon	Placebo	Exelon	Placebo	Exelon	Placebo
Markedly improved (1)	4%	2%	5%	2%	6%	2%
Moderately improved (2)	16%	12%	16%	12%	18%	12%
Minimally improved (3)	21%	15%	23%	16%	23%	15%
Unchanged (4)	26%	28%	25%	28%	25%	29%
Minimally worse (5)	21%	19%	20%	19%	19%	19%
Moderately worse (6)	11%	16%	9%	17%	8%	17%
Markedly worse (7)	2%	7%	2%	6%	2%	6%
p-value	0.007*		<0.001*		<0.001*	

p-value (Exelon vs. placebo) based on van Elteren test blocking for country. *: p<0.05

3.1.5.6 Review's Analysis

According to the protocol, the primary objective of the study requires demonstration of a statistically significant difference at the two-sided 5% level of significance between the Exelon group and the placebo group for each of the two primary endpoints, ADAS-Cog and ADCS-CGIC. This reviewer performed primary efficacy analyses independently following the methods specified in the protocol, and the results agree with those reported by the sponsor, treatment differences are statistically significant different in favor of the investigated drug. It needs to be pointed out though some issues have to be considered.

One requirement for the ANCOVA is the normality of the data. This reviewer tested the residuals using Shapiro-Wilk's test. The hypothesis of normality of the residual was rejected (P values = 0.0072 for Week 16 and < 0.0072 for Week 24) so that a nonparametric method (Wilcoxon rank test) was also performed. The results using the nonparametric method agree with those reported by the sponsors. For both weeks, the p-values are less than 0.05 in favor of Exelon.

For the ADAS-Cog endpoint, the sponsor proposed ANCOVA method using baseline total ADAS-Cog score, treatment and country as independent variables. The interpretation of the treatment effect is meaningful only if the regression relationships among two treatment groups are the same. Regression relationships that differ among two groups indicate an interaction between the treatment groups and the independent variable, the baseline measurement, and this interaction makes it hard to interpret the final treatment effect due to the drug. This reviewer performed a test to test for the heterogeneity of the slopes. Table 9 displays the results of the test for ADAS-Cog endpoint at both Week 16 and Week 24 among ITT+RDO population. It turns out that the two slopes at Week 16 are very similar; however, the slopes among two treatment groups at Week 24 are statistically significant different.

Therefore, if relying on the ANCOVA model to predict the treatment effect due to the drug, at low baseline values, the drug effect turns to be underestimated; whereas at the high baseline values, the drug effect will be overestimated.

Table 9 Estimates of the slopes in each treatment group and the P values for testing the heterogeneity of the slopes (Source: Reviewer's Analysis for study 2311)

		Estimate	Standard Error	P values for the Heterogeneity of slopes
Slope for Exelon	Week 16	0.216	0.037	
Slope for placebo		0.215	0.051	0.982
Slope for Exelon	Week 24	0.270	0.041	
Slope for placebo		0.120	0.057	0.034

For another primary endpoint, ADCS-CGIC, the sponsor proposed to use Van Elteren nonparametric method to test for the treatment effect using country as the blocking variable. At both Week 16 and Week 24, the results across all the countries are not consistent in terms of percentage of improvement after treatment. The total percentage changes from baseline after each treatment for each country are listed in Tables 10 & 11. Because of small sample sizes, three countries, Austria, Norway and Portugal were combined. As can be seen from both tables, in most countries, Exelon is better than placebo; however, in some countries, placebo performs better than Exelon. Since the results per country were not consistent, the final results should be interpreted with caution.

Table 10 ADCS CGIC – patients improving by treatment and country (Week 16) (Source: Reviewer's Analysis for study 2311)

		Exelon		Placebo	
	N	# Impr. (% Impr.)	N	# Impr. (% Impr.)	P values
Belgium	13	4 (30.77)	8	2 (25)	0.369
Canada	29	14 (48.28)	9	5 (55.56)	0.277
Austria, Norway, Portugal	13	8 (61.54)	8	1 (12.50)	0.035
Germany	42	16 (38.10)	21	3 (14.29)	0.036
Spain	37	13 (15.14)	20	5 (25)	0.178
France	35	17 (48.57)	20	6 (30)	0.094
United Kingdom	33	13 (39.39)	14	3 (21.43)	0.139
Italy	77	28 (36.36)	39	13 (33.33)	0.156
Netherlands	10	5 (50.0)	7	1 (14.29)	0.143

