RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART A

# The Efficacy, Safety, and Tolerability of Donepezil for the Treatment of Young Adults With Down Syndrome'

**Priya S. Kishnani,[1]* Barbara R. Sommer,[2] Benjamin L. Handen,[3] Benjamin Seltzer,[4] George T. Capone,[5] Gail A. Spiridigliozzi,[1] James H. Heller,[6] Sharon Richardson,[7] and Thomas McRae[8]**

[1]Department of Pediatrics, Duke University Medical Center, Durham, North Carolina

[2]Stanford University School of Medicine, Stanford, California

[3]University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[4]Department of Psychiatry and Neurology, Tulane University School of Medicine, New Orleans, Louisiana

[5]Division of Neurology & Developmental Medicine, Department of Pediatrics, Johns Hopkins School of Medicine, Kennedy Krieger Institute, Baltimore, Maryland

[6]Department of Surgery, Duke University Medical Center, Durham, North Carolina

[7]Eisai, Inc., Teaneck, New Jersey

[8]Pfizer Pharmaceuticals Group, New York, New York

Received 14 November 2008; Accepted 21 April 2009

The objective of our study was to assess the efficacy and safety of donepezil in young adults with Down syndrome (DS) but no evidence of Alzheimer disease (AD). A 12-week, randomized, double-blind, placebo-controlled study with a 12-week, open-label extension was conducted. The intervention consisted of donepezil (5–10 mg/day) in young adults (aged 18–35 years) with DS, but no AD. The primary measure was the Severe Impairment Battery (SIB) test and secondary measures were the Vineland Adaptive Behavior Scales (VABS), the Rivermead Behavioral Memory Test for Children, and the *Clinical Evaluation of Language Fundamentals*, Third Edition. At baseline, 123 subjects were randomly assigned treatment with donepezil or placebo. During the double-blind phase, SIB scores improved significantly from baseline in both groups, with no significant between-group differences. During the open-label phase, SIB scores in the original donepezil group remained stable; the original placebo group showed an improvement similar to that seen in the double-blind phase. VABS scores improved for donepezil, but not placebo, during the double-blind phase (observed cases, $P = 0.03$; last observation carried forward, $P = 0.07$). Post hoc responder analyses were significant for donepezil using three of five response definitions ($P \leq 0.045$). Adverse event rates were comparable to AD studies. In this first large-scale, multicenter trial of a pharmacological agent for DS, donepezil appears safe. Efficacy interpretation was limited for the primary measure due to apparent learning/practice and ceiling effects. Outcomes in post hoc analyses suggested efficacy in some, but not all subjects, consistent with phenotypic variability of DS. Additional studies are required to confirm potential benefits of donepezil in this population. 

**How to Cite this Article:**
Kishnani PS, Sommer BR, Handen BL, Seltzer B, Capone GT, Spiridigliozzi GA, Heller JH, Richardson S, McRae T. 2009. The efficacy, safety, and tolerability of donepezil for the treatment of young adults with Down syndrome.
Am J Med Genet Part A 149A:1641–1654.

**Key words:** Down syndrome; cognition; donepezil; clinical trials

## INTRODUCTION

Individuals with Down syndrome (DS) (trisomy 21) are living longer and have enhanced functional potential due to improved

*Correspondence to:
Priya S. Kishnani, M.D., Division of Medical Genetics/Administration & Research, Department of Pediatrics, Duke University Medical Center, Box 103856 DUMC, 4th Floor GSRBI, 595 LaSalle Street, Durham, NC 27710.
E-mail: kishn001@mc.duke.edu
Published online 15 July 2009 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/ajmg.a.32953

healthcare services and better integration within society. As they live longer, fuller lives, enhancing their cognitive potential is of great importance. In individuals with DS, an innate cholinergic deficit or dysfunction may occur, with abnormalities of both peripheral and central function [Sacks and Smith, 1989; Florez et al., 1990; Beccaria et al., 1998]. Studies have found exaggerated pupillary and cardiac responses to atropine [Harris and Goodman, 1968], chronotropic incompetency (the inability to increase heart rate after exercise) [Guerra et al., 2003], and gall bladder dysfunction [Tasdemir et al., 2004]. Many such studies have included children and young adults, and while not all have been consistent, they imply that the third copy of chromosome 21 is associated with some cholinergic dysfunction early in the life of individuals with DS. Given the critical role of the central cholinergic system in cognition, memory, attention, and mood, its dysfunction could be responsible for some of the cognitive limitations seen in DS.

Other evidence supporting the hypothesis of a cholinergic deficit innate to DS includes studies using the Ts16 transgenic mouse model of DS, where the animals have a complete third copy of chromosome 16, the mouse analog of human chromosome 21, but do not survive beyond birth. In this model, fetal brains showed marked cholinergic deficits compared with age-matched controls, and cultured cells from these models showed cholinergic dysfunction [Corsi and Coyle, 1991; Fiedler et al., 1994; Allen et al., 2000].

A possible cause of cholinergic deficiency in DS is through the overproduction of amyloid precursor protein (APP) via the extra copy of the amyloid precursor protein gene (*App*) [Hardy and Higgins, 1992]. Recent studies using transgenic mouse models with and without the *App* triplication found that APP seems to inhibit axonal nerve growth factor (NGF) transport in basal forebrain cholinergic neurons (BFCN), which is critical to their function and ultimate survival [Salehi et al., 2006]. Furthermore, in this animal model, APP and NGF accumulate, forming early endosomes in the axon terminals of the BFCN, similar to those seen in human fetal DS brains. Thus, regardless of the process that ultimately causes Alzheimer disease (AD) in older DS individuals, the overproduction of APP due to trisomy 21 may result in cholinergic deficit or dysfunction.

If an innate cholinergic deficit or dysfunction does contribute to the cognitive and functional limitations seen in DS, enhancing central cholinergic function is an appropriate target for treatment. Cholinesterase inhibitors (ChEIs) increase the availability of acetylcholine to post-synaptic receptors by limiting its breakdown in the synapse. There have now been many studies showing that ChEIs improve cognition and function in individuals with AD [Rosler et al., 1999; Tariot et al., 2000; Feldman et al., 2001; Winblad et al., 2001], as well as individuals with both DS and AD [Prasher et al., 2002], apparently by enhancing central cholinergic function [Wilkinson et al., 2004]. Thus, if cholinergic deficits or abnormalities are evident early in children with DS, ChEIs could be expected to improve the functional capacity of these individuals. Two human pharmacological studies evaluating effects of ChEIs on cognition in children with DS support this theory [Heller et al., 2006a; Spiridigliozzi et al., 2007]. Although the numbers were small in these two open-label studies, patterns of cognitive improvement following ChEI therapy were similar.

Donepezil is a ChEI approved for the treatment of mild, moderate, and severe AD in the general population [Rogers et al., 1998a,b; Burns et al., 1999; Winblad et al., 2001, 2006; Aricept, 2006; Black et al., 2007]. In patients with AD, donepezil has shown benefits in cognitive function that include the language domain [Winblad et al., 2006; Black et al., 2007]. This is notable because in studies of DS adults without signs of AD, improvements in language performance were observed in both a 12-week double-blind [Johnson et al., 2003] and a 24-week open-label trial of donepezil [Heller et al., 2003]. Similar language performance benefits were seen in donepezil-treated children with DS [Heller et al., 2004]. The rigor of these studies of donepezil in DS, however, was limited by their size and/or their design.

The current study was planned to address these limitations. To our knowledge, this is the first large-scale study evaluating the cognitive effects of a medication with a scientific rationale in the DS population. Moreover, it is the first large, multicenter, randomized, double-blind, placebo-controlled trial to assess the cognitive and functional efficacy and the safety of a ChEI in young adults (aged 18–35) with DS and no signs of AD. This report details the methods and results of this study along with those from a 12-week open-label extension study designed to collect additional safety data.

## METHODS

### Study Population

Eligible individuals were men and women, aged 18–35 years, and diagnosed with DS documented by karyotyping. The targeted age range was based on the limited likelihood of AD having developed in these subjects with DS [Lai and Williams, 1989]. Individuals were required to have a Severe Impairment Battery (SIB) [Panisset et al., 1994] score of $<95$ at the screening visit (to limit ceiling effects), to be in generally good health, and to be capable of participating in study assessments. Each subject needed a reliable caregiver or family member to accompany him/her to all visits. Women were documented not to be pregnant at screening, and repeat pregnancy tests were carried out at all subsequent visits. Sexually active women of childbearing potential were required to use reliable methods of birth control. No subject had experienced any particular recent change in mental state, and there was no reason to believe any patient was in the early stages of AD.

