CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

022568

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 22, 568 (0000)

Drug Name: Aricept 23 mg Sustained Release

Indication(s): Alzheimer's Disease

Applicant: Eisai

Date(s): Submission Date: Sept 24, 2009

Review Priority: Standard

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Tristan Massie, Ph.D.

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

Jim (Hsien Ming) Hung, Ph.D., Division Director

Medical Division: Division of Neurology (HFD-120)

Clinical Team: Ranjit Mani, M.D., Reviewer and Team Leader

Russell Katz, M.D, Division Director

Project Manager: Teresa Wheelous

Keywords: Active Control; Assay Sensitivity; Co-primary endpoints; Post-hoc subgroup result

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

This application for a new higher dose formulation of Donepezil rests on a single study which utilized the approved dose of the original formulation as the control group rather than a placebo control. The trial demonstrated a statistically significant effect on the co-primary cognitive endpoint ($p<0.001$) as determined by the pre-specified primary analysis, but the treatment difference on the co-primary global endpoint, CIBIC+, was not significant, $p=0.179$. The sponsor reported some post-hoc subgroups in which the treatment difference on the CIBIC+ was nominally significant, but this does not meet the usual standards; in addition, the results of various subgroups are inconsistent (see Sections 1.3 and 3.1.1.5), they would need to be replicated.

It is uncertain whether the trial had assay sensitivity. For example, the cognitive endpoint was 10-20 points higher on average than in the previous Donepezil studies in which it was used. The CIBIC+ was also not a good choice of co-primary endpoint in the absence of a placebo control since the earlier placebo controlled results for the low dose of Donepezil were mixed so that it would be questionable to make a non-inferiority argument. Unless there is some compelling prior reason to believe that there is a dose response between 10 mg IR (immediate release) and 23 mg SR (suspended release) the data from this trial does not seem to provide enough support for the efficacy of the 23 mg SR formulation.

1.2 Brief Overview of Clinical Studies

.

Study 326, the only efficacy study for this new formulation of Aricept, was performed at 220 sites in Asia (including India), Oceania, Europe, North America, Africa, and South America. Four hundred sixty five (32%) of the 1467 randomized patients in the trial were randomized in the U.S. Study 326 is the single, pivotal Phase III study to support efficacy of the donepezil SR 23 mg formulation for the treatment of moderate-to-severe Alzheimer's Disease (AD). The study used Donepezil 10 mg IR as a control rather than placebo. The study was conducted between 06 June 2007 and 27 March 2009 and was a randomized, double-blind, multicenter study. To be eligible for the study, patients were required to be on a stable dose of donepezil IR 10 mg (Aricept or a bioequivalent generic) for at least 3 months prior to screening in order to assure that the maximum theoretical therapeutic benefit had been achieved on this dose.

1.3 Statistical Issues and Findings

Overall, the treatment difference on the co-primary Severe Impairment Battery (SIB) change at week 24 was statistically significant ($p<0.0001$). Effects were numerically larger in the U.S. which accounted for 32% of the randomized patients. An exploratory test for an interaction between U.S. and treatment group on the change from baseline in SIB at Week 24 yielded a p-value of 0.0539. A similar trend was seen for the co-primary CIBIC+ endpoint. The overall result for the CIBIC+ was not statistically significant but in the U.S. subgroup the exploratory result reached the nominal significance level. There were several nominally significant differences at baseline between the U.S. and non-U.S. populations which raise questions about the

generalizability of the U.S. subgroup result for the CIBIC+ at week 24 even if one were to assume that it was real and not merely due to chance.

Previous trials did not demonstrate a consistently significant effect on CIBIC+ of Aricept 10 mg IR vs. placebo. Therefore, it would not be reasonable to consider a non-inferiority approach to show that 23 mg was superior to putative placebo for the CIBIC+. In this light, the fact that the overall co-primary CIBIC+ result comparing 23 mg to 10 mg did not reach statistical significance may be a major problem unless we can appeal to prior beliefs that the treatment effect should increase as a function of dose. Even that might be questioned if the effect of 10 mg on CIBIC+ is in doubt because the 23 mg dose might not be high enough to observe an effect. Treatment differences on other secondary endpoints, Alzheimer's Disease Cooperative Study-Activities of Daily Living (ADCS-ADL) and Mini-Mental Status Exam (MMSE), did not reach nominal significance either.

In the meeting minutes from an 11 April 2008 meeting the FDA questioned the appropriateness of the SIB as an endpoint for patients with MMSEs >=15 saying that they were not fully convinced, but based on the submitted data it was, nevertheless, acceptable to proceed. This study randomized patients with MMSEs up to 22. Perhaps related to this concern, the sponsor claims nominal significance of the CIBIC+ treatment difference in a post-hoc subgroup, those with baseline MMSE scores from 0-16. This seems problematic first and foremost because it is post-hoc, but also because results for subgroups (3-14 N=728, p=0.0508; 0-14 N=769, p=0.1663 0-15 N=858, p=0.0938; 0-17, N=1063, p=0.0687) within this group do not reach nominal significance. The insignificant overall CIBIC+ result together with the lack of consistency of these additional subgroup results seem to undermine the sponsor's post-hoc result in the 0-16 subgroup.

There were significantly more dropouts in the 23 mg SR group (30%) than in 10 mg IR group (18%). While the primary Intent-to-Treat Population-Last observation Carried forward (ITT-LOCF) analysis and the secondary Observed Cases (OC) analysis of the change from baseline in SIB at Week 24 were nominally significant in favor of the high dose, significance was lost for an analysis assigning the lowest rank for dropouts and some other similar sensitivity analyses. In most cases where significance was lost the high dose was still numerically better than the low dose, but these analyses still raise the importance of the question of whether the trial had assay sensitivity. The mean SIB was between 10 and 20 points higher at baseline in this trial than in the previous Donepezil trials in which the SIB was used as an endpoint; this may further complicate cross trial comparisons and the lack of a placebo control issue.

An interim analysis after 400 patients had completed 24 months with possible stopping for efficacy or futility was originally planned for this trial. The stopping rule was to be specified in the Indpendent Data Monitoring Committee (IDMC) charter document. The final protocol amendment dated 20 Jun 2008 stated that interim efficacy would not be assessed because based on the enrollment rate seen it was likely that all patients would be enrolled before the interim analysis results would be available. The sponsor's study report also states that no interim efficacy analysis was done and, therefore, no adjustment to the alpha level for the final analysis is necessary. This reviewer notes that all patients had been randomized by the date associated with the interim analysis plan, 21 November 2008. For more details see section 3.1.1.5.5 (page

28).

2 INTRODUCTION

2.1 Overview

Donepezil (also called Aricept) is currently marketed as the immediate-release (IR) formulation, at doses of 5 mg and 10 mg. The IR formulation of 10 mg donepezil has been shown to be effective in multiple well-controlled randomized clinical trials as a treatment for mild to moderate and severe dementia in patients with Alzheimer's Disease (see NDA 20690). Eisai developed a modified formulation of donepezil that was designed to provide sustained release over an 8-hour time period and allow acceptable tolerability of a higher, more effective dose in those patients who might benefit. Because the effectiveness of donepezil is established, and donepezil SR 23 mg represents a new dose and formulation of an approved product, a single randomized, double-blind pivotal efficacy Phase III study (E2020-G000-326) was judged sufficient to support the claim that donepezil SR 23 mg is superior to IR 10 mg as a treatment for moderate to severe AD. The 23 mg dose, i.e., a dose slightly higher than 20 mg, was selected for the SR formulation based on the PK profile from the first Phase I study that found slightly reduced area under the plasma concentration-time curve (AUC) for the SR relative to equivalent doses of the IR formulation.

The study was performed at 220 sites in Asia (including India), Oceania (including Australia), Europe, North America, Africa, and South America. Four hundred sixty five (32%) of the 1467 randomized patients in the trial were randomized in the U.S.

2.2 Data Sources

At the time of review the sponsor's study data was contained in the following directories.
\\Cdsesub1\evsprod\NDA022568**0000**\m5\datasets\e2020-g000-**326**\analyses
\\Cdsesub1\evsprod\NDA022568**0000**\m5\datasets\e2020-g000-**326**\tabulations

At the time of review the sponsor's study report was contained in the following directory.
\\cdsesub1\evsprod\NDA022568\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud\moderate-to-severe-alzheimers-disease\5351-stud-rep-contr\e2020-g000-326\body.pdf

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study 326

3.1.1.1 Study Design and Analysis Plan

The study was initiated on 06 June 2007 and the last subject completed on 27 March 2009. The original protocol was dated 30 October 2006. It was amended 6 times including the final version, dated 20 June 2008. The statistical analysis plan is dated March 31, 2009.

The original protocol, dated 30 October 2006, underwent four amendments during the study: Amendment 1, dated 12 February 2007; Amendment 2, dated 10 May 2007; Amendment 3, dated 14 February 2008; and Amendment 4, dated 20 June 2008. Amendment 2 was incorporated after 3 patients had been enrolled into the study. Table 10 presents the numbers of patients enrolled, by protocol amendments. The primary objective of this study was to compare 23 mg donepezil sustained release (SR) with 10 mg donepezil immediate release (IR) in the treatment of subjects with moderate to severe Alzheimer's disease.

Table 1 Number of Patients Enrolled by Time of Protocol Amendment (Randomized Population)

	Number (%) of Patients		
Amendments	Donepezil SR 23 mg (n=963)	Donepezil IR 10 mg (n=471)	Total (N=1434)
Amendment 1	2 (0.2)	1 (0.2)	3 (0.2)
Amendment 1-2	601 (61.3)	309 (63.6)	910 (62.0)
Amendment 1-3	945 (96.3)	468 (96.3)	1413 (96.3)
Amendment 1-4	981 (100.0)	486 (100.0)	1467 (100.0)

Note: This table was copied from the sponsor's study report, page 63
The reviewer was unable to verify this table

This study consists of a randomized, double-blind, double-dummy, parallel-group comparison of 23 mg donepezil SR with the currently marketed donepezil formulation (10 mg donepezil IR) in subjects with moderate to severe Alzheimer's disease. Subjects must have been taking Aricept® 10 mg IR (or a bioequivalent generic) for at least 3 months prior to Screening. The study was planned to consist of 24 weeks of daily administration of study medication, with clinic visits at Screening, Baseline, 3 weeks (safety only), 6 weeks, 12 weeks, 18 weeks and 24 weeks or early termination. Subjects were to receive either 10 mg donepezil IR in combination with a placebo corresponding to 23 mg donepezil SR, or 23 mg donepezil SR in combination with a placebo corresponding to 10 mg donepezil IR.

A total of approximately 1200 patients were to be randomized. During the Baseline visit, patients were to be randomized in a 2:1 ratio (23 mg donepezil SR to 10 mg donepezil IR). Within each treatment group, patients were to be stratified according to whether they received donepezil alone or donepezil plus memantine during the study treatment period. The study was to be performed at approximately 200 global sites (Asia, Oceania, Europe, India, Israel, North America, South Africa, and South America).

Sample Size and Power Considerations

The sample size determination is based on a Type I error = 0.05, two-sided t-test, using the pooled standard deviations from the ITT-LOCF Week 24 analysis. The overall study sample is estimated on the basis of the primary efficacy variables, SIB and CIBIC+.