Turkey	29	17 (58.62)	13	11 (84.62)	0.077

Table 11 ADCS CGIC – patients improving by treatment and country (Week 24) (Source: Reviewer's Analysis for study 2311)

Week 24		Exelon		Placebo	
	N	# Impr. (% Impr.)	N	# Impr. (% Impr.)	P values
Belgium	13	3 (23.08)	8	2 (25)	0.394
Canada	31	15 (48.39)	9	5 (55.56)	0.275
Austria, Norway, Portugal	14	5 (35.71)	9	5 (55.56)	0.221
Germany	42	18 (42.86)	21	3 (14.29)	0.017
Spain	38	9 (23.68)	20	3 (15.00)	0.208
France	38	20 (52.63)	23	6 (26.09)	0.028
United Kingdom	34	15 (44.12)	14	4 (28.57)	0.161
Italy	77	23 (29.87)	40	12 (30.00)	0.168
Netherlands	11	5 (45.45)	7	1 (14.29)	0.174
Turkey	31	21 (67.74)	14	8 (57.14)	0.206

3.2 Evaluation of Safety

Please refer to Clinical Review by Dr. Ranjit Mani for Evaluation of Safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

The primary efficacy measures were analyzed in subgroups with regard to gender. There were a total of 190 female patients (128 females in the Exelon group and 117 females in the placebo group) and 351 male patients (234 males in the Exelon group and 117 males in the placebo group) in the study.

The subgroup efficacy results for ADAS-Cog are listed in Table 12. The results were consistent with overall findings even though some results for female do not meet the 0.05 nominal level.

The results for ADCS-CGIC are listed in Tables 13 & 14. When the subgroup analysis was performed by gender, the p-values for testing the difference of ADCS-CGIC ratings for both male and female PDD patients at Week 16 and for female patients at Week 24 for the primary analysis population are greater than 0.05.

It needs to be noted that the subgroup analysis was a post hoc analysis, without power and sample size properly adjusted for the significant testing.

Since all the patients were 50 years or older and 539 out of 541 enrolled patients were Caucasians, the subgroup analyses by age and by race are not performed.

Table 12 ADAS-Cog - Change from Baseline (Source: Reviewer's Analysis for study 2311)

	Exelon Mean (SD)	Placebo	p-value
Female			
ITT+RDO			
Week 16	2.2 (7.7)	0.6 (6.5)	0.166
Week 24	1.9 (8.4)	-0.9 (8.0)	0.027
ITT+LOCF			
Week 16	2.7 (7.9)	0.6 (6.2)	0.075
Week 24	2.6 (8.6)	-1.0 (8.0)	0.010
OC			
Week 16	2.8 (8.0)	0.6 (6.2)	0.066
Week 24	3.3 (8.5)	-1.7 (7.9)	0.004
Male			
ITT+RDO			
Week 16	2.4 (7.1)	0.1 (6.9)	0.005
Week 24	2.2 (8.1)	-0.6 (7.2)	0.001
ITT+LOCF			
Week 16	2.8 (7.1)	0.1 (7.0)	<0.001
Week 24	2.5 (8.3)	-0.7 (7.3)	<0.001
OC			
Week 16	2.8 (7.1)	0.1 (7.2)	<0.001
Week 24	2.6 (8.3)	-0.7 (7.4)	0.001

Table 13 ADCS-CGIC at Week 16 and Week 24 for Female (Source: Reviewer's Analysis for study 2311)