Subjects were excluded from the study if they were anticipated to need any institutional care within the 6 months of study initiation; if they had any clinically significant and/or clinically unstable medical conditions, including pre-existing AD; or if they were unable to perform the outcome measures or to comply with the protocol for any reason. Subjects with histories of active or clinically significant conditions affecting the absorption, distribution, or metabolism of the study drug, such as swallowing difficulties or inflammatory bowel disease, were also excluded.

The study was approved by the institutional review board of each institution at which subjects were enrolled. Written, informed consent was obtained from each individual and from the caregiver, who provided patient data. If the individual was unable to give written, informed consent, it was obtained from an appropriate

family member and/or a legal representative, and verbal assent from the individual was obtained in all cases.

## Study Design

This was a 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study of donepezil (5–10 mg/day) with a 12-week, open-label extension. Individuals were recruited by 24 centers across the United States that had prior experience with the target population. Training in the use of study instruments was provided for all centers by research staff from Duke University Medical Center and the University of Kentucky.

In the double-blind phase, individuals were randomly assigned in a 1:1 ratio to placebo or donepezil 5 mg/day (one 5-mg tablet) for 6 weeks followed by 10 mg/day (two 5-mg tablets) for 6 weeks. Those unable to tolerate 10 mg donepezil or two placebo tablets were temporarily given a 5-mg donepezil tablet or one placebo tablet per day, with a rechallenge to 10 mg or two placebo tablets per day within 7–10 days. If the individual could not tolerate the higher dose after rechallenge, he/she could continue the study on 5 mg donepezil or one placebo tablet per day. Assessments for treatment efficacy, safety, and tolerability occurred at baseline, at Week 6, and at Week 12. Subjects completing the double-blind phase were eligible to enter the open-label phase. In those cases, the final visit of the double-blind phase served as the baseline visit of the open-label phase.

Throughout both phases of the study, medication compliance was assessed at each 6-week visit by counting the number of tablets removed from the bottle and dividing by the number of days since the medication was dispensed. An 80% adherence rate was required at each visit for efficacy assessments to be considered fully evaluable.

## Efficacy Variables

With few previous studies evaluating the cognitive effect of pharmacological treatments in individuals with DS, there were no widely accepted scales available to serve as dependent measures. One clear consideration was to try to use developmentally appropriate scales rather than the usual scales for AD outcome studies, which would not be expected to be sensitive in this population [Heller et al., 2006b]. Please see the Appendix for a full description of, and rationale for, the outcome measures used: SIB (the primary outcome measure) [Panisset et al., 1994], the Vineland Adaptive Behavior Scales (VABS) [Sparrow et al., 1984], the Rivermead Behavioral Memory Test for Children (RBMT-C) [Wilson et al., 1989, 1991], and subtests of the *Clinical Evaluation of Language Fundamentals*, 3rd Edition (CELF-3) [Semel et al., 1995].

## Safety Variables

Safety data were collected at screening, baseline, and weeks 6 and 12 of the double-blind phase, and at weeks 6 and 12 of the open-label phase. Variables measured included incidence and severity of any adverse events (AEs), as well as those that met criteria for serious adverse events (SAEs). In addition, clinically significant laboratory test abnormalities, laboratory test values, electrocardiogram (ECG) results, vital signs, medication use, and physical and neurological examination results were collected.

## Post Hoc Analyses

In an attempt to better understand the results of the pre-specified analyses, a variety of post hoc analyses were undertaken. Groups were examined for evidence of ceiling, floor, and practice/learning effects. Correlation analyses were performed with age and body mass index (BMI) as independent variables and each week 12 last observation carried forward (LOCF) outcome as a dependent variable. In addition, treatment effects were compared for subgroups based on age (18–24, $\geq$25) and sex, a combination of both, and on baseline SIB scores using various cutoffs. Responder analyses using chi-squares were conducted using five responder definitions (see the Appendix for definitions). Finally, the SIB data were examined for evidence of the contribution to the results of regression to the mean (RTM) based on the method detailed by Yudkin and Stratton [1996].

## Statistical Assessments

Analyses of both the primary and secondary efficacy measures were conducted using the intent-to-treat (ITT) and the efficacy evaluable (EE) populations. The primary efficacy end point of the double-blind phase was the difference between the donepezil and placebo groups in the SIB least-squares (LS) mean change from baseline to week 12 of the double-blind phase in the ITT population, with the last post-baseline observation carried forward to replace any missing values. The LS mean change from the baseline of open-label treatment to week 12 of the open-label phase was also performed for each original treatment group, but no statistical comparisons were made because of the open-label nature of the data. The safety population—subjects who received at least one dose of study drug (donepezil or placebo)—was used for all safety analyses. Baseline demographic characteristics were summarized by descriptive statistics for each group, but no statistical comparisons were performed. All statistical tests of efficacy measures were two-sided unless specified otherwise. The baseline values for efficacy variables were compared using analysis of variance (ANOVA). For continuous efficacy end point variables, an analysis of covariance (ANCOVA) model was used; for comparison of efficacy values within a treatment group, a paired Student's *t*-test was used. For comparing treatment using categorical efficacy variables, a Cochran–Mantel–Haenszel test was used with the study center as the stratifying variable. No statistical tests were performed on the safety data because of the large number of variables and the likelihood of chance findings, which would be difficult to interpret.

During the open-label phase, all data were summarized by descriptive statistics, but statistical tests were not performed on the demographic and baseline characteristics, SIB data, or safety analyses.

Power calculations were based on SIB data from a study of patients with moderate-to-severe AD [Feldman et al., 2001], as no data were available for subjects with DS when the current study was designed and initiated. Eighty subjects per treatment group were estimated to provide 80% power to detect a between-group mean

difference of five points on the SIB with respect to change from baseline to week 12 end point (LOCF). Because of slow individual enrollment into the study, recruitment was stopped after 123 individuals instead of the planned 160.

## RESULTS

### Demographic and Other Baseline Characteristics

Patient numbers throughout the study are given in Figure 1. Three protocol deviations were revealed during the open-label phase of the study in two individuals. However, none of these deviations were judged to be of clinical significance, and these subjects, who had received placebo during the double-blind phase, were included in all analyses.

The treatment groups in the double-blind and the open-label phases were similar with regard to demographic characteristics (age, sex, race, and weight; Table I). They were also similar with respect to IQ score, education, use of nicotine and alcohol, marital status, childbearing potential, occupational status, education, and drug allergies.

In the double-blind phase, the donepezil group (DB-D) and the placebo group (DB-P) were comparable in terms of all baseline efficacy measures (Table II). Concomitant medication use was similar between the two groups during the double-blind phase: 87.1% in the DB-D group and 83.6% in the DB-P group. Rates of concomitant use of individual medications are shown in Table III.

During both the double-blind and open-label phases, mean treatment compliance was 97–98% for both groups, median treat-