The following power calculations support an overall statistical power of at least 80% (99% for SIB) to show a difference, between treatment groups, of 0.20 (S.D.=1.053) points on the CIBIC+ and a difference, between treatment groups, of 3.0 (S.D.=9.544) points on the change from baseline in the SIB. An estimated sample size of 981 subjects is needed (654:327 for 23 mg SR to 10 mg IR). To allow for an approximately 80% rate of completion, the planned total number of subjects to be randomized was approximately 1200 subjects (800:400; 23 mg SR to 10 mg IR). The sample sizes also support a statistical power of approximately 80% to detect a difference, between treatment groups, of 1.37 (S.D.=7.18) points on the ADCS-ADL (severe version).

Note that Amendment 2 (dated 08 May 2007) increased the sample size from 1200 (N=800 and 400) to 1600 (N=1067 and 533) total patients. The sponsor stated that the change was based on a decision to increase the desired power of the study from 80 to 90%. No patients had been randomized by the date of amendment 2.

Amendment 4 (dated 20 Jun 2008) revised the sample size back to the originally planned 1200 with the justification that “the industry standard for statistical power is 80%”. Note that 60% percent of patients had provided consent, 52% been randomized and 21% had finished by the date of amendment 4.

Definitions of analysis populations (analysis sets)

The Safety Population was to be used in the statistical analyses of safety. Subjects included in the Safety Population were to be those who were randomized, took at least one dose of study medication, and who had at least 1 post-Baseline safety assessment. In the event that a subject received study drug different from the one to which he/she was randomized, the subject’s safety data was to be analyzed “as treated.”

The Intent-to-Treat Population (ITT) population was to be used in the statistical analyses of efficacy. This population was to consist of all randomized subjects who are in the Safety Population and for whom either (a) SIB data are available at Baseline and at least one subsequent SIB data point is available post-Baseline, or (b) Clinician’s Interview-Based Impression of Severity (Plus Caregiver Input Version, CIBIS+) data are available at Baseline, and at least one subsequent CIBIC+ data point is available post-Baseline. In the event that a subject received study drug different from the one to which he/she was randomized, the subject’s efficacy data was to be analyzed “as randomized.”

Primary Endpoints
The co-primary efficacy endpoints, using the Last Observation Carried Forward (LOCF) approach, are:

Severe Impairment Battery (SIB) total score change from Baseline to Week 24

Clinician's Interview-Based Impression of Change (Plus Version, CIBIC+) score at Week 24.

Secondary endpoint(s)
Secondary efficacy endpoints (using LOCF approach) include:
• Mini-Mental State Examination (MMSE) total score change from Baseline to Week 24
• ADCS-ADL total score change from Baseline to Week 24 (In the EU only, the co-primary endpoints are the CIBIC+ and the ADCS-ADL scales.)

Pooling of centers
This study is a multi-center, international study conducted in the following regions:

Asia
Oceania (e.g., Australia)
Europe
India
Israel
North America
South Africa
South America

Since some centers have low enrollment (only one randomized subject), data from all centers within a country were to be pooled together for analysis purposes. Countries would then be sorted in descending order by the number of subjects. The next step was to identify the largest country without an ITT subject in at least one treatment group. If that country was the smallest country on the list, then that country was to be pooled with the next country (or countries) above it on the list so that the pooled country would have at least one ITT subject per treatment group. If the identified country was not the smallest on the list, then that country was to be pooled with other countries below it on the list and, if needed, with other countries above it on the list so that the pooled country would have at least one ITT subject per treatment group. This process was to be continued until all (pooled) countries had at least one ITT subject per treatment group.

Adjustments for covariates
Efficacy analyses were to adjust for the corresponding baseline value as a covariate and a factor for country. With regards to Multiple Comparisons/Multiplicity, since both co-primary endpoints must demonstrate superiority for 23 mg donepezil SR as compared to 10 mg donepezil IR in the ITT-LOCF population in order for the outcome to be declared positive, the overall Type I error rate of 0.05 is controlled.

Primary efficacy analysis
For the continuous efficacy endpoint of SIB change from Baseline to Week 24, an analysis of covariance (ANCOVA) model with terms for baseline, country, and treatment was to be used as

the primary model for estimating and testing treatment effects. For the categorical endpoint of CIBIC+ impression-of-change score at Week 24, a nonparametric ANCOVA method, with a Cochran-Mantel-Haenszel test component, as described by Koch et al. (1998), was to be performed. The analysis was to adjust for CIBIS+ at baseline with a stratification adjustment for country.
Note that the protocol was not entirely clear on how this analysis would be done up until the time of the final protocol. Prior to that it just said that it would be specified in the analysis plan before unblinding. In the final protocol it characterized the analysis of CIBIC+ as an analysis of covariance combined with a Cochran-Mantel-Haenszel test and provided the following additional details.

Nonparametric ANCOVA
A non-parametric ANCOVA with a Cochran-Mantel-Haenszel test component, as described by Koch et al. (1998), was to be performed on the CIBIC+ impression-of-change score at Week 24. The analysis was to adjust for CIBIS+ at baseline with a stratification adjustment for country. Since it is expected that the sample sizes are not large for all strata, the following SAS code using the MIXED and FREQ procedures was to be used, as suggested by Koch et. al. in Appendix VII of their paper.

```
proc mixed;
class country;
model endpoint=baseline country/outp=res;
run;
proc freq data=res;
tables country*trt*resid/cmh;
run;
```

Handling of missing efficacy data, drop-outs, and outliers
The primary approach to handling missing post-baseline efficacy data at each visit is the LOCF method. If a subject is missing a Week 24 endpoint observation, then the last post-baseline observed value was to be carried forward and used as the Endpoint visit observation. The primary analysis of efficacy was to be conducted on the data set for the ITT-LOCF Population. The data set from Week 24-LOCF is the scheduled Endpoint data set for this study.

INTERIM ANALYSIS
The interim analysis was planned to be conducted after approximately the first 400 subjects (with efficacy and safety data) had been randomized and had completed the study (24 weeks or early termination). The objective of this interim analysis was to assess safety, including any unexpected toxicity. If the results of the interim analyses indicated serious safety concerns, the Sponsor would consult with health regulatory authorities (HRAs) regarding stopping the trial. Safety assessments were to include summaries of incidence rates of adverse events, changes in vital signs and weight, changes in laboratory parameters, rates of abnormal overall Electrocardiogram (ECG) interpretations, rates of concomitant medication use, and premature termination. According to the interim analysis plan dated 21 November 2008 there were no planned efficacy analyses for the interim analysis and no efficacy data were being included in the

interim analysis. Furthermore, there would be no inflation of type I error since no statistical testing was to be performed.

However, this was not always the plan. The original protocol stated that an efficacy analysis would be conducted at the interim for the purposes of stopping early due to efficacy. Early stopping on the grounds of efficacy was to be considered by the Independent Data Monitoring Committee (IDMC) according to statistical criteria pre-specified in the IDMC charter. In contrast, in the statistical analysis plan written later it was stated that only safety data would be analyzed in the interim analysis. According to the final study report because of the late surge of enrollment into the study, by the time sufficient efficacy data were available for interim analysis (400 patients) all patients had been enrolled. It stated that it was, therefore, inappropriate to conduct the interim efficacy analysis. This decision was endorsed by the IDMC. In particular, Amendment 4 (20 Jun 2008), which was the final protocol amendment, removed the potential for stopping for efficacy at the interim analysis stating that "it is likely that all patients would be enrolled by the time the results of the interim analysis became available". According to the study report, as a result, only safety data were reviewed during the interim analysis.

The interim analysis plan has a date of 21 November 2008. The document seems to be inconsistent in terms of whether or not efficacy data were to be included in the interim analysis. In particular, while it includes the statement "no efficacy data are included in the interim analysis" it also describes a futility analysis which it would not be possible to conduct without looking at efficacy data. One possible interpretation would be that there was to be no possibility of stopping early for superior efficacy of Aricept 23 mg SR at the time of the interim analysis but efficacy data was to be looked at in the interim analysis to see if the chance of demonstrating such efficacy at the end was so small as to make it futile to continue the trial.

3.1.1.2 Disposition of Subjects

Table 2 presents a summary of patient disposition. A total of 2186 patients were screened for the study and 1467 patients were randomized in a 2:1 ratio to treatment: 981 (66.9%) to donepezil SR 23 mg and 486 (33.1%) to donepezil IR 10 mg, respectively. Of these, 1084 (73.9%) patients completed the study; 296 (30.2%) patients in the SR 23 mg group and 87 (17.9%) patients in the IR 10 mg group discontinued from the study prematurely. A total of 182 (18.6%) patients in the donepezil SR 23 mg group and 39 (8.0%) patients in the donepezil IR 10 mg group discontinued due to treatment-emergent adverse events (AEs), and 114 (11.6%) patients in the SR 23 mg group and 48 (9.9%) patients in the IR 10 mg group discontinued for other reasons including "patient withdrew consent," protocol violation, medication non-compliance, request of investigator or sponsor, lack of efficacy, or "other." Overall, 963 of 981 (98.2%) patients in the donepezil SR 23 mg group and 471 of 486 (96.9%) patients in the donepezil IR 10 mg group were evaluable for safety as part of the Safety Population, and 909 (92.7%) patients in the donepezil SR 23 mg group and 462 (95.1%) patients in the donepezil IR 10 mg group were evaluable for efficacy as part of the ITT Population. Seventeen patients (9 donepezil SR 23 mg and 8 donepezil IR 10 mg) received at least one dose of study medication but were not included in the Safety Population due to lack of post-baseline safety assessment and 16 patients (9

donepezil SR 23 mg and 7 donepezil IR 10 mg) discontinued from the study prior to receiving study drug.

Table 2 Patient Disposition: All Patients

	Number (%) of Patients		
	Donepezil SR 23 mg	Donepezil IR 10 mg	Total
Screened			2186
Failed screening			719
Randomized[a]	981	486	1467
Completed	685 (69.8)	399 (82.1)	1084 (73.9)
Discontinued	296 (30.2)	87 (17.9)	383 (26.1)
Reason for discontinuation			
AE[b]	182[c] (18.6)	39[d] (8.0)	221 (15.1)
Medication non-compliance	9 (0.9)	4 (0.8)	13 (0.9)
Protocol violation	15 (1.5)	5 (1.0)	20 (1.4)
Request of investigator or sponsor	4 (0.4)	5 (1.0)	9 (0.6)
Patient withdrew consent	61 (6.2)	22 (4.5)	83 (5.7)
Lack of efficacy	1 (0.1)	0 (0.0)	1 (0.1)
Other	24 (2.4)	12 (2.5)	36 (2.5)
Deaths	7 (0.7)	5 (1.0)	12 (0.8)

Source: Tables 14.1.1.2 and 14.1.1.3.1
a: Number of randomized patients was used as the denominator for calculating percentages.
b: Includes SAEs.
c: Patient 60771009 experienced a single AE after randomization but prior to receiving double-blind study medication. The patient discontinued from the study due to the events (see Data List 16.2.1.3, Subject Disposition), but was not included in Table 14.3.2.3.1 (Listing of Adverse Events Leading to Discontinuation) since the patient was not included in the Safety Population.
d: Patient 60941001 experienced multiple AEs after randomization but prior to receiving double-blind study medication. The patient discontinued from the study due to the events (see Data List 16.2.1.3, Patient Disposition), but was not included in Table 14.3.2.3.1 (Listing of Adverse Events Leading to Discontinuation) since the patient was not included in the Safety Population.
Abbreviations: IR – immediate release; SR – sustained release; AE – adverse event.
Copied from page 72 of sponsor's study report

3.1.1.3 Demographic Characteristics

Table 3 presents a summary of demographic characteristics at baseline. Mean (± SD) age of patients in the Safety Population was 73.8 ± 8.53 years (73.9 ± 8.53 in the donepezil SR 23 mg group and 73.8 ± 8.56 in the donepezil IR 10 mg group). The majority of patients were female (donepezil SR 23 mg, 63.0%; donepezil IR 10 mg, 62.4%). Most of the patients were White (73.5% in both treatment groups). Racial distribution was comparable between groups. The numbers of patients were approximately evenly distributed among the four weight groups (< 55 kg, 55 to < 65 kg, 65 to < 75 kg, and ≥ 75 kg) for both treatment groups; the mean BMI was 25.20 ± 4.39 kg/m2 in the donepezil SR 23 mg group and 25.04 ± 4.30 kg/m2 in the donepezil IR 10 mg group. The majority of patients (79.8%) lived with their caregivers (81.0% donepezil SR 23 mg; 77.5% donepezil IR 10 mg). The percentage of patients on concomitant memantine was 36.3% (36.6% donepezil SR 23 mg; 35.7% donepezil IR 10 mg).