	ITT+RDO		ITT+LOCF		OC	
Female, Week 16	Exelon	Placebo	Exelon	Placebo	Exelon	Placebo
N	116	57	96	52	96	52
Mean ± SD	3.9 ± 1.4	4.2 ± 1.4	3.7 ± 1.4	4.2 ± 1.3	3.7 ± 1.4	4.2 ± 1.3
Markedly Improved (1)	4	4	4	4	4	4
Moderately improved (2)	13	11	15	10	15	10
Minimally improved (3)	28	9	30	10	30	10
Unchanged (4)	27	31	27	31	27	31
Minimally worse (5)	13	31	11	33	11	33
Moderately worse (6)	14	13	11	12	11	12
Markedly worse (7)	3	2	1	2	1	2
p-value	0.245		0.049		0.049	
Female, Week 24						
N	116	57	99	54	81	50
Mean ± SD	3.9 ± 1.5	4.3 ± 1.4	3.7 ± 1.4	4.4 ± 1.4	3.6 ± 1.4	4.2 ± 1.3
Markedly Improved (1)	2	2	2	0	2	0
Moderately improved (2)	19	14	20	13	25	14
Minimally improved (3)	19	11	23	11	21	12
Unchanged (4)	28	30	30	30	30	32
Minimally worse (5)	14	21	11	22	12	24
Moderately worse (6)	15	19	11	20	9	16
Markedly worse (7)	3	4	2	4	1	2
p-value	0.350		0.035		0.012	

Table 14 ADCS-CGIC at Week 16 and Week 24 for Male (Source: Reviewer's Analysis for study 2311)

	ITT+RDO		ITT+LOCF		OC	
Male, Week 16	Exelon	Placebo	Exelon	Placebo	Exelon	Placebo
N	206	104	186	101	186	101
Mean ± SD	3.7 ± 1.3	4.0 ± 1.4	3.6 ± 1.3	4.0 ± 1.4	3.6 ± 1.3	4.0 ± 1.4
Markedly Improved (1)	4	2	4	2	4	2
Moderately improved (2)	15	13	16	13	16	13
Minimally improved (3)	23	21	24	21	24	21
Unchanged (4)	29	30	28	30	28	30
Minimally worse (5)	23	19	23	20	23	20
Moderately worse (6)	4	12	3	12	3	12
Markedly worse (7)	2	4	1	3	1	3
p-value	0.167		0.06		0.06	
Male, Week 24						
N	213	108	190	104	171	95
Mean ± SD	3.8 ± 1.4	4.3 ± 1.5	3.8 ± 1.4	4.2 ± 1.5	3.7 ± 1.4	4.2 ± 1.5
Markedly Improved (1)	6	3	6	3	7	3
Moderately improved (2)	14	11	14	12	15	12
Minimally improved (3)	22	18	23	18	23	17
Unchanged (4)	24	27	23	27	23	27
Minimally worse (5)	24	19	24	17	23	17
Moderately worse (6)	8	15	8	15	8	17
Markedly worse (7)	2	8	2	8	2	7
p-value	0.045		0.055		0.025	

4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The sponsor proposed to use the ITT+RDO as their primary analysis population. Normally it is recommended by the agency to use the ITT+LOCF as the primary analysis population. RDO patients discontinued study treatment early but came back for the efficacy evaluations. The ITT+LOCF population only carried forward the results if their assessment were done within 2 days after the last dose of study drug. In study 2311, values of 41 patients in Exelon group and 7 patients in Placebo group were not carried forward since the assessment were done 2 days after the last dose of the study drug (the ratio is almost 6 between the two treatment groups). The sponsor did perform the same analyses for ITT+LOCF population and the results were consistent with the findings based on the analysis from ITT+RDO population. The reviewer's analysis agrees with the reported findings.

The results based on the subgroup analyses (by gender) show that in some situations, the magnitude of the treatment difference between male and female is different. For instance, for the primary endpoint ADAS-Cog, the data did not show a difference between the two groups for female at Week 16 at a nominal level 0.05. For another primary endpoint ADCS-CGIC, among the female patients, at both week 16 and 24, the data did not show a difference between the two treatments at a significant level 0.05 based on ITT+RDO population. Among the male patients, no differences between Exelon and Placebo were detected at Week 16 based on all the three analysis populations and at Week 24 based on ITT+LOCF population at a nominal level 0.05. As mentioned above, the subgroup analysis is a post hoc analysis.

5.2 Conclusions and Recommendations

The data based on Study 2311 support the efficacy of 3-12 mg/day of Exelon® (rivastigmine) in patients with Parkinson's disease dementia based on the statistical methods proposed in the original protocol. Some sensitivity analyses still support the efficacy of Exelon.