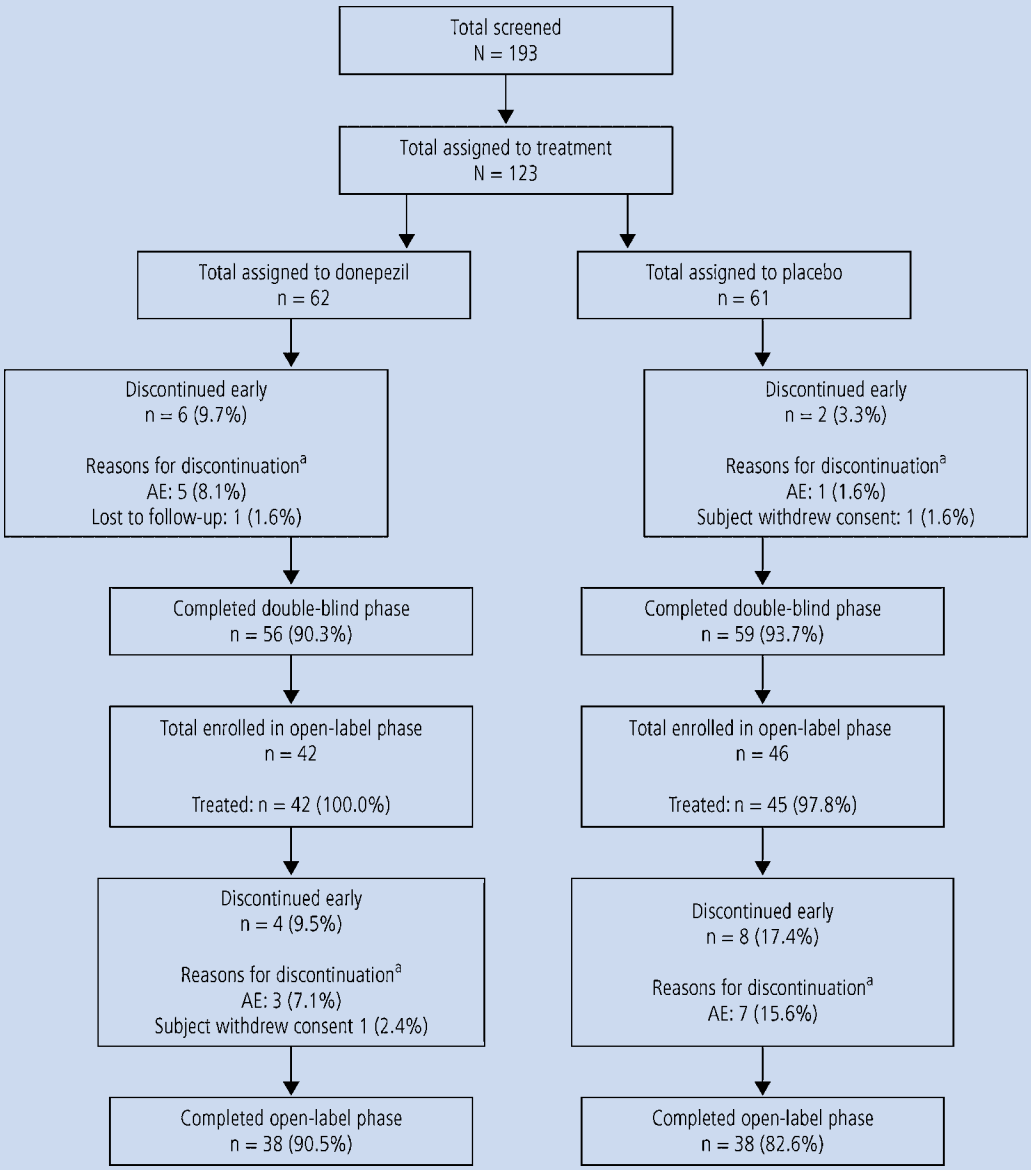


**FIG. 1. Flow diagram of patient disposition. [a]Figures are given only for patients who received study drug. AE, adverse event.**

TABLE I. Summary of Baseline Demographic Characteristics for the Safety Population

| | Double-blind phase (treatment group) | | Open-label phase (original treatment group) | |
|---|---|---|---|---|
| **Assessment** | **Donepezil (n = 62)** | **Placebo (n = 61)** | **Donepezil (n = 42)** | **Placebo (n = 45)** |
| Age (years) | | | | |
| Mean (SD) | 24.2 (5.1) | 26.0 (5.5) | 24.1 (4.8) | 26.7 (5.5) |
| Range | 18–36 | 18–38 | 18–35 | 18–36 |
| Age category (n) | | | | |
| ≤35 years | 60 | 60 | 42 | 44 |
| >35 years | 2 | 1 | 0 | 1 |
| Sex (n) | | | | |
| Male | 38 | 39 | 27 | 28 |
| Female | 24 | 22 | 15 | 17 |
| Race (n) | | | | |
| White | 54 | 58 | 37 | 45 |
| Black | 4 | 1 | 2 | 0 |
| All other | 4 | 2 | 3 | 0 |
| Weight (kg) | | | | |
| Mean (SD) | 72.9 (17.2) | 79.7 (19.3) | 74.0 (17.2) | 79.5 (18.2) |
| Range | 44–120 | 46–130 | 41–115 | 48–129 |
| BMI (mean ± SD) | 31.4 (7.15) | 33.9 (7.49) | n.a. | n.a. |

SD, standard deviation; BMI, body mass index.

ment duration was 83–84 days, and average treatment dose was approximately 7 mg/day.

## Primary Efficacy Measure

During the double-blind phase, SIB scores significantly improved (LS mean change from baseline at Week 12-LOCF) by a similar amount for both the DB-D ($P = 0.0302$) and DB-P ($P = 0.0091$) groups, with no significant between-group difference ($P = 0.71$; Fig. 2). SIB scores (mean ± SD) at baseline of the open-label phase were similar at 86.8 ± 11.53 for the original donepezil group (OL-DD; n = 41) and 87.3 ± 8.51 for the original placebo group (OL-PD; n = 43). During the open-label phase, the OL-PD group (i.e., subjects newly exposed to donepezil) showed an improvement in cognition based on the SIB (Week 12 mean change from open-label baseline) that was comparable to the improvement observed in both groups during the double-blind phase (Fig. 2). For individuals who received donepezil throughout the study (OL-DD), the improvement in SIB scores observed during the double-blind phase was maintained in the open-label phase.

## Secondary Efficacy Measures

Results from the secondary efficacy measures during the double-blind phase are summarized in Table IV. For the VABS composite supplemental norm score, the DB-D group improved significantly from baseline while the DB-P group was unchanged. There was a trend toward a significant difference between groups at week 12 (ITT-LOCF; $P = 0.07$). This difference achieved significance in the ITT-observed cases (OC) ($P = 0.03$) and EE ($P = 0.04$) populations.

On the RBMT-C, there was a trend toward improvement from baseline in the DB-D group ($P = 0.0669$), but this was not significantly different from the DB-P group ($P = 0.25$). Four subtests and one supplemental subtest of the CELF-3 were used in this study. The supplemental subtest, Rapid Automatic Color-Shape Naming, has two components—time spent, which was analyzed as mean change from baseline, and number of errors, which was analyzed categorically. The improvement on the recalling sentences subtest was significant versus baseline ($P = 0.0464$) in the DB-D group, but not significant between treatment groups (Table IV). For the Rapid Automatic Color-Shape Naming number of errors subtest, about one-third of both groups increased their number of errors, and the groups were not significantly different.

## Post Hoc Analyses

Each subject's performance-based assessment was examined for possible floor and ceiling effects. The floor for each instrument was defined in two ways: (1) the absolute minimum score (0 in all cases) and (2) as ≤ (the absolute minimum score +5% of the absolute maximum score) as measured at baseline. Similarly, the ceiling was defined as (1) the absolute maximum for each instrument and (2) as ≥ (the absolute maximum score −5% of the absolute maximum score). The results of this assessment are shown in Table V. Only the SIB had a substantial number of subjects at the ceiling for the instrument. Conversely, it had no subjects at the floor. The VABS and RBMT-C had few subjects at either the ceiling or floor. In contrast, the CELF-3 had substantial numbers of subjects at the floor but none at the ceiling.

Possible practice and learning effects were evaluated by counting the number of subjects in each group who improved from screening

TABLE II. Summary of Clinical Characteristics at Baseline for the Double-Blind Phase (ITT Population)

| | DB-D | DB-P | *P*-value |
|---|---|---|---|
| SIB | n = 59 | n = 61 | |
| Mean (SD) | 85.0 (11.86) | 86.0 (9.19) | 0.54 |
| Observed range | 44–100 | 53–99 | |
| Scale range | 0–100 | 0–100 | |
| VABS—Composite Supplemental Norm Score | n = 53 | n = 59 | |
| Mean | 57.4 | 64.1 | 0.11 |
| Observed range | 10–99 | 30–99 | |
| Scale range | 1–99 | 1–99 | |
| RBMT-C—Standardized Profile Score | n = 59 | n = 61 | |
| Mean (SD) | 10.4 (5.3) | 9.5 (4.9) | 0.41 |
| Observed range | 0–21 | 0–18 | |
| Scale range | 0–22 | 0–22 | |
| CELF-3—Raw Score | | | |
| Concepts and Directions | n = 57 | n = 60 | |
| Mean (SD) | 4.0 (4.7) | 4.1 (5.1) | 0.95 |
| Observed range | 0–17 | 0–20 | |
| Scale range | 0–30 | 0–30 | |
| Formulated Sentences | n = 59 | n = 61 | |
| Mean (SD) | 4.8 (6.1) | 4.9 (5.4) | 0.93 |
| Observed range | 0–21.5 | 0–19 | |
| Scale range | 0–44 | 0–44 | |
| Word Classes | n = 57 | n = 60 | |
| Mean (SD) | 5.7 (6.3) | 5.7 (5.3) | 0.89 |
| Observed range | 0–24 | 0–19 | |
| Scale range | 0–34 | 0–34 | |
| Recalling Sentences | n = 57 | n = 61 | |
| Mean (SE) | 5.9 (6.1) | 6.0 (5.8) | 0.93 |
| Observed range | 0–27 | 0–29 | |
| Scale range | 0–78 | 0–78 | |
| Errors in Rapid Automatic Color-Shape Naming (ITT-OC) | n = 25 | n = 24 | |
| Number with 0 errors at baseline (%) | 11 (44) | 11 (46) | 0.84 |
| Time spent in Rapid Automatic Color-Shape Naming (ITT-OC) (sec) | n = 25 | n = 24 | |
| Mean (SD) | 106.2 (46.0) | 99.0 (48.6) | 0.49 |
| Range | 50–202 | 53–272 | |