Table 3 Summary of Demographic Characteristics: Safety Population

	Donepezil SR 23 mg (n=963)	Donepezil IR 10 mg (n=471)	Total (N=1434)
Age (years)			
Mean (SD)	73.9 (8.53)	73.8 (8.56)	73.8 (8.53)
Median	75.0	75.0	75.0
Min, Max	47, 90	49, 90	47, 90
Gender: n (%)			
Male	356 (37.0)	177 (37.6)	533 (37.2)
Female	607 (63.0)	294 (62.4)	901 (62.8)
Race: n (%)			
White	708 (73.5)	346 (73.5)	1054 (73.5)
Asian/Pacific	161 (16.7)	87 (18.5)	248 (17.3)
Hispanic	67 (7.0)	26 (5.5)	93 (6.5)
Black	22 (2.3)	9 (1.9)	31 (2.2)
Native American	0 (0.0)	0 (0.0)	0 (0.0)
Other	5 (0.5)	3 (0.6)	8 (0.6)
Weight (kg): n (%)			
< 55	218 (22.6)	111 (23.6)	329 (22.9)
55 to < 65	245 (25.4)	129 (27.4)	374 (26.1)
65 to < 75	240 (24.9)	110 (23.4)	350 (24.4)
≥ 75	259 (26.9)	121 (25.7)	380 (26.5)
Missing	1 (0.1)	0 (0.0)	1 (0.1)
Body Mass Index			
Mean (SD)	25.20 (4.39)	25.04 (4.30)	25.15 (4.36)
Median	24.85	24.65	24.78
Min, Max	13.22, 49.60	14.69, 42.81	13.22, 49.60
Missing	5 (0.5)	3 (0.6)	8 (0.6)
Type of residence: n (%)			
Lives alone	34 (3.5)	30 (6.4)	64 (4.5)
Lives with caregiver	780 (81.0)	365 (77.5)	1145 (79.8)
Lives with relative, friend	97 (10.1)	45 (9.6)	142 (9.9)
Senior residence or retirement home	14 (1.5)	5 (1.1)	19 (1.3)
Assisted living facility	20 (2.1)	20 (4.2)	40 (2.8)
Intermediate nursing care facility	3 (0.3)	0 (0.0)	3 (0.2)
Skilled nursing care facility	7 (0.7)	2 (0.4)	9 (0.6)
Other	8 (0.8)	4 (0.8)	12 (0.8)
Pre-study dose IR duration of donepezil in Weeks			
Mean (SD)	112.17 (108.18)	104.76 (98.98)	109.73 (105.27)
Concomitant memantine use: n (%)			
Memantine	352 (36.6)	168 (35.7)	520 (36.3)
None	611 (63.4)	303 (64.3)	914 (63.7)
CYP2D6 phenotype: n (%)[a]			
Analyzed (n)	543[b]	266	809
Extensive metabolizer	502 (92.4)	248 (93.2)	750 (92.7)
Poor metabolizer	30 (5.5)	12 (4.5)	42 (5.2)
Ultra-rapid metabolizer	11 (2.0)	6 (2.3)	17 (2.1)
$APOE_4$ genotype: n (%)[a]			
Analyzed (n)	544	266	810
Average risk (E2/E2, E2/E3, E3/E3)	230 (42.3)	107 (40.2)	337 (41.6)
Increased risk (E2/E4, E3/E4)	240 (44.1)	128 (48.1)	368 (45.4)
High risk (E4/E4)	74 (13.6)	31 (11.7)	105 (13.0)

Note: The above table was copied from page 83 of the sponsor's study report

The two treatment groups were comparable with respect to the distribution of SIB, ADCS-ADL, MMSE, QoL-AD by patients and caregivers, EQ-5D, and CIBIS+ scores as shown in Table 4.

Table 4 Baseline Efficacy Assessments: Safety Population

	Donepezil SR 23 mg	Donepezil IR 10 mg	Total
Number of patients / number of caregivers	963 / 984	471 / 486	1434 / 1470
SIB total score	n=963	n=471	n=1434
Mean (SD)	74.3 (17.55)	75.4 (16.47)	74.7 (17.20)
Median	81.0	82.0	81.0
Min, Max	4, 97	4, 91	4, 97
ADCS-ADL total score	n=962	n=471	n=1433
Mean (SD)	34.1 (10.98)	34.5 (11.23)	34.3 (11.06)
Median	36.0	36.0	36.0
Min, Max	1, 54	3, 54	1, 54
MMSE total score	n=963	n=471	n=1434
Mean (SD)	13.1 (4.99)	13.0 (4.75)	13.1 (4.91)
Median	14.0	14.0	14.0
Min, Max	0, 22	0, 20	0, 22
QoL-AD total score by patients	n=898	n=433	n=1331
Mean (SD)	34.4 (7.43)	34.1 (7.22)	34.3 (7.36)
Median	35.6	34.7	35.0
Min, Max	13, 52	13, 52	13, 52
QoL-AD total score by caregivers	n=954	n=464	n=1418
Mean (SD)	31.1 (6.75)	30.9 (6.59)	31.0 (6.70)
Median	31.0	31.0	31.0
Min, Max	13, 51	13, 50	13, 51
TES for caregivers: n (%)	n=952	n=466	n=1418
1 (Very helpful)	272 (28.6)	132 (28.3)	404 (28.5)
2	180 (18.9)	86 (18.5)	266 (18.8)
3 (Somewhat helpful)	435 (45.7)	221 (47.4)	656 (46.3)
4	52 (5.5)	22 (4.7)	74 (5.2)
5 (Not helpful at all)	13 (1.4)	5 (1.1)	18 (1.3)
EQ-5D index score	n=962	n=471	n=1433
Mean (SD)	0.74 (0.18)	0.75 (0.18)	0.74 (0.18)
Median	0.80	0.80	0.80
Min, Max	-0.02, 1.00	0.08, 1.00	-0.02, 1.00
CIBIS+	n=958	n=470	n=1428
Mean (SD)	4.41 (0.85)	4.39 (0.89)	4.40 (0.86)
Median	4.00	4.00	4.00
Min, Max	2.0, 7.0	2.0, 7.0	2.0, 7.0
CIBIS+: n (%)	n=958	n=470	n=1428
Normal	0 (0.0)	0 (0.0)	0 (0.0)
Borderline mentally ill	7 (0.7)	5 (1.1)	12 (0.8)
Mildly mentally ill	99 (10.3)	55 (11.7)	154 (10.8)
Moderately mentally ill	451 (47.1)	217 (46.2)	668 (46.8)
Markedly mentally ill	300 (31.3)	138 (29.4)	438 (30.7)
Severely mentally ill	97 (10.1)	53 (11.3)	150 (10.5)
Among most extremely ill	4 (0.4)	2 (0.4)	6 (0.4)

Note: This table was copied from page 87 of sponsor's study report

3.1.1.4 Sponsor's Results

The primary efficacy analyses were performed using the LOCF method to account for missing data. Analyses of the SIB and CIBIC+ were also conducted on the Observed Cases (OC) Population at endpoint and intermediate time points (Weeks 6, 12, and 18). At Week 24, patients assigned to donepezil SR 23 mg demonstrated a statistically significant superior mean change from baseline scores compared with patients randomized to donepezil IR 10 mg on the SIB. On the CIBIC+, a small numerical difference (not significant) favoring donepezil SR 23 mg was observed.

Analysis of Primary Efficacy Variables

(a) Severe Impairment Battery

SIB scores at Baseline and LS mean changes from Baseline to Week 24 (LOCF) are summarized by treatment group for the ITT Population in Table 5. At Baseline, there was no statistically significant difference in LS mean ± (SE) SIB total scores between the donepezil SR 23 mg (75.4 ± 1.01) and donepezil IR 10 mg (76.8 ± 1.15) treatment groups, (LS mean difference -1.4; 95% CI: -3.30, 0.50; p=0.1495). At Week 24 (LOCF), mean SIB scores improved among patients in the donepezil SR 23 mg group (Mean: 2.2 ± 9.38) and were nearly unchanged in the donepezil IR 10 mg group (Mean: 0.1 ± 10.63). The LS mean difference between treatments for the change from Baseline to Week 24 was 2.2 (95% CI: 1.06, 3.24; p=0.0001), indicating a statistically significant treatment benefit in favor of donepezil SR 23 mg over donepezil IR 10 mg for the SIB co-primary endpoint. Similarly, in the OC Population at Week 24, SIB scores improved among patients in the donepezil SR 23 mg group (LS Mean ± SE: 3.3 ± 0.69) and changed little in the donepezil IR 10 mg group (LS Mean ± SE: 0.9 ± 0.75). The LS mean difference between treatments for the change from Baseline to Week 24 was 2.4 (95% CI: 1.16, 3.55; p=0.0001), indicating a nominally significant treatment benefit in favor of donepezil SR 23 mg over donepezil IR 10 mg for the SIB co-primary endpoint.

Table 5 Sponsor's Analysis of Change from baseline to week 24 in SIB

Analysis	Donepezil SR 23 mg	Donepezil IR 10 mg
Primary: ITT population, LOCF		
SIB change from Baseline to Week 24	n = 907	n = 462
LS mean (SE)	2.6 (0.58)	0.4 (0.66)
p value compared to donepezil IR[a]	0.0001	
Exploratory: ITT population, OC		
SIB change from Baseline to Week 24	n = 684	n = 397
LS mean (SE)	3.3 (0.69)	0.9 (0.75)
p value compared to donepezil IR[a]	0.0001	
Concomitant Memantine Use, LOCF		
SIB change from Baseline to Week 24	n=338	n=163
LS mean (SE)	-0.2 (1.27)	-3.0 (1.36)
p value compared to donepezil IR[a]	0.0033	
No Concomitant Memantine Use, LOCF		
SIB change from Baseline to Week 24	n=569	n=299
LS mean (SE)	3.1 (0.61)	1.3 (0.72)
p value compared to donepezil IR[a]	0.0069	
Post Hoc : US Population, LOCF		
SIB change from Baseline to Week 24	n = 292	n = 141
LS mean (SE)	2.7 (0.59)	-1.2 (0.85)
p value compared to donepezil IR[a]	0.0002	
Post Hoc: MMSE Baseline Score of 3-14, LOCF		
SIB change from Baseline to Week 24	n=476	n=256
LS mean (SE)	1.1 (0.97)	-2.0 (1.08)
p value compared to donepezil IR[a]	0.0005	
Post Hoc: MMSE Baseline Score of 5-14, LOCF		
SIB change from Baseline to Week 24	n=436	n=244
LS mean (SE)	1.2 (0.96)	-1.4 (1.07)
p value compared to donepezil IR[a]	0.0034	
Post Hoc: MMSE Baseline Score of 0-16, LOCF		
SIB change from Baseline to Week 24	n=641	n=331
LS mean (SE)	1.6 (0.78)	-1.5 (0.88)
p value compared to donepezil IR[a]	<0.0001	

Source: Tables 14.2.1.1.1.1 to 14.2.1.1.1.3 and 14.2.1.1.1.10 to 14.2.1.1.1.11 and 14.2.1.1.1.14 to 14.2.1.1.1.15

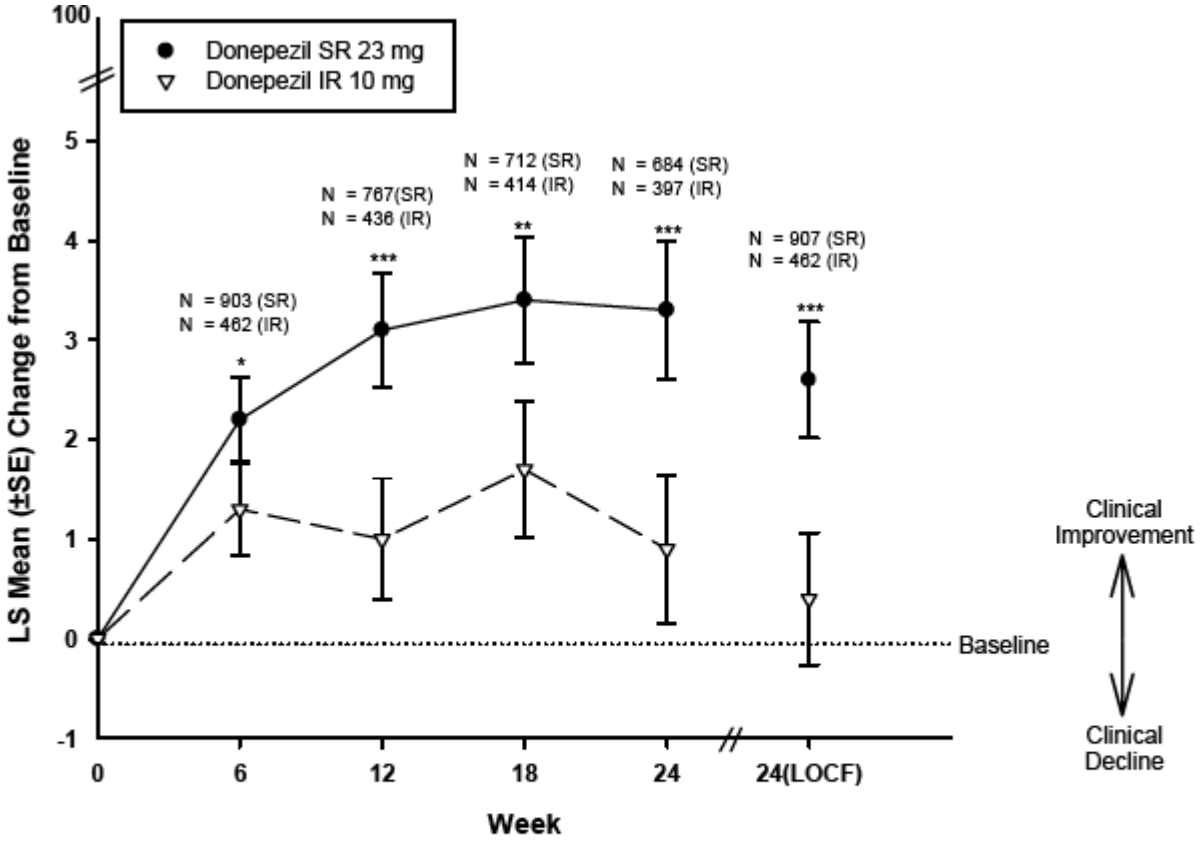
[a] Analysis method was ANCOVA model with terms for Baseline, country, and treatment.

Abbreviations: IR – immediate release; SR – sustained release; ANCOVA – analysis of covariance; ITT – intent-to-treat population; LOCF – last observation carried forward; LS – least squares; MMSE – Mini-Mental State Examination; SIB – Severe Impairment Battery.

Note: This table was copied from page 98 of the sponsor's study report

Figure 3 shows the time course for the change from Baseline in SIB scores for the two treatment groups over the 24-week study using OC data at each time point and Week 24 (LOCF) data. In both the ITT and OC Populations, the (LS mean) difference between treatments was nominally significantly different for the change from Baseline as early as Week 6 and then throughout the study duration at Weeks 12, 18, and 24, in favor of treatment with donepezil SR 23 mg.

Figure 1 Severe Impairment Battery (SIB) Total Score Change from Baseline by Visit (OC) and at Endpoint (ITT: LOCF)



Source: Tables 14.2.1.1.1.1 and 14.2.1.1.1.2
Note: * = p<0.05, ** = p<0.01, *** = p< 0.001. p-values other than LOCF are exploratory and not adjusted for multiplicity
Abbreviations: IR – immediate release; SR – sustained release; ITT – Intent-to-Treat; LOCF – last observation carried forward; LS – least squares; OC – observed cases; SE – standard error.

Note: This table was copied from page 99 of sponsor's study report

b)CIBIC+

Based on the CIBIS+ ratings, patients in the two treatment groups began the study with approximately equivalent clinical status (Table 6). At Week 24 (LOCF), mean (± SD) CIBIC+ overall change ratings were 4.23 ± 1.07 in the donepezil SR 23 mg group and 4.29 ± 1.07 in the donepezil IR 10 mg group (p=0.1789). In the OC Population, Week 24 mean (± SD) CIBIC+ overall change ratings were 4.18 ± 1.11 in the donepezil SR 23 mg group and 4.28 ± 1.09 in the

donepezil IR 10 mg group. This between-treatment difference approached the level of nominal significance (p=0.0592).

Table 6 Sponsor's Analyses of CIBIC+

Analysis	Donepezil SR 23 mg	Donepezil IR 10 mg
Primary: ITT population, LOCF		
CIBIC+ change in assessment at Week 24 overall change	n = 908	n = 459
Mean (SD)	4.23 (1.07)	4.29 (1.07)
p value compared to donepezil IR[a]	0.1789	
Exploratory: OC Population		
CIBIC+ change in assessment at Week 24 overall change	n = 682	n = 395
Mean (SD)	4.18 (1.11)	4.28 (1.09)
p value compared to donepezil IR[a]	0.0592	
Concomitant Memantine Use, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n=338	n=161
Mean (SD)	4.40 (1.02)	4.52 (0.94)
p value compared to donepezil IR[a]	0.1372	
No Concomitant Memantine Use, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n=570	n=298
Mean (SD)	4.12 (1.09)	4.16 (1.12)
p value compared to donepezil IR[a]	0.3795	
US Population, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n = 292	n = 141
Mean (SD)	4.38 (0.97)	4.57 (0.89)
p value compared to donepezil IR[a]	0.0330	
MMSE Baseline Score of 3-14, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n=478	n=254
Mean (SD)	4.37 (1.10)	4.47 (1.14)
p value compared to donepezil IR[a]	0.0508	
MMSE Baseline Score of 5-14, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n=438	n=242
Mean (SD)	4.34 (1.11)	4.45 (1.16)
p value compared to donepezil IR[a]	0.0469	
MMSE Baseline Score of 0-16, ITT-LOCF		
CIBIC+ change in assessment at Week 24 overall change	n = 642	n = 329
Mean (SD)	4.31 (1.09)	4.42 (1.10)
p value compared to donepezil IR[a]	0.0279	

Source: Tables 14.2.1.1.2.1, 14.2.1.1.2.2, 14.2.1.1.2.5, 14.2.1.1.2.12, 14.2.1.1.2.13, 14.2.1.1.2.16, 14.2.1.1.2.17

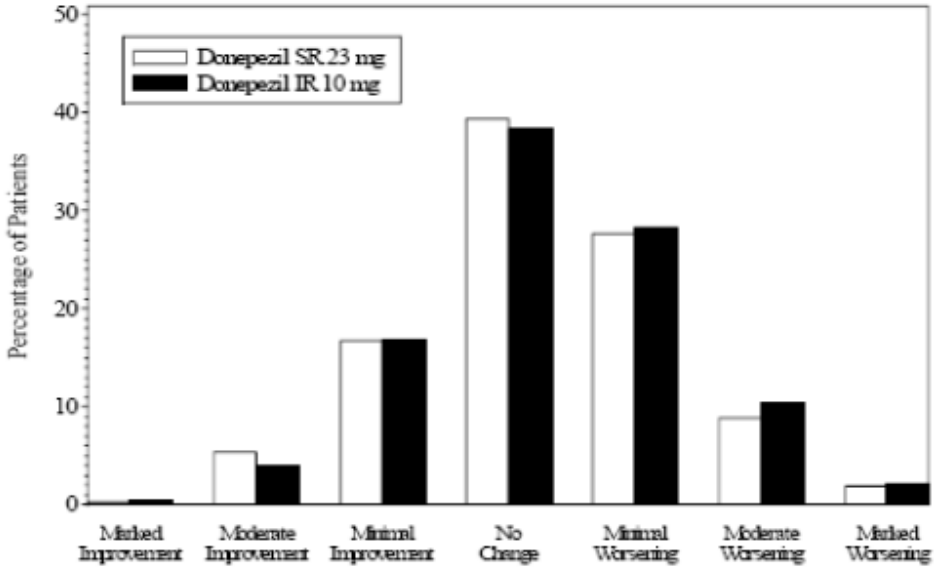
[a] Analysis method was non-parametric ANCOVA combined with a Cochran-Mantel-Haenszel test, adjusted for CIBIS+ at Baseline with a stratification adjustment for countries.

Abbreviations: IR – immediate release; SR – sustained release; ANCOVA – analysis of covariance; CIBIC+ – Clinician's Interview-Based Impression of Change (Plus Version); ITT – intent-to-treat population; LOCF – last observation carried forward; MMSE – Mini-Mental State Examination.