DB-D, double-blind donepezil group; DB-P, double-blind placebo group; SIB, Severe Impairment Battery; VABS, Vineland Adaptive Behavior Scales; RBMT-C, Rivermead Behavioral Memory Test for Children; CELF-3, Clinical Evaluation of Language Fundamentals, 3rd edition; ITT, intent-to-treat; OC, observed cases.

to baseline in the case of practice effects and from both screening to baseline and baseline to week 12 in the case of learning effects for the SIB, the RBMT-C, and four of the five CELF-3 subtests. These results are shown in Table VI. The occurrence of practice effects ranged from about one-third (CELF-3 Concepts and Directions, both groups) to about two-thirds (SIB, both groups) of the total sample for each test and was greater for placebo in five of six tests. Learning effects ranged from about 5% (CELF-3 Concepts and Directions placebo group) to about one-third (SIB, both groups) of the total sample for each test and was greater for placebo in four of six tests. As a percentage of the group that improved from screening to baseline, the number who also improved from baseline to week 12 ranged from 13% (CELF-3 Concepts and Directions placebo group) to about 50% (SIB, both groups); this percentage was greater for placebo in three of six tests.

Correlation analyses between age and the week 12 LOCF change score for each of the assessments except Rapid Automatic Color-Shape Naming found no substantial correlations for either group (all r-values $<0.2$). Similar results were found with BMI as the independent variable (all r-values $<0.2$).

ANCOVA analyses similar to those used for the a priori efficacy variable analyses were repeated for subgroup analyses based on age (ages 18–24 years and $\geq$25 years) and sex separately and combined. They were also run for groups using baseline SIB score cutoffs of 80, 90, and 95. None of these analyses found significant differences between treatment groups in mean changes from baseline (data not shown).

Chi-square analyses were run for five responder definitions; Table VII lists the utilized definitions of response and associated results. The definitions that excluded the SIB were both significant

TABLE III. Most Commonly Used Concomitant Medications (Classes Used by >20% and >10%, Respectively, of Subjects in Either Treatment Group)

| | Treatment group, n (%) | |
|---|---|---|
| | DB-D | DB-P |
| Classes used by >20% of subjects in either group | | |
| Antibacterial drugs | 19 (30.6) | 14 (23.0) |
| Drugs used in allergic disorders | 14 (22.6) | 15 (24.6) |
| Thyroid and antithyroid drugs | 16 (25.8) | 16 (26.2) |
| Vitamins | 23 (37.1) | 22 (36.1) |
| Classes used by >10% and <20% of subjects in either group | | |
| Analgesics | 4 (6.5) | 7 (11.5) |
| Drugs used in rheumatic diseases and gout | 6 (9.7) | 10 (16.4) |
| Respiratory corticosteroids | 3 (4.8) | 7 (11.5) |
| Systemic treatment for symptoms of UTI | 11 (17.7) | 12 (19.7) |

DB-D, double-blind donepezil group; DB-P, double-blind placebo group; UTI, urinary tract infection.

($P = 0.04$, 0.008), as was the VABS analysis ($P = 0.00042$). The definitions that included the SIB approached significance ($P = 0.09$, 0.12).

RTM analyses were performed on donepezil and placebo subgroups based on baseline SIB scores. For the donepezil subgroup with baseline SIB scores <80, the calculated RTM effect was 1.23; the observed difference between the baseline mean and the week 12 mean for this group was 1.5. For the comparable placebo group, the calculated RTM effect was 1.76, and the observed difference was 2.73. For the donepezil subgroup with SIB cutoff scores >90, the calculated RTM effect was −1.125; the observed difference at week 12 was +0.61. For the comparable placebo group, the calculated RTM effect was −1.34; the observed value was +0.20. These results suggest that for individuals in both groups with baseline SIB scores both above and below the group mean, factors in addition to the RTM effect contributed to their improvement.

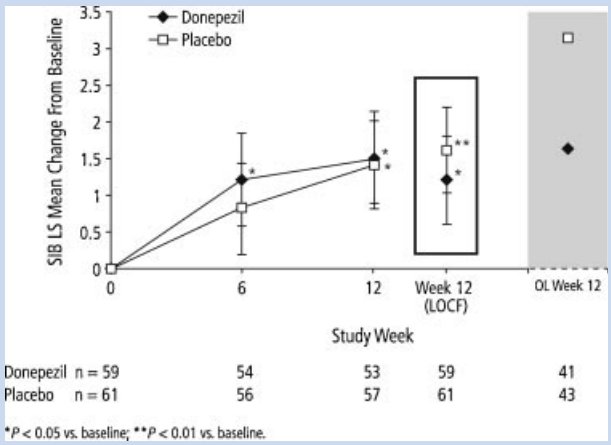


**FIG. 2. SIB scores in the double-blind phase (LS mean change from baseline ± SE) and open-label phase (mean change from baseline) for patients receiving placebo and donepezil. SIB, Severe Impairment Battery; LS, least squares; SE, standard error; LOCF, last observation carried forward; OL, open label.**

## Safety Evaluations

Table VIII summarizes the AEs reported during both phases of the study. No deaths or other SAEs were reported in either treatment group throughout the course of either the double-blind or the open-label phases. Individuals in the DB-D group were more likely to experience AEs than those in the DB-P group. Similarly, individuals in the OL-PD group were more likely to experience AEs than those in the OL-DD group. The most common AEs (those reported by >5% of donepezil-treated individuals and at least twice the rate of placebo) in the double-blind phase were asthenia, anorexia, dyspepsia, nausea, vomiting, and insomnia (Table VIII).

In the double-blind phase, more individuals receiving donepezil experienced AEs considered related to study drug (26 [41.9%]) compared with individuals receiving placebo (13 [21.3%]). The most common study drug-related AEs (those reported by >5% of donepezil-treated subjects and at least twice the rate of placebo) were abdominal pain (9.7% vs. 3.3%), nausea (6.5% vs. 0%), vomiting (8.1% vs. 1.6%), and insomnia (8.1% vs. 0%). In the double-blind phase, most AEs were transient and mild (70.1% in the DB-D group, 87.5% in the DB-P group) or moderate (26.9% in the DB-D group, 12.5% in the DB-P group) in severity. Only two AEs (hypertension and emotional lability) were reported as severe; these both occurred in the DB-D group and were considered possibly or probably related to the study drug. In both cases, the subject was discontinued from the study. In the open-label phase, four treatment-emergent AEs were rated as severe, three in the OL-DD group (decreased activities of daily living, diarrhea, and paresthesia) and one in the OL-PD group (personality disorder).

TABLE IV. Secondary Efficacy Measures: LS Mean Change From Baseline at Week 12

| | DB-D | | DB-P | | |
|---|---|---|---|---|---|
| | n | Mean ± SE | n | Mean ± SE | *P*-value |
| VABS—Composite Supplemental Norm Score | | | | | |
| ITT_LOCF | 53 | 3.4 ± 1.1 | 59 | 0.6 ± 1.1 | 0.07 |
| OC | 53 | 3.9 ± 1.1 | 59 | 0.7 + 1.0 | 0.03 |
| EE | 45 | 3.6 ± 1.2 | 54 | 0.3 + 1.1 | 0.04 |
| RBMT-C—Standardized Profile Score | | | | | |
| ITT_LOCF | 59 | 1.0 ± 0.4 | 61 | 0.3 ± 0.4 | 0.25 |
| OC | 52 | 1.2 ± 0.4 | 56 | 0.4 ± 0.4 | 0.18 |
| EE | 47 | 0.9 ± 0.5 | 55 | 0.3 ± 0.4 | 0.33 |
| CELF-3—Raw Score | | | | | |
| Concepts and Directions | | | | | |
| ITT_LOCF | 57 | −0.2 ± 0.4 | 60 | −0.2 ± 0.3 | 0.95 |
| OC | 51 | 0.0 ± 0.4 | 56 | −0.0 ± 0.4 | 0.89 |
| EE | 46 | −0.0 ± 0.4 | 55 | −0.1 ± 0.4 | 0.96 |
| Formulated Sentences | | | | | |
| ITT-LOCF | 59 | 0.2 ± 0.3 | 61 | −0.2 ± 0.3 | 0.31 |
| OC | 53 | 0.3 ± 0.4 | 57 | −0.1 ± 0.4 | 0.33 |
| EE | 48 | 0.1 ± 0.4 | 56 | −0.2 ± 0.3 | 0.55 |
| Word Classes Score | | | | | |
| ITT-LOCF | 57 | 0.8 ± 0.5 | 60 | 0.9 ± 0.5 | 0.90 |
| OC | 51 | 0.4 ± 0.5 | 55 | 0.4 ± 0.5 | 0.94 |
| EE | 46 | 0.4 ± 0.5 | 54 | 0.5 ± 0.5 | 0.91 |
| Recalling Sentences Score | | | | | |
| ITT-LOCF | 57 | 0.9 ± 0.5 | 61 | 0.2 ± 0.4 | 0.30 |
| OC | 50 | 1.3 ± 0.5 | 53 | 0.3 ± 0.4 | 0.15 |
| EE | 45 | 1.1 ± 0.5 | 52 | 0.4 ± 0.4 | 0.25 |
| Errors in Rapid Automatic Color-Shape Naming—number with 0 errors at week 12 | | | | | |
| OC | 25 | 14 (56%) | 24 | 13 (54%) | 0.84 |
| EE | 21 | 11 (52%) | 24 | 13 (54%) | 0.92 |
| Time spent in Rapid Automatic Color-Shape Naming | | | | | |
| OC | 25 | 9.0 ± 8.8 | 24 | 0.6 ± 9.2 | 0.48 |
| EE | 21 | 10.3 ± 9.9 | 24 | 0.6 ± 9.7 | 0.45 |

LS, least squares; ITT, intent-to-treat; LOCF, last observation carried forward; OC, observed cases; EE, efficacy evaluable; DB-D, double-blind donepezil group; DB-P, double-blind placebo group; VABS, Vineland Adaptive Behavior Scales; RBMT-C, Rivermead Behavioral Memory Test for Children; CELF-3, Clinical Evaluation of Language, 3rd edition.
LOCF was not done for Rapid Automatic Color-Shape Naming.

TABLE V. Percentage Ceiling and Floor Effects at Baseline

| | Donepezil, ceiling 1 | Donepezil, ceiling 2 | Donepezil, floor 1 | Donepezil, floor 2 | Placebo, ceiling 1 | Placebo, ceiling 2 | Placebo, floor 1 | Placebo, floor 2 |
|---|---|---|---|---|---|---|---|---|
| SIB | 1.7 | 20.3 | 0.0 | 0.0 | 0.0 | 16.4 | 0.0 | 0.0 |
| VABS (SN) | 1.8 | 9.1 | 0.0 | 0.0 | 5.1 | 13.6 | 0.0 | 0.0 |
| RMBT-C | 0.0 | 1.6 | 1.6 | 3.2 | 0.0 | 0.0 | 5.0 | 8.3 |
| CELF CD | 0.0 | 0.0 | 24.6 | 52.5 | 0.0 | 0.0 | 21.3 | 57.4 |
| CELF WC | 0.0 | 0.0 | 31.1 | 45.9 | 0.0 | 0.0 | 22.9 | 39.3 |
| CELF FS | 0.0 | 0.0 | 26.2 | 50.8 | 0.0 | 0.0 | 22.9 | 42.6 |
| CELF RS | 0.0 | 0.0 | 24.6 | 45.9 | 0.0 | 0.0 | 23.3 | 50.0 |

Ceiling 1 = absolute maximum; Ceiling 2 = ≥(absolute maximum − 5% of absolute maximum); Floor 1 = 0; Floor 2 = ≤(0 + 5% of absolute maximum); SIB, Severe Impairment Battery; VABS (SN), Vineland Adaptive Behavior Scales Supplemental Norm; RBMT-C, Rivermead Behavioral Memory Test for Children; CELF, Clinical Evaluation of Language; CD, Concepts and Directions; WC, World Classes; FS, Formulated Sentences; RS, Recalling Sentences.

TABLE VI. Practice and Learning Effects*

| | Practice effect | | Learning effect | | | |
|---|---|---|---|---|---|---|
| | Donepezil, % total sample | Placebo, % total sample | Donepezil, % total sample | Placebo, % total sample | Donepezil, % practice effect sample | Placebo, % practice effect sample |
| SIB | 66.1 | 62.3 | 32.2 | 32.8 | 48.7 | 52.6 |
| RBMT-C | 45.8 | 57.6 | 19.6 | 20.0 | 40.7 | 35.3 |
| CELF CD | 32.2 | 39.0 | 12.7 | 5.2 | 36.8 | 13.0 |
| CELF WC | 39.3 | 48.3 | 9.8 | 8.6 | 22.7 | 17.8 |
| CELF FS | 45.0 | 51.7 | 10.5 | 16.7 | 22.2 | 32.2 |
| CELF RS | 39.7 | 50.8 | 13.5 | 16.9 | 30.4 | 33.3 |

SIB, Severe Impairment Battery; RBMT-C, Rivermead Behavioral Memory Test for Children; CELF, Clinical Evaluation of Language; CD, Concepts and Directions; WC, World Classes; FS, Formulated Sentences; RS, Recalling Sentences.
*Practice effect is defined as the improvement from screening to baseline; learning effect is defined as the improvement from screening to baseline and from baseline to week 12 (last observation carried forward).

Premature discontinuations due to AEs were higher in the DB-D group than the DB-P group (five vs. one; Fig. 1) and higher in the OL -PD group than in the OL-DD group (seven vs. three; Fig. 1).

There were no clinically significant changes in laboratory parameters, ECG findings, vital signs, or physical or neurological examination findings in the double-blind or open-label phases for either treatment group.

## DISCUSSION

To our knowledge, this is the first large, multicenter, randomized, double-blind, placebo-controlled trial of a pharmacological agent that assesses treatment efficacy with regard to elements of cognition and global function in young adults with DS but no signs of AD. While donepezil and other ChEIs have been tested in both adults [Hemingway-Eltomey and Lerner, 1999; Kishnani et al., 1999, 2001; Lott et al., 2002; Prasher et al., 2002, 2003, 2005; Heller et al., 2003; Prasher, 2004] and children with DS [Heller et al., 2004], none of these studies are comparable to the current study in either size or design. The challenge in designing this study was how to measure the possible treatment effects, given the phenotypic variability, such as the wide range of IQ scores and language abilities in the DS population and the lack of widely accepted measures of treatment efficacy in clinical use or reported in the literature for DS studies. The results of this study suggest that donepezil may have a beneficial effect on cognition and global function in young adults with DS. However, it is also apparent from this study that the choice of assessment instruments, their frequency of administration, and the methods of analyzing outcomes can limit the detection of efficacy and the interpretation of results.

TABLE VII. Responder Analyses Outcomes*

| | $\chi^2$ | *P*-value |
|---|---|---|
| SIB + x[a] | 2.78 | 0.095 |
| SIB + xV[b] | 2.38 | 0.123 |
| Non-SIB xV[c] | 4.00 | 0.045 |
| RBMT-C and/or CELF RS[d] | 6.98 | 0.008 |
| VABS >10% | 12.44 | 0.00042 |

SIB, Severe Impairment Battery; RBMT-C, Rivermead Behavioral Memory Test for Children; CELF, Clinical Evaluation of Language; CD, Concepts and Directions; WC, World Classes; FS, Formulated Sentences; RS, Recalling Sentences; VABS, Vineland Adaptive Behavior Scales.
*All analyses are 2 × 2 tables—Donepezil/Placebo, meets definition yes/no; all df = 1.
[a]Responder defined as week 12 (LOCF) improvement on SIB and ≥3 other patient-performance-based measures (RBMT-C, CELF CD, CELF WC, CELF FS, CELF RS).
[b]Responder defined as week 12 (LOCF) improvement on SIB and ≥3 other patient-performance-based measures (RBMT-C, CELF CD, CELF WC, CELF FS, CELF RS).
[c]Responder defined as week 12 (LOCF) improvement ≥3 patient-performance-based measures excluding the SIB (RBMT-C, CELF CD, CELF WC, CELF FS, CELF RS) or the VABS.
[d]Responder defined as week 12 (LOCF) improvement on either the RBMT-C or the CELF RS or both.