Note: This table was copied from page 107 of the sponsor's study report

Figure 2 shows the distribution of CIBIC+ scores at Week 24 (LOCF) for the two treatment groups. For the categorical analyses of CIBIC+ overall change at Week 24 (LOCF), 22.3% of patients in the donepezil SR 23 mg group received ratings of < 4 (improved) compared with 21.1% in the donepezil IR 10 mg group; 39.3% of patients in the donepezil SR 23 mg group received ratings of 4 (no change) compared with 38.3% in the donepezil IR 10 mg group; and 38.3% of patients in the donepezil SR 23 mg group received ratings of > 4 (worsened) compared with 40.5% in the donepezil IR 10 mg group.

Figure 2 Week 24 (or LOCF) CIBIC+ by Treatment Group



Source: Figure 14.2.6
Abbreviations: IR – immediate release; SR – sustained release; ITT – Intent-to-Treat; LOCF – last observation carried forward.
Note: This figure was copied from page 108 of the sponsor's study report

Exploratory Analyses of CIBIS+/CIBIC+: Impact of Baseline Severity

Patients with MMSE Scores of 3-14

Mean (± SD) CIBIC+ overall change scores at Week 24 (LOCF) for the donepezil SR 23 mg and donepezil IR 10 mg groups were: 4.37 ± 1.10 and 4.47 ± 1.14, respectively. There was nearly nominally significant treatment benefit in favor of donepezil SR 23 mg over donepezil IR 10 mg for the overall change CIBIC+ co-primary endpoint at Week 24 (LOCF) (p=0.0508) in the subgroup of patients with baseline MMSE scores of 3-14 (see Table 6).

Patients with MMSE Scores of 5-14

Mean (± SD) CIBIC+ overall change scores at Week 24 (LOCF) for the donepezil SR 23 mg and donepezil IR 10 mg groups were: 4.34 ± 1.11 and 4.452 ± 1.16, respectively. There was a

nominally significant treatment benefit in favor of donepezil SR 23 mg over donepezil IR 10 mg for the overall change CIBIC+ co-primary endpoint at Week 24 (LOCF) (p=0.0469) in the subgroup of patients with baseline MMSE scores of 5-14 (see Table 6).

Patients with MMSE Scores of 0-16

In patients with baseline MMSE score of 0-16, mean (± SD) CIBIC+ overall change scores at Week 24 (ITT) for the donepezil SR 23 mg and donepezil IR 10 mg groups were: 4.31 ± 1.09 and 4.42 ± 1.10, respectively (p=0.0279 in favor of the higher dose, see Table 6). Results were similar in the OC Population (4.29 ± 1.13 and 4.42 ± 1.11, respectively; p=0.0213).

CIBIS+/CIBIC+: Patients with MMSE Scores of 17-20

There was no significant difference in the overall change of CIBIC+ mean scores between the donepezil SR 23 mg and donepezil IR 10 mg treatment groups at Week 24 (p=0.5949). The overall change of CIBIC+ mean scores (± SD) from Baseline to Week 24 (LOCF) for the donepezil SR 23 mg and donepezil IR 10 mg groups were: 4.02 ± 1.01 and 3.95 ± 0.91, respectively. These data support the same conclusion drawn from the SIB data. This does not differ from the overall ITT Population with the LOCF analyses.

Secondary Efficacy Parameters: ADCS-ADL and MMSE

There was no significant difference in LS mean (± SE) ADCS-ADL total scores between the donepezil SR 23 mg (34.1 ± 0.64) and donepezil IR 10 mg (34.5 ± 0.72) treatment groups at Baseline. At Week 24, mean ADCS-ADL total scores declined among patients in both the donepezil SR 23 mg group (-1.2 ± 6.83) and the donepezil IR 10 mg group (-1.2 ± 6.78). The LS mean difference between treatments for the change from Baseline to Week 24 was not statistically significant (LS mean difference = -0.1, 95% CI: -0.81, 0.69; p=0.8822). Results were similar in the OC Population.

At Baseline, there was no significant difference in MMSE total scores between the donepezil SR 23 mg (LS Mean ± SE: 13.5 ± 0.28) and donepezil IR 10 mg (LS Mean ± SE: 13.6 ± 0.32) treatment groups; LS mean difference between treatments of -0.1 (95% CI: -0.60, 0.47; p=0.8077). At Week 24, mean MMSE ± SD total score was numerically higher in the donepezil SR 23 mg group (0.6 ± 2.93) than in the donepezil IR 10 mg group (0.4 ± 3.20), but the LS mean difference of 0.2 (95% CI: -0.14, 0.53; p=0.2443) was not significantly different.

3.1.1.5 Reviewer's Results

3.1.1.5.1 Primary Analysis of the Co-primary SIB

The primary analysis included 462 from the 10 mg IR group and 907 from the 23 mg SR group because 24/486 (4.9%) of those randomized to 10 mg IR and 74/981 (7.5%) of 23 mg SR had no post-baseline efficacy data available.

This reviewer verified the statistical significance of the sponsor's primary analysis of the change from baseline in SIB at week 24, using LOCF where applicable.

3.1.1.5.2 Assessment of the Impact of Missing Data

There was more missing week 24 SIB data in the 23 mg SR group than in the 10 mg IR group (24% vs. 13%) and the 23 mg dropouts also tended to dropout earlier than the 10 mg group. As noted by the sponsor and verified by this reviewer the analysis of week 24 SIB change using the Observed Cases population supported the primary result. The average last post-baseline change from baseline in SIB for non-completers was -1.85 for Donepezil IR 10 mg and 0.38 for Donepezil SR 23 mg. For completers the average week 24 SIB change from baseline was 0.39 for 10 mg and 2.82 for 23 mg. The group difference between dropouts was 2.23 and between completers was 2.43. A mixed model for repeated measures (MMRM) data model conducted by this reviewer yielded a result similar to those for LOCF and OC. This model adjusted for scheduled visit and treatment by visit interaction in addition to the other effects included in the primary ANCOVA model. For the MMRM model the within patient covariance structure was assumed to be as general as possible, i.e., "unstructured".

Another way to assess the impact of the dropouts is assuming the worst case with respect to the Week 24 SIB for dropouts. First, note that a Wilcoxon rank sum test of the LOCF change from baseline yields a p-value of 0.006 in favor of Donepezil 23 mg which is similar to that obtained by the primary analysis method. However, if we assume the worst case for dropouts, by assigning the worst rank for the change in SIB at week 24, then the test is no longer nominally significant, p=0.3423. Thus, it appears that the handling of dropouts may affect the analysis results for the SIB. The same is true (p=0.3388) for a Rank ANCOVA analysis which, unlike the Wilcoxon test, permits adjustments for country and baseline SIB as used in the primary analysis. In fact, this analysis suggests that the 23 mg group is numerically worse but not nominally significantly worse than the 10 mg group.

Another exploratory less extreme imputation method uses the multiple imputation method. The imputations made by this reviewer were based on a regression model assuming that week 24 change for the dropouts depends only on the baseline SIB value. The imputations are drawn from the observed data, but are restricted to be below a certain quantile of the Week 24 SIB change distribution, which makes this a non-missing at random (missing data) model. Each complete data set is analyzed after imputation and supposing that we fill in all the missing data 10 separate times we then average the resulting 10 separate analyses to summarize all of the multiple imputations. The point of this is to account for the variability across sets of imputations. If we do this and choose the imputations to be at or below the 50^{th} percentile (or lower) of the Week 24 SIB change the resulting analysis of the imputed data does not find a significant treatment effect, LS Mean diff=0.92, p=0.087. However, if we only exclude the best 30% of the observed SIB Week 24 changes from being potential imputations then the result is significant LS Mean diff=1.19, p=0.0304. These sensitivity analyses like the worst case analysis suggest that the missing data could potentially alter the significance of the primary result if it were known. The sign of the difference does numerically favor the 23 mg group, so assuming the study had assay sensitivity since the 10 mg dose was previously shown to be effective we may not have to worry too much about this missing data problem for the SIB data.

3.1.1.5.3 Primary Analysis of Co-Primary CIBIC+

It is interesting to note that the three studies that were the basis for extending the indication of Donepezil 10 mg to moderate to severe Alzheimer's (see NDA 20690) had mixed results for the treatment comparison of the CIBIC+ or other global rating outcome. In study A2501017 the p-value for the Clinician's Global Impression (CGI) was 0.0547 (note that the CGI was not a co-primary endpoint in this study; the corresponding observed treatment difference was 0.3). For study 315, the p-value for the CIBIC+ was 0.0905 (a corresponding mean difference of 0.2) and in this case the CIBIC+ was a co-primary endpoint. In study 231, which only included Japanese patients, the p-value for the co-primary CIBIC+ was 0.007 (10 mg vs. placebo difference of 0.4). Because there were three studies two of which demonstrated significance on their prespecified co-primary analyses this presence of some inconsistent results on the CIBIC+ (or other global) endpoint was not a serious issue for that application.

3.1.1.5.4 Sensitivity Analyses of CIBIC+ data

As the sponsor indicated, the overall result for the CIBIC+ at week 24 (or LOCF) was not statistically significant. Exploratory analyses of the CIBIC+ at earlier timepoints did not show any group differences reaching nominal significance either. The mean CIBIC+ is displayed across time in Figure 3. Notice that the LOCF mean is higher than the week 24 mean despite the upward trend over time. This is due to the fact that at each time dropouts tended to have scores that were above the mean for completers, as seen in Table 7. The average last post-baseline CIBIC+ for non-completers was 4.37 for Donepezil IR 10 mg and 4.36 for Donepezil SR 23 mg. For completers the average week 24 CIBIC+ was 4.28 for 10 mg and 4.18 for 23 mg. Because the treatment group difference was smaller between dropouts than completers the observed cases result may be biased in favor of Donepezil 23 mg.

A higher proportion of the 23 mg group dropped out and they also tended to dropout earlier than the 10 mg group. Given the observed trend towards worsening over time in both groups (see Figure 3) this also could bias in favor of 23 mg. A Wilcoxon test stratified by time of last assessment leads to the same conclusion as the primary analysis, giving a p-value of 0.145 for the treatment comparison of CIBIC+.

Figure 3 Mean CIBIC+ by Treatment Group over Time

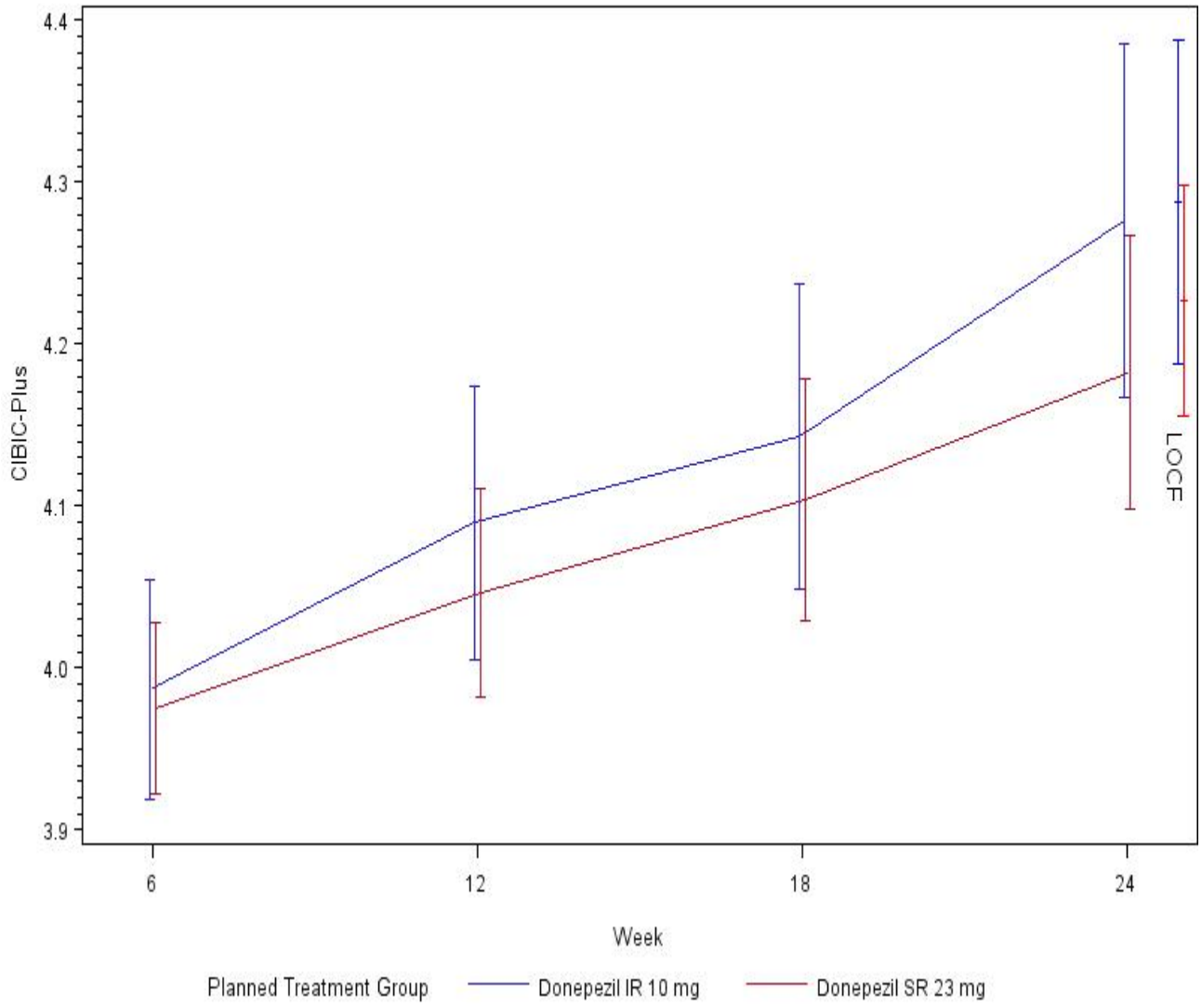


Table 7 Mean CIBIC+ by Completion Status Over Time

		Planned Treatment Group			
		Donepezil IR 10 mg		Donepezil SR 23 mg	
		N	Mean CIBIC+	N	Mean CIBIC+
Week	Completer				
6	No	60	4.27	215	4.17
	Yes	397	3.94	676	3.91
12	No	42	4.45	91	4.41
	Yes	392	4.05	678	4.00
18	No	16	4.31	37	4.51
	Yes	397	4.14	676	4.08
24	Yes	399	4.28	685	4.18
All		1703	4.12	3058	4.07

The sponsor highlighted the fact that the difference in CIBIC+ although not significant overall, reached nominal significance in the patients from the U.S.

This reviewer discovered that there were numerous differences between U.S. and non-U.S randomized patients at baseline. It is possible that some of these differences could affect the generalizability of the post-hoc U.S. subgroup result, even if we were to assume it was not due to chance alone.

- In U.S. patients the baseline SIB was 3.49 points higher in the Donepezil 10 mg IR group than in the 23 mg SR group, a baseline severity difference which was not too far from nominal significance, p=0.072.
- There was a dramatic difference in the use of stable Memantine at baseline:75.3% in the U.S. as compared with 18.7% outside the U.S.

- Baseline severity as measured by the CIBIS+ was also significantly higher in the U.S. than in the non-U.S. (4.37 vs. 4.49, p=0.0194).
- APOE risk status tended to be higher in the U.S than non-U.S. (17.3% vs. 10.6% had "greatly increased risk").
- A higher proportion of White (88% vs. 68%) and a lower proportion of Asian/Pacific (0.2% vs. 24.2%) patients in the U.S. than in the non-U.S..
- Education background was different for U.S. and non-U.S. : 5% vs. 42% had 0-8 years of education.
- The high dose group had a numerically higher proportion of patients residing with the caregiver (80.1 vs. 73.8%) and a lower proportion residing in an assisted living facility without skilled nursing (4.5 vs. 11.3%) than the low dose group.

The nominal significance of the U.S. subgroup result was also sensitive to handling of missing data as follows. The proportion of US patients that completed the trial was 87.2% for 10 mg IR and 73.6% for 23 mg SR. If we assume the worst possible CIBIC+ week 24 outcome for those who dropped out then the mean CIBIC+ becomes 4.64 for 10 mg IR and 4.87 for 23 mg SR. This numerically favors 10 mg IR. Also, in U.S. patients a Wilcoxon test stratified by time of last assessment gives a p-value of 0.089 for the treatment comparison of CIBIC+. Thus, there is some sensitivity of the U.S. subgroup result to the handling of dropouts.

In addition, it was found that if we adjust for race in the U.S. subgroup, but otherwise do the analysis corresponding to the primary analysis, the p-value for a treatment difference on CIBIC+ is 0.0633.

These alternative analyses reveal a lack of robustness of the **post-hoc** U.S. subgroup result.

3.1.1.5.5 Interim Analysis Plan

Originally, an interim analysis was planned after 400 patients had completed the study. As described earlier the sponsor reported that they decided not to do the interim analysis for efficacy because of faster than expected enrollment. This reviewer determined that about 51% of patients had been randomized and about 20% had completed by the time of protocol amendment 4 which removed the possibility of stopping early for efficacy at the interim. Also, 100% had provided consent and 82% had been randomized by the time the first 400 completed or withdrew. If an interim analysis for efficacy had been done after 405 patients completed (>400 to account for patients randomized on the same day as the 400^{th} patient) this reviewer found that the interim results would have been as follows. The conditional power at the interim for the CIBIC-plus was calculated to be 88% based on revising only the estimate of the variance and 69% based on revising both the variance and the group difference according to the observed interim results. For ease of calculation these conditional power calculations were based on a t-test rather than the nonparametric primary analysis method for the CIBIC-plus.

Treatment Difference at Originally Planned Interim (N=405)

Week 24 SIB Change 5.41 +/- 1.11 (S.E.), $p<0.0001$

CIBIC-plus $p=0.1410$

3.2 Evaluation of Safety

Safety was not reviewed here please see the medical review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

4.1.1 Gender

About sixty three (62.8%) percent of randomized patients were female.
Treatment effects on the SIB change from Baseline to week 24 appear to be roughly consistent across gender (Table 8).

Table 8 SIB Change from Baseline to week 24 by Gender in MITT-LOCF Population

		GROUP FEMALE			GROUP MALE			ALL		
TREAT	N	Baseline MEAN (SD)	MEAN (SD)	Pvalue 23 vs. 10	N	Baseline MEAN (SD)	MEAN (SD)	Pvalue 23 vs. 10	N	Mean (SD)
10 mg	287	75.3 (17.0)	-0.1 (10.7)	0.001	175	76.0 (15.1)	0.5 (10.5)	0.027	462	0.1 (10.7)
23 mg	572	74.2 (18.0)	1.9 (9.4)	.	335	74.2 (18.0)	2.7 (9.4)	.	907	2.2 (9.4)

Gender by Treatment Interaction test p= 0.7809

For the Week 24 CIBIC+ treatment group comparison there was a hint of an effect in males, but this may be due to chance alone since overall there was no statistically significant effect (Table 9).

Table 9 Week 24 (or LOCF) CIBIC+ by Gender in MITT Population

		GROUP FEMALE			GROUP MALE				ALL	
TREAT	N	Baseline MEAN (SD)	MEAN (SD)	Pvalue 23 vs. 10	N	Baseline MEAN (SD)	MEAN (SD)	Pvalue 23 vs. 10	N	Mean (SD)
10mg	287	4.41 (0.90)	4.25 (1.08)	.	172	4.41 (0.90)	4.34 (1.07)	.	459	4.3 (1.1)
23mg	574	4.44 (0.85)	4.30 (1.04)	0.796	334	4.39 (0.86)	4.11 (1.12)	0.008	908	4.2 (1.1)

Gender by Treatment Interaction test based on ANCOVA p=0.0155

4.1.2 RACE

About 74.2% of randomized patients were White, 16.6% were Asian, 6.6% were Hispanic, 2.1% were Black, and 0.5% were classified as Other. The observed treatment group differences in the change from baseline to week 24 (or LOCF) in SIB appeared to be reasonably consistent over the different race groups in the study (Table 10).

Table 10 SIB Change from Baseline to week 24 by Race in MITT-LOCF Population

RACE	STATISTIC	10 MG	23 MG
WHITE	n	339.	677.
	Baseline Mean (SD)	76.3 (16.0)	74.6 (17.2)
	Change Mean (SD)	0.5 (10.3)	2.2 (9.4)
	p-value 23 vs. 10		0.004
ASIAN	n	85.	143.
	Baseline Mean (SD)	73.1 (18.1)	72.3 (18.0)
	Change Mean (SD)	-0.4 (11.2)	1.9 (9.4)
	p-value 23 vs. 10		0.086
OTHER	n	38.	87.
	Baseline Mean (SD)	74.8 (14.5)	73.6 (19.7)
	Change Mean (SD)	-2.2 (12.1)	2.9 (9.0)
	p-value 23 vs. 10		0.005
ALL	n	462.	907.
	Baseline Mean (SD)	75.6 (16.3)	74.1 (17.6)
	Change Mean (SD)	0.1 (10.6)	2.2 (9.4)

Treatment By Race Interaction test p= 0.2347

There was no evidence of an interaction between treatment and race on the Week 24 (or last post-baseline) CIBIC+. Like the overall result for the Week 24 CIBIC+ there were no nominally significant treatment group differences within any race subgroups (Table 11).

Table 11 Week 24 (or LOCF) CIBIC+ by Race in MITT Population

RACE	STATISTIC	10 MG IR	23 MG SR
WHITE	n	338.	677.
	Baseline Mean (SD)	4.36 (0.89)	4.41 (0.88)
	Change Mean (SD)	4.37 (1.01)	4.31 (1.04)
	p-value 23 vs. 10	0.275	.
ASIAN	n	84.	143.
	Baseline Mean (SD)	4.41 (0.89)	4.45 (0.73)
	Change Mean (SD)	3.93 (1.14)	3.92 (1.13)
	p-value 23 vs. 10	0.793	.
OTHER	n	37.	88.
	Baseline Mean (SD)	4.55 (0.83)	4.44 (0.84)
	Change Mean (SD)	4.35 (1.32)	4.06 (1.11)
	p-value 23 vs. 10	0.162	.
ALL	n	459.	908.
	Baseline Mean (SD)	4.38 (0.89)	4.42 (0.85)
	Change Mean (SD)	4.29 (1.06)	4.22 (1.06)

Treatment by Race Interaction test based on ANCOVA p= 0.5167

4.1.3 AGE

Ages ranged from 47 to 90. The average age was 73.8, while the median age was 75.