The SIB was chosen as the primary efficacy measure for this study because it has been validated in an adult DS population [Witts and Elders, 1998], and because it has shown treatment effects in a study of moderate-to-severe AD [Feldman et al., 2001]. It was not known, however, how it would perform in this population, and powering this study had to be based on AD study results. Here, individuals receiving donepezil treatment showed a significant improvement on the SIB in the double-blind phase (DB-D group), which stabilized during the open-label phase (OL-DD group). This would be an expected clinical result if the hypothesis of a cholinergic deficit/dysfunction inherent to DS were correct. The clinical expectation for the placebo (DB-P) group would be little or no change in the double-blind phase and improvement at the end of the open-label phase. However, individuals in the DB-P group also showed an improvement on the SIB in the double-blind phase. Interestingly, on entering the open-label phase (OL-PD) they showed further improvement comparable in magnitude to that seen for both groups in the double-blind phase. Donepezil may, therefore, afford a cognitive benefit to young adults with DS, but the improvement of the placebo group during the double-blind phase prevents this from being a clear conclusion. As discussed below, practice and ceiling effects likely contributed to these observed results. Because the SIB was the a priori declared primary efficacy measure, efforts were made to better understand these data, which led to a variety of post hoc analyses.

Ceiling effects were anticipated when designing this study (SIB scores >95 at screening were not allowed), but practice/learning

TABLE VIII. AEs Reported by >5% of Individuals in Either Treatment Group

| | DB-D (n = 62), n (%) | DB-P (n = 61), n (%) | OL-DD (n = 42), n (%) | OL-PD (n = 45), n (%) |
|---|---|---|---|---|
| Any AE | 46 (74.2) | 29 (47.5) | 24 (57.1) | 38 (84.4) |
| Abdominal pain | 6 (9.7) | 3 (4.9) | 1 (2.4) | 8 (17.8) |
| Asthenia | 4 (6.5) | 0 (0) | 0 (0) | 0 (0) |
| Headache | 4 (6.5) | 3 (4.9) | 1 (2.4) | 3 (6.7) |
| Sinus bradycardia | 0 (0) | 0 (0) | 1 (2.4) | 3 (6.7) |
| Anorexia | 4 (6.5) | 0 (0) | 0 (0) | 0 (0) |
| Diarrhea | 10 (16.1) | 6 (9.8) | 5 (11.9) | 5 (11.1) |
| Dyspepsia | 4 (6.5) | 1 (1.6) | 0 (0) | 0 (0) |
| Nausea | 7 (11.3) | 0 (0) | 2 (4.8) | 3 (6.7) |
| Vomiting | 8 (12.9) | 1 (1.6) | 2 (4.8) | 3 (6.7) |
| Agitation | 0 (0) | 0 (0) | 2 (4.8) | 3 (6.7) |
| Insomnia | 5 (8.1) | 0 (0) | 0 (0) | 3 (6.7) |
| Somnolence | 0 (0) | 0 (0) | 0 (0) | 3 (6.7) |
| Cough increased | 0 (0) | 0 (0) | 2 (4.8) | 3 (6.7) |
| Respiratory tract infection | 6 (9.7) | 5 (8.2) | 4 (9.5) | 4 (8.9) |
| Urinary incontinence | 0 (0) | 0 (0) | 0 (0) | 3 (6.7) |
| Vaginitis | 0 (0) | 0 (0) | 0 (0) | 1 (5.9) |

AE, adverse event; DB-D, double-blind donepezil group; DB-P, double-blind placebo group; OL-DD, open-label donepezil group who received donepezil in the double-blind phase; OL-PD, open-label donepezil group who received placebo in the double-blind phase.

effects were not, perhaps reflecting a bias toward low-performance expectations for this group. In hindsight, given that the SIB has only one form and that it was given approximately 2 weeks apart from screening to baseline, the occurrence of practice/learning effects and as a consequence, ceiling effects, is not surprising. One can also speculate that individuals with DS may be prone to practice effects as they become more comfortable with the research setting and process. The continued influence of practice/learning effects is evidenced by examining the contribution of RTM to the change in the mean value at week 12 compared with baseline. RTM is a phenomenon that occurs with all repeated measures due to a variety of reasons, including the inherent variability in individuals and in assessment instruments. The expected amount of difference between the means due to RTM can be calculated given certain known parameters from a data set. This value can then be compared to the difference between the actual observed means for specified subgroups. In this study, for both the donepezil and the placebo groups with baseline SIB scores <80 (21.7% of the study population) or >90 (40% of the study population), more improvement was seen in the actual data than RTM predicted. Practice/learning effects are one factor that could explain this; actual drug effect could be another factor in the donepezil group.

Other factors that might help to explain the SIB results were also explored. These included age, sex, age and sex combined, BMI (given the potential contribution of sleep apnea to cognitive performance in obese subjects), and various baseline SIB score cutoffs. None of these analyses were informative. It is now clear that the SIB was not the appropriate choice to measure a clinically relevant treatment effect in this study population.

The secondary efficacy measures in this study were chosen to supplement the more general cognitive measure of the SIB. The VABS has a long history of use with this population, and there was some limited preliminary evidence that it might show a treatment effect. Memory is an area affected by cholinergic deficit/dysfunction, and the RBMT-C had been validated in adults with DS. Language function, as measured by an earlier version of the CELF, is another area for which there was preliminary evidence of a potential benefit from donepezil. Of the three secondary efficacy measures, only the VABS appeared free from the performance variability limitations caused by floor and ceiling effects or practice/learning effects. The availability of supplementary norm scores standardized with a comparable population provided a full range of scores with a near-normal distribution. The result was that the DB-D group change was significant compared with baseline, and the difference compared with the DB-P group approached significance in the LOCF analysis and was significant for the OC and EE analyses. This result is perhaps the strongest evidence for a donepezil treatment benefit in this study and is consistent with the results of previous clinical studies as well as with the cholinergic hypothesis of DS.

The performance of the RBMT-C was surprising. It has four different forms that were used at the four visits in this study, and it did not have substantial numbers of subjects at the floor or ceiling. Despite this, nearly 46% of the DB-D group and 58% of the DB-P group improved from screening to baseline, showing a practice effect, and approximately 41% and 35% of these subgroups, respectively, showed continued improvement at Week 12, suggesting a learning effect.

The performance of the CELF-3 subtests is notable for the prominence of floor effects. Roughly one-quarter of both groups were at the absolute floor and about half were at the relative floor of the two receptive and the two expressive language subtests. Less

than 45% of the subjects in each group were able to complete Rapid Automatic Color-Shape Naming, which is the only part of the Rapid Automatic Naming subtest that is scored. In addition, from about one-third to one half of each group showed practice effects, depending on the subtest, although generally less than 20% showed learning effects. Thus, it is not surprising that none of these measures showed significant treatment differences. It is perhaps notable that while none of the DB-P group's changes from baseline were significant, recalling sentences (a memory as well as an expressive language assessment) did show significant improvement compared to baseline for the DB-D group.

The performance limitations of these assessment instruments prompted further post hoc analyses similar to those described for the SIB. As with the SIB, none of these were informative. Examination of the data revealed a wide range of individual responses, which suggested the possible utility of responder analyses. DS is well known for its phenotypic variability, despite its consistent genetic cause. Within this variation, each individual with DS has a unique set of strengths and weaknesses. To the extent that cholinergic function or deficit/dysfunction underlies various aspects of cognitive performance, it seems clinically reasonable that response to cholinergic enhancement will also be variable, so that where memory may improve for one individual, language function may improve for another. Thus, rather than defining responders by the amount of improvement on any given assessment, we instead chose to define them based on the number of tests showing any improvement. The first two responder definitions required improvement on the SIB, because it was the primary outcome measure, as well as at least three other patient-performance-based measures. Using these definitions favored the DB-D group, but did not reach statistical significance ($P = 0.095$ and 0.123, respectively), probably because they included placebo responders by definition. The non-SIB + xV definition avoided this issue by excluding the SIB and showed significantly more responders in the DB-D group ($P = 0.045$). The next definition focused on improvement on one or more of the two memory items, the RBMT-C and the CELF-3 Recalling Sentences, which both suggested mean improvement from baseline for the DB-D group but not the DB-P group. Significantly more donepezil subjects showed response by this definition ($P = 0.008$). The final responder analysis was done to further explore the results on the VABS. Because the mean improvement for the DB-D group was only about 3 percentile points, it could reasonably be asked if this is a clinically meaningful result. We believe that greater than 10 percentile points improvement on the adaptive behavior composite supplementary norm scale would represent a clinically meaningful outcome. Using this responder definition, nearly 40% of the DB-D group responded compared with 10% of the DB-P group; this difference was highly significant ($\chi^2 = 12.44$; df = 1; $P = 0.00042$).

Donepezil was generally well tolerated in this population of young adults with DS. Safety and tolerability outcomes are important in a vulnerable population such as individuals with DS. No deaths or other SAEs were reported in either treatment group throughout the course of the double-blind or open-label phases of this study. The rates of individual AEs in this study were similar to those observed in the pivotal trials for patients with AD (~74%) [Rogers et al., 1998a,b; Burns et al., 1999; Winblad et al., 2001; Aricept, 2006], indicating that the safety profile of donepezil in DS is comparable. While AE rates and premature discontinuations due to AEs were higher in the DB-D group than in those in the DB-P group, AE rates in the OL-PD group were similar to those of the DB-D group. Discontinuations due to AEs were higher in the OL-PD group than in the DB-D group. The most common AEs associated with donepezil treatment were related to the digestive system and are likely to be related to the cholinergic effects of donepezil. There were also no notable safety findings in laboratory values, ECGs, vital signs, and physical examinations. It must be noted, however, that the longest collection of safety data was 6 months, so that no conclusions regarding the safety of long-term use in this population can be made. In addition, as this is a developmentally delayed population, the reported AE rate may underestimate the actual frequency of AEs.

A potential limitation of this study was the lengthy 3-year enrollment period, which may have introduced additional variability into the results due to the length of time between completing subjects and their assessments at individual sites. Anecdotal reports suggested that a main reason for the slow enrollment was that, given their age, many of the subjects had reached a period of relative stability in their lives and did not want to risk disrupting this stability. Furthermore, due to enrollment issues, the study was underpowered, which may also have limited our potential to show treatment-mediated effects. We also acknowledge that post hoc analyses must necessarily be viewed with caution. They were useful in this study for illuminating the limitations of the assessment instruments and the standard clinical trial efficacy analyses that were chosen a priori.

The intellectual disability and limitations of adaptive behavior experienced by people with DS represent a significant unmet medical need that warrants further study. The hypothesis that a cholinergic deficit or dysfunction due to the extra copy of chromosome 21 contributes to the disability of DS is scientifically defendable. That the phenotypic variability of this syndrome extends to the underlying cholinergic deficit/dysfunction and leads to a variable response to donepezil is clinically reasonable. The response pattern of both study groups on the SIB, the results of the VABS, and the results of the responder analyses all suggest that donepezil may benefit some young adults with DS, but no signs of AD, and seem consistent with a cholinergic deficit/dysfunction hypothesis. However, as these results cannot be extended to children with DS, studies of ChEIs in this population are clearly needed, especially given the supposition that earlier treatment may provide greater benefits. Furthermore, additional studies are needed to further explore both the results of this trial and the hypothesis of inherent cholinergic deficit/dysfunction in DS. The lessons learned from this study regarding selection of assessment instruments, frequency of administration, and choice of analysis methodology may help address some of the challenges identified in conducting clinical trials in DS populations [Heller et al., 2006b] and should also help to design better future trials for both adults and children with DS.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the essential contribution of Yikang Xu, PhD (Pfizer, Inc.) in performing the statistical analyses

reported within the article. The authors would also like to acknowledge the editorial assistance provided by M. Seymour, PhD, and R. Daniel, PhD of PAREXEL, which was funded by Eisai, Inc. and Pfizer, Inc.

*Financial support and disclosures*: This study was sponsored by Eisai, Inc. and Pfizer, Inc. Study data were initially analyzed by Drs. McRae and Richardson, with subsequent analytical contribution from Drs. Kishnani, Spiridigliozzi, Capone, and Handen and Mr. Heller. Statistical analyses were performed by Dr. Yikang Xu of Pfizer, Inc. The manuscript was written by the authors with editorial assistance from M. Seymour, PhD, and R. Daniel, PhD of PAREXEL, which was funded by Eisai, Inc. and Pfizer, Inc.

Drs. Kishnani and Spiridigliozzi and Mr. Heller have received research grants and honoraria and/or have served as consultants for various companies including: Pfizer, Johnson & Johnson, Eisai, and Novartis. In addition, Drs. Kishnani and Spiridigiozzi and Mr. Heller will receive royalties from Pfizer for the use of a language measure (TOVER), developed at Duke University Medical Center, in several clinical trials and they have applied for a patent for the use of ChEIs in a population other than individuals with Down syndrome. Dr. Sommer has received honoraria from Pfizer, Inc. and Forest Pharmaceuticals, has received funding from Novartis, and is a co-investigator on a GlaxoSmithKline clinical study. Dr. Handen is a consultant to Eisai Medical Research, Forest Research Institute, and Bristol-Myers Squibb. He has received research funding from Eli Lilly, Pfizer, Bristol-Myers Squibb, Pediamed, Janssen, Cerebrus, and Neuropharm. Dr. Seltzer has received research support and honoraria for speaking from Eisai/Pfizer, Novartis, Forest, Myriad, Janssen, and Bristol-Myers-Squibb. Dr. Richardson is an employee of Eisai, Inc. and Dr. McRae is an employee of Pfizer, Inc. Dr. Capone has no disclosures.

## APPENDIX

The SIB was chosen as the primary outcome measure and has validation in a Down syndrome (DS) population aged 22–53 years, with both test–retest and criterion validity [Witts and Elders, 1998]. The SIB is a 40-item assessment (some items have multiple parts) designed to evaluate cognitive function in severely demented patients [Panisset et al., 1994], and evaluates domains of memory, language, orientation, attention, praxis, visuospatial awareness, construction skills, orientation to name, and social interaction. Scores range from 0 (greatest impairment) to 100 (no impairment).

Secondary efficacy measures in the study included the VABS [Sparrow et al., 2006], the RBMT-C [Wilson et al., 1989, 1991], and five subtests of the CELF-3 [Semel et al., 1995].

The VABS measures personal and social skills such as communication, daily living skills, and socialization, and provides a composite score reflecting an individual's overall function [Sparrow et al., 2006]. It is presented to the caregiver in a semi-structured interview format. Lower VABS scores indicate greater functional impairment. The VABS was selected as an efficacy measure because it is a validated, caregiver-rated, global measure often used in the DS population for general assessment. It was intended to provide context for improvement in individual performance-based measures. The VABS is designed to assess individuals based on comparison to a national sample of normal, non-mentally handicapped individuals. In this process, raw scores are converted to standardized scores with a standardized score of 100 equaling the 50th percentile. In this study population, the mean standardized scores fell below the 0.1 percentile, creating a floor effect, which limited valid data comparisons. To address this, raw scores were instead transformed to a supplementary norm score based on a comparable population of mentally handicapped adults aged 18 years or older living in non-residential facilities. This transformation yielded a range of scores that was adequate for the comparison of treatment differences.

The RBMT-C is a test battery for detecting and monitoring everyday memory problems, and tests activities such as recognizing a face or recalling a message [Wilson et al., 1989, 1991]. The test was originally developed for use in children, but has been validated for use in adults with DS [Hon et al., 1998; Wilson and Ivani-Chalian, 1995]. The validation of the RBMT-C in a DS population, and the potential for donepezil treatment to provide improvements in memory, suggested that this scale would be a useful outcome measure in this study.

The standard battery of the CELF-3 consists of six subtests, standardized for children up to 18 years of age. Four of the six subtests were administered in this protocol. Raw score results of these tests were reported. Two of the subtests are measures of receptive language function: Concepts and Directions (raw score range, 0–30) and Word Classes (raw score range, 0–34), and two are measures of expressive language function: Formulated Sentences (raw score range, 0–44) and Recalling Sentences (raw score range, 0–78). A supplemental subtest, Rapid Automatic Naming, which measures the processes involved in automatic naming and resistance to interference from multiple dimensions of stimuli, was also administered.

Responder analyses performed in the current analysis utilized the following definitions: (1) week 12 LOCF improvement on the SIB + improvement on $\geq$3 additional patient-performance-based measures consisting of the RBMT-C and four subtests of the CELF -3, but excluding the VABS (SIB + x); (2) week 12 LOCF improvement on the SIB + improvement on $\geq$3 additional patient-performance-based measures or the VABS (SIB + xV); (3) week 12 LOCF improvement on $\geq$3 patient-performance-based measures or the VABS, but excluding the SIB (non-SIB + xV); (4) week 12 LOCF improvement on either the RBMT-C or the CELF-3 recalling sentences or both; (5) week 12 LOCF improvement of greater than 10 percentile points on the VABS composite supplementary norm scale.

## REFERENCES

Allen DD, Martin J, Arriagada C, Cardenas AM, Rapoport SI, Caviedes R, Caviedes P. 2000. Impaired cholinergic function in cell lines derived from the cerebral cortex of normal and trisomy 16 mice. Eur J Neurosci 12:3259–3264.

Aricept [package insert]. 2006. Eisai, Inc., Teaneck, NJ. Available at: www.eisai.com or www.aricept.com.