Assuming a linear functional relationship between change in SIB at Week 24 and age (analyzed as a continuous variable) there is some evidence that the treatment group differences vary significantly with age (interaction test: p=0.0217), in particular, there are bigger differences for lower ages (Table 12). However, because the differences all favored 23 mg SR over 10 mg IR the interaction is not a big concern.

Table 12 SIB Change from Baseline to week 24 by Age Group in MITT-LOCF Population

AGE GROUP	STATISTIC	10 MG	23 MG
< 65	N	82.	142.
	Baseline: MEAN (S.D.)	71.3 (19.2)	69.4 (19.1)
	Change: MEAN (S.D.)	-2.1 (10.2)	1.8 (9.0)
	p-value 23 vs. 10	.	0.0028
65-74	N	127.	290.
	Baseline: MEAN (S.D.)	73.9 (17.5)	72.6 (18.6)
	Change: MEAN (S.D.)	-0.6 (13.3)	2.1 (10.8)
	p-value 23 vs. 10	.	0.0215
75-84	N	215.	404.
	Baseline: MEAN (S.D.)	77.2 (14.7)	76.5 (16.1)
	Change: MEAN (S.D.)	1.1 (9.4)	2.6 (8.7)
	p-value 23 vs. 10	.	0.0552
85-90	N	38.	71.
	Baseline: MEAN (S.D.)	81.8 (9.6)	76.4 (16.2)
	Change: MEAN (S.D.)	1.6 (6.7)	1.7 (7.8)
	p-value 23 vs. 10	.	0.7034
All Ages	N	462.	907.
	Baseline: MEAN (S.D.)	75.6 (16.0)	74.1 (17.4)
	Change: MEAN (S.D.)	0.1 (10.6)	2.2 (9.4)

Note: Interaction test between Treatment and Age Group p= 0.3891

There was no evidence of an interaction between treatment and age group (or continuous age) on the Week 24 (or last post-baseline) CIBIC+. Like the overall result for the Week 24 CIBIC+ there were no nominally significant treatment group differences within any age subgroups (Table 13).

Table 13 Week 24 (or LOCF) CIBIC+ by Age Group in MITT Population

AGE GROUP	STATISTIC	10 MG IR	23 MG SR
< 65	N	82.	142.
	Baseline: MEAN (S.D.)	4.45 (0.83)	4.51 (0.92)
	Change: MEAN (S.D.)	4.18 (1.35)	4.13 (1.03)
	p-value 23 vs. 10	0.413	.
65-74	N	125.	290.
	Baseline: MEAN (S.D.)	4.39 (0.95)	4.47 (0.84)
	Change: MEAN (S.D.)	4.29 (1.20)	4.24 (1.19)
	p-value 23 vs. 10	0.726	.
75-84	N	214.	405.
	Baseline: MEAN (S.D.)	4.35 (0.89)	4.35 (0.84)
	Change: MEAN (S.D.)	4.33 (0.91)	4.23 (0.97)
	p-value 23 vs. 10	0.208	.
85-90	N	38.	71.
	Baseline: MEAN (S.D.)	4.37 (0.79)	4.43 (0.83)
	Change: MEAN (S.D.)	4.26 (0.76)	4.31 (1.18)
	p-value 23 vs. 10	0.896	.
All Ages	N	459.	908.
	Baseline: MEAN (S.D.)	4.38 (0.89)	4.42 (0.85)
	Change: MEAN (S.D.)	4.29 (1.07)	4.22 (1.07)

Age by Treatment Interaction test based on ANCOVA p= 0.9025

4.2 Other Special/Subgroup Populations

4.2.1 Individual Sites

Figure 4 shows the observed treatment group differences (based on observed means) by individual study site for the change from baseline in SIB to week 24 (or LOCF). The exclusion of data from any single site did not alter the significance of the treatment difference result for the change from baseline in SIB to week 24 (or LOCF). Note that in the figure the curve that contains most of the bubble symbols indicates roughly the level needed for nominal significance as a function of sample size. The size of the bubbles is proportional to the sample size at the given site.

Figure 4 Treatment Group Differences in Week 24 (LOCF) SIB Change by Individual Site

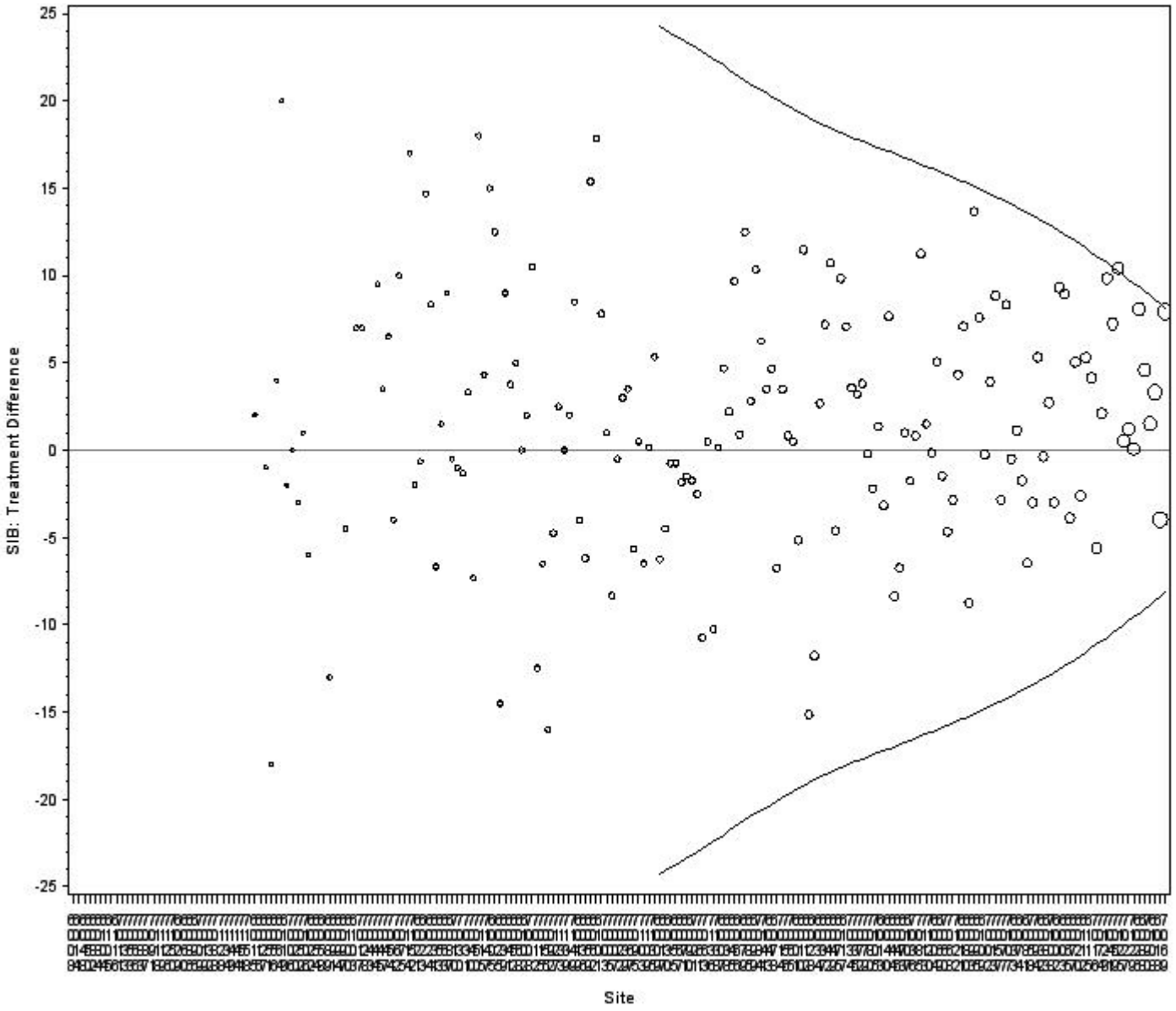


Figure 5 shows the observed treatment group differences (based on observed means) by country for the change from baseline in SIB to week 24 (or LOCF). The USA and Chile had nominally significant differences. If we combine the non-US countries and then test for an interaction between country (U.S. vs. non-U.S.) and treatment we obtain a p-value of 0.0539. The estimated difference in the U.S. was 3.81. The estimated treatment difference in the combined non-U.S. countries was estimated to be 1.48, which is nominally significant, p=0.0285. Therefore, although the observed effect was numerically bigger in the U.S. than in the combined non-US, both region's results were nominally significant favoring 23 mg SR. So if there is in fact an interaction between treatment and U.S. vs. non-U.S. on SIB change at Week 24 since both U.S. and non-U.S. have nominally significant results it is not a real concern.

Figure 5 Treatment Group Differences in Week 24 (LOCF) SIB Change by Individual Site

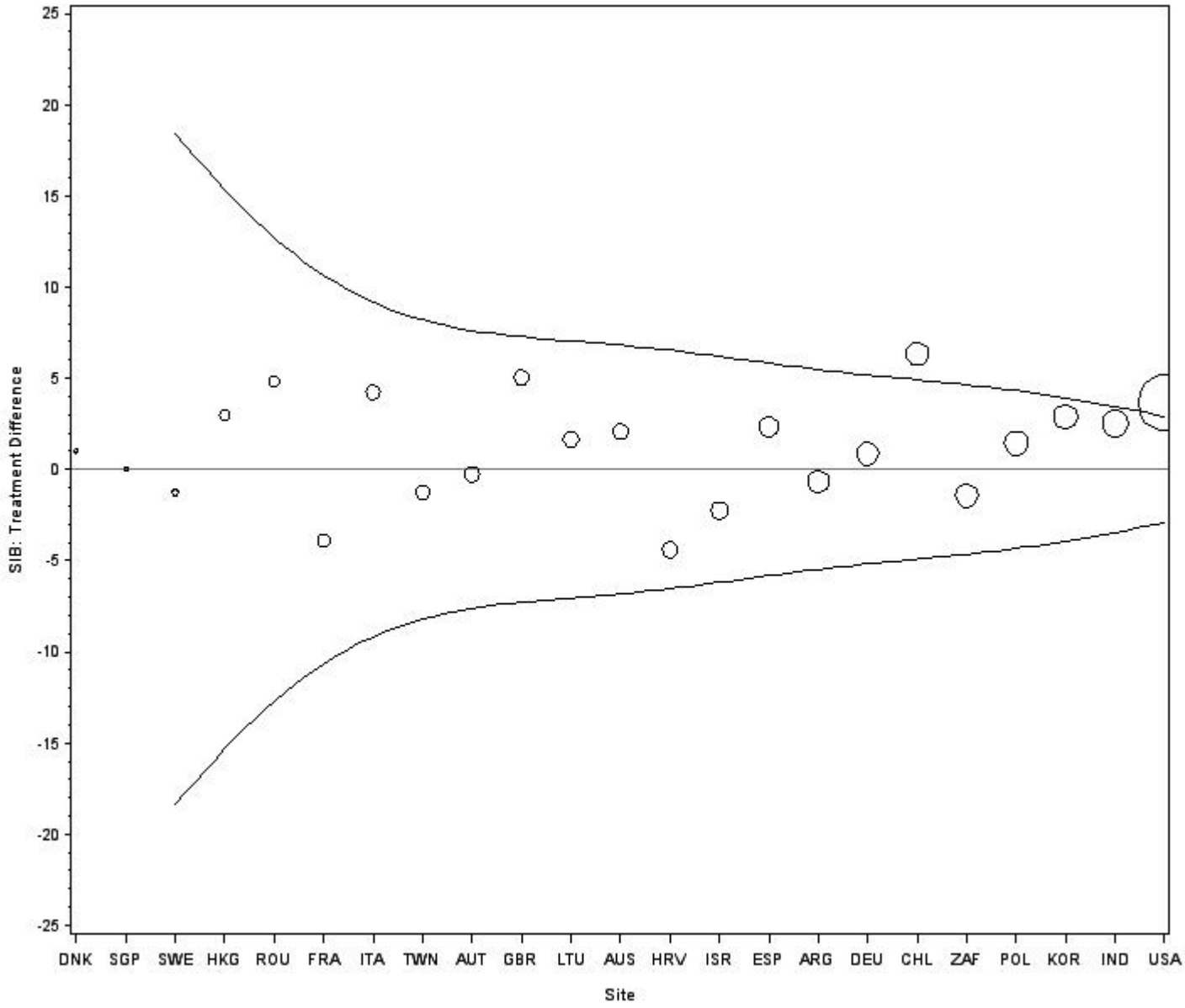


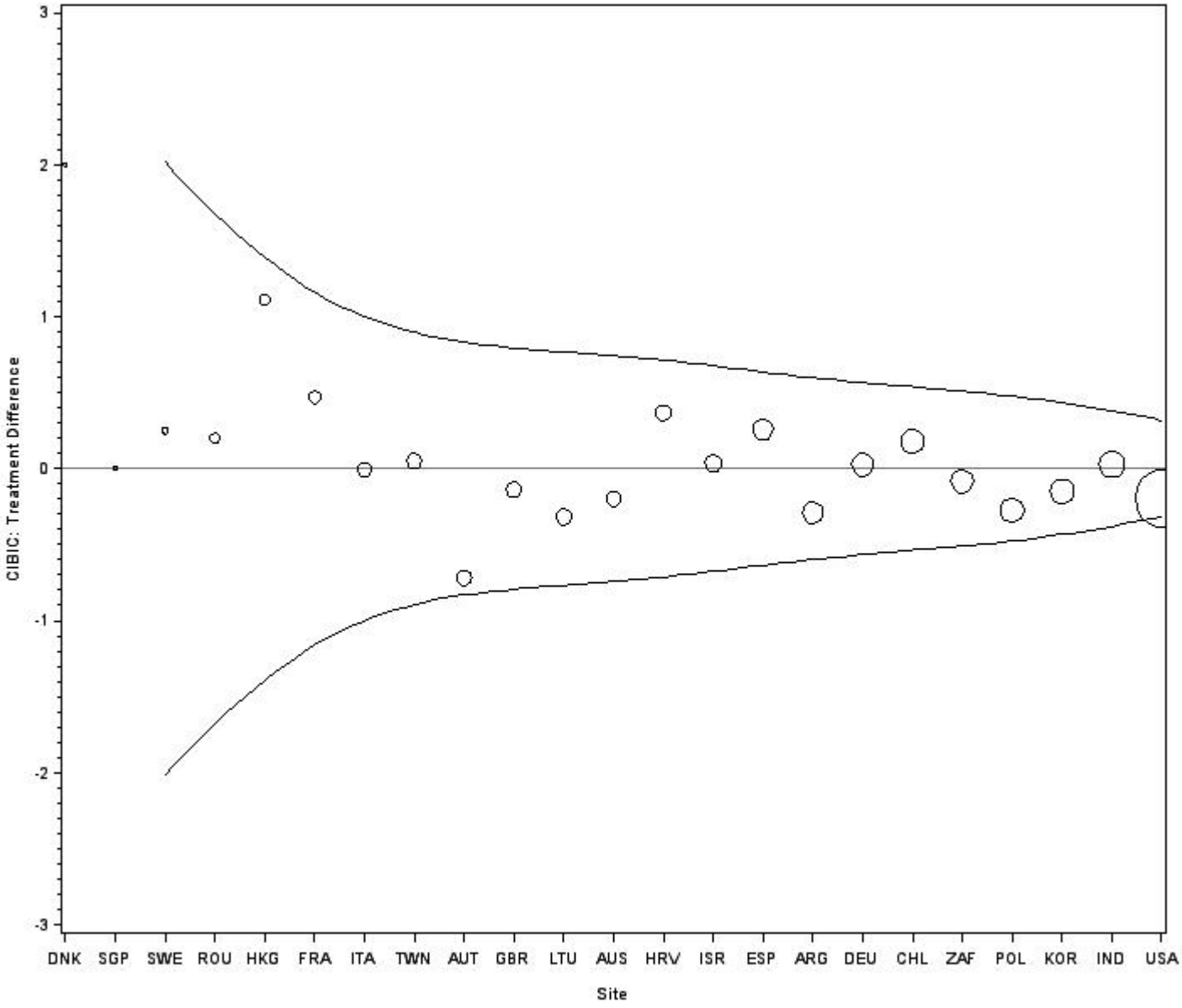
Figure 6 shows the observed treatment group differences (based on observed means) by country for the Week 24 (or LOCF) CIBIC+.

Twelve (12) countries had a mean CIBIC+ that was lower in the 10 mg IR group, 1 had equal means in both groups, and 10 had a lower mean in the 23 mg SR group. USA and AUT had nominally significant results favoring 23 mg SR for the treatment difference in CIBIC+ at week 24 and none were nominally significant favoring 10 mg IR.

The Probability of 2 or more out of 23 countries having a nominally significant result on the CIBIC+ when no difference really exists is 0.321. A few of the 23 countries had sample sizes that were too small for it to be even possible to observe a statistically significant difference at the 0.05 level. If we restrict to countries with a certain minimum number of patients we find the following.

The Probability of 2 or more out of 20 countries with 5 or more patients in each group having a nominally significant result on the CIBIC+ when no difference really exists is 0.283.
The Probability of 2 or more out of 17 countries with 8 or more patients in each group having a nominally significant result on the CIBIC+ when no difference really exists is 0.230.
From the size of these probabilities we see that it is certainly possible to observe 2 nominally significant treatment differences from the countries in the study when all true differences are zero. However, it would be more likely to observe 0 or 1 significant difference if all differences are truly zero.

Figure 6 Observed Mean Treatment Group Difference in Week 24 (or LOCF) CIBIC+ by Country



5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Overall, the treatment difference on the co-primary SIB change at week 24 was statistically significant ($p<0.0001$). Effects were numerically larger in the U.S. which accounted for 32% of the randomized patients. An exploratory test for an interaction between U.S. and treatment group on the change from baseline in SIB yielded a p-value of 0.0539. A similar trend was seen for the CIBIC+. The overall result for the CIBIC+ was not statistically significant but in the U.S. subgroup the exploratory result reached the nominal significance level. There were several nominally significant differences at baseline between the U.S. and non-U.S. populations which raise questions about the generalizability of the U.S. subgroup result for the CIBIC+ at week 24 even if one were to assume that it was real and not merely due to chance.

Previous trials did not demonstrate a consistently significant effect on CIBIC+ of Aricept 10 mg vs. placebo. Therefore, it would not be reasonable to consider a non-inferiority approach to show that 23 mg was superior to putative placebo for the CIBIC+. In this light, the fact that the overall co-primary CIBIC+ result comparing 23 mg to 10 mg did not reach statistical significance may be a major problem unless we can appeal to prior beliefs that the treatment effect should increase as a function of dose. Even that might be questioned if the effect of 10 mg on CIBIC+ is in doubt because the 23 mg dose might not be high enough to observe an effect. Treatment differences on other secondary endpoints, ADCS-ADL and MMSE, did not reach nominal significance either.

In the meeting minutes from the 11 April 2008 meeting the FDA questioned the appropriateness of the SIB as an endpoint for MMSEs >=15 saying that they were not fully convinced, but based on the submitted data it was, nevertheless, acceptable to proceed. This study randomized patients with MMSEs up to 22. Perhaps related to this concern, the sponsor claims nominal significance of the CIBIC+ treatment difference in a post-hoc subgroup, those with baseline MMSE scores from 0-16. This seems problematic first and foremost because it is post-hoc, but also because results for subgroups (3-14 N=728, p=0.0508; 0-14 N=769, p=0.1663 0-15 N=858, p=0.0938; 0-17, N=1063, p=0.0687) within this group do not reach nominal significance. The insignificant overall CIBIC+ result together with the lack of consistency of these additional subgroup results seem to undermine the sponsor's post-hoc result in the 0-16 subgroup.

There were significantly more dropouts in the 23 mg SR group (30%) than in 10 mg IR (18%). While the primary ITT-LOCF analysis and the secondary Observed Cases analysis of the change from baseline in SIB at Week 24 were nominally significant in favor of the high dose, significance was lost for an analysis assigning the lowest rank for dropouts and some other similar sensitivity analyses. In most cases where significance was lost the high dose was still numerically better than the low dose, but these analyses still raise the importance of the question of whether the trial had assay sensitivity. The mean SIB was between 10 and 20 points higher at baseline in this trial than in the previous Donepezil trials in which the SIB was used as an endpoint; this may further complicate cross trial comparisons and the lack of a placebo control issue.

An interim analysis after 400 patients had completed 24 months with possible stopping for efficacy or futility was originally planned. The stopping rule was to be specified in the IDMC

charter document. The final protocol amendment dated 20 Jun 2008 stated that interim efficacy would not be assessed because based on the enrollment rate seen it was likely that all patients would be enrolled before the interim analysis results would be available. The sponsor's study report also states that not interim efficacy analysis was done and, therefore, no adjustment to the alpha level for the final analysis is necessary.

5.2 Conclusions and Recommendations

This application for a new higher dose formulation of Donepezil rests on a single study which utilized the approved dose of the original formulation as the control group rather than a placebo control. The trial demonstrated a statistically significant effect on the co-primary cognitive endpoint ($p<0.001$) as determined by the pre-specified primary analysis, but the treatment difference on the co-primary global endpoint, CIBIC+, was not significant, $p=0.179$. The sponsor reported some post-hoc subgroups in which the treatment difference on the CIBIC+ was nominally significant, but this does not meet the usual standards; in addition, the results of various subgroups are inconsistent (see Sections 1.3 and 3.1.1.5), they would need to be replicated.

It is uncertain whether the trial had assay sensitivity. For example, the cognitive endpoint was 10-20 points higher on average than in the previous Donepezil studies in which it was used. The CIBIC+ was also not a good choice of co-primary endpoint in the absence of a placebo control since the earlier placebo controlled results for the low dose of Donepezil were mixed so that it would be questionable to make a non-inferiority argument. Unless there is some compelling prior reason to believe that there is a dose response between 10 mg IR and 23 mg SR the data from this trial does not seem to provide enough support for the efficacy of the 23 mg SR formulation.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22568	ORIG-1	EISAI MEDICAL RESEARCH INC	DONEPEZIL HYDROCHLORIDE

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

/s/

TRISTAN S MASSIE
07/09/2010

KUN JIN
07/09/2010
I concur with the review.

HSIEN MING J J HUNG
07/09/2010