Beccaria L, Marziani E, Manzoni P, Arvat E, Valetto MR, Gianotti L, Ghigo E, Chiumello G. 1998. Further evidence of cholinergic impairment of the

neuroendocrine control of the GH secretion in Down's syndrome. Dement Geriatr Cogn Disord 9:78–81.

Black SE, Doody R, Li H, McRae T, Jambor KM, Xu Y, Sun Y, Perdomo CA, Richardson S. 2007. Donepezil preserves cognition and global function in patients with severe Alzheimer disease. Neurology 69:459–469.

Burns A, Rossor M, Hecker J, Gauthier S, Petit H, Moller H-J, Rogers SL, Friedhoff LT, the international Donepezil Study Group. 1999. The effects of donepezil in Alzheimer's disease—Results from a multinational trial. Dement Geriatr Cogn Disord 10:237–244.

Corsi P, Coyle JT. 1991. Nerve growth factor corrects developmental impairments of basal forebrain cholinergic neurons in the trisomy 16 mouse. Proc Natl Acad Sci USA 88:1793–1797.

Feldman H, Gauthier S, Hecker J, Vellas B, Subbiah P, Whalen E. 2001. A 24 -week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease. Neurology 57:613–620.

Fiedler JL, Epstein CJ, Rapoport SI, Caviedes R, Caviedes P. 1994. Regional alteration of cholinergic function in central neurons of trisomy 16 mouse fetuses, an animal model of human trisomy 21 (Down syndrome). Brain Res 658:27–32.

Florez J, del Arco C, Gonzalez A, Pascual J, Pazos A. 1990. Autoradiographic studies of neurotransmitter receptors in the brain of newborn infants with Down syndrome. Am J Med Genet Suppl 7:301–305.

Guerra M, Llorens N, Fernhall B. 2003. Chronotropic incompetence in persons with Down syndrome. Arch Phys Med Rehabil 84:1604–1608.

Hardy JA, Higgins GA. 1992. Alzheimer's disease: The amyloid cascade hypothesis. Science 256:184–185.

Harris WS, Goodman RM. 1968. Hyper-reactivity to atropine in Down's syndrome. N Engl J Med 279:407–410.

Heller JH, Spiridigliozzi GA, Sullivan JA, Doraiswamy PM, Krishnan RR, Kishnani PS. 2003. Donepezil for the treatment of language deficits in adults with Down syndrome: A preliminary 24-week open trial. Am J Med Genet Part A 116A:111–116.

Heller JH, Spiridigliozzi GA, Doraiswamy PM, Sullivan JA, Crissman BG, Kishnani PS. 2004. Donepezil effects on language in children with Down syndrome: Results of the first 22-week pilot clinical trial. Am J Med Genet Part A 130A:325–326.

Heller JH, Spiridigliozzi GA, Crissman BG, Sullivan JA, Eells RL, Li JS, Doraiswamy PM, Krishnan KR, Kishnani PS. 2006a. Safety and efficacy of rivastigmine in adolescents with Down syndrome: A preliminary 20-week, open-label study. J Child Adolesc Psychopharmacol 16:755–765.

Heller JH, Spiridigliozzi GA, Crissman BG, Sullivan-Saarela JA, Li JS, Kishnani PS. 2006b. Clinical trials in children with Down syndrome: Issues from a cognitive research perspective. Am J Med Genet Part C 142C:187–195.

Hemingway-Eltomey JM, Lerner AJ. 1999. Adverse effects of donepezil in treating Alzheimer's disease associated with Down's syndrome. Am J Psychiatry 156:1470.

Hon J, Huppert FA, Holland AJ, Watson P. 1998. The value of the Rivermead Behavioural Memory Test (Children's Version) in an epidemiological study of older adults with Down syndrome. Br J Clin Psychol 37:15–29.

Johnson N, Fahey C, Chicoine B, Chong G, Gitelman D. 2003. Effects of donepezil on cognitive functioning in Down syndrome. Am J Ment Retard 108:367–372.

Kishnani PS, Sullivan JA, Walter BK, Spiridigliozzi GA, Doraiswamy PM, Krishnan KR. 1999. Cholinergic therapy for Down's syndrome. Lancet 353:1064–1065.

Kishnani PS, Spiridigliozzi GA, Heller JH, Sullivan JA, Doraiswamy PM, Krishnan KR. 2001. Donepezil for Down's syndrome. Am J Psychiatry 158:143.

Lai F, Williams RS. 1989. A prospective study of Alzheimer disease in Down syndrome. Arch Neurol 46:849–853.

Lott IT, Osann K, Doran E, Nelson L. 2002. Down syndrome and Alzheimer disease: Response to donepezil. Arch Neurol 59:1133–1136.

Panisset M, Roudier M, Saxton J, Boller F. 1994. Severe impairment battery. A neuropsychological test for severely demented patients. Arch Neurol 51:41–45.

Prasher VP. 2004. Review of donepezil, rivastigmine, galantamine and memantine for the treatment of dementia in Alzheimer's disease in adults with Down syndrome: Implications for the intellectual disability population. Int J Geriatr Psychiatry 19:509–515.

Prasher VP, Huxley A, Haque MS. 2002. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Down syndrome and Alzheimer's disease—Pilot study. Int J Geriatr Psychiatry 17:270–278.

Prasher VP, Adams C, Holder R. 2003. Long term safety and efficacy of donepezil in the treatment of dementia in Alzheimer's disease in adults with Down syndrome: Open label study. Int J Geriatr Psychiatry 18:549–551.

Prasher VP, Fung N, Adams C. 2005. Rivastigmine in the treatment of dementia in Alzheimer's disease in adults with Down syndrome. Int J Geriatr Psychiatry 20:496–497.

Rogers SL, Doody RS, Mohs RC, Friedhoff LT, the Donepezil Study Group. 1998a. Donepezil improves cognition and global function in Alzheimer disease: A 15-week, double-blind, placebo-controlled study. Arch Intern Med 158:1021–1031.

Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT, the Donepezil Study Group. 1998b. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology 50:136–145.

Rosler M, Anand R, Cicin-Sain A, Gauthier S, Agid Y, Dal Bianco P, Stähelin HB, Hartman R, Gharabawi M, on behalf of the B303 Exelon Study Group. 1999. Efficacy and safety of rivastigmine in patients with Alzheimer's disease: International randomised controlled trial [published erratum appears in: BMJ 2001; 322: 1456]. Br Med J 318:633–638.

Sacks B, Smith S. 1989. People with Down's syndrome can be distinguished on the basis of cholinergic dysfunction. J Neurol Neurosurg Psychiatry 52:1294–1295.

Salehi A, Delcroix JD, Belichenko PV, Zhan K, Wu C, Valletta JS, Takimoto-Kimura R, Kleschevnikov AM, Sambamurti K, Chung PP, Xia W, Villar A, Campbell WA, Kulnane LS, Nixon RA, Lamb BT, Epstein CJ, Stokin GB, Goldstein LS, Mobley WC. 2006. Increased App expression in a mouse model of Down's syndrome disrupts NGF transport and causes cholinergic neuron degeneration. Neuron 51:29–42.

Semel E, Wilig E, Second W. 1995. Clinical evaluation of language fundamental. 3rd edition. San Antonio, TX: The Psychological Corporation.

Sparrow SS, Balla DA, Cicchetti DV. 1984. Vineland Adaptive Behavior Scales—Interview Edition (Survey Form). Livonia, MN: American Guidance Service, Inc. Available at: http://www.pearsonassessments.com/assessments/technical/vineland.asp (accessed March 12, 2009).

Sparrow SS, Cicchetti DV, Balla DA. 2006. Vineland Adaptive Behavior Scales: Second Edition. Livonia, MN: American Guidance Services, Inc. Available at: http://www.pearsonassessments.com/vinelandadapt.aspx (accessed March 12, 2009).

Spiridigliozzi GA, Heller JH, Crissman BG, Sullivan-Saarela JA, Eells R, Dawson D, Li J, Kishnani PS. 2007. Preliminary study of the safety and efficacy of donepezil hydrochloride in children with Down syndrome: A clinical report series. Am J Med Genet Part A 143A:1408–1413.

Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S, Ding C, the Galantamine USA-10 Study Group. 2000. A 5-month, randomized, placebo-controlled trial of galantamine in AD. Neurology 54:2269–2276.

Tasdemir HA, Cetinkaya MC, Polat C, Belet U, Kalayci AG, Akbas S. 2004. Gallbladder motility in children with Down syndrome. J Pediatr Gastroenterol Nutr 39:187–191.

Wilkinson DG, Francis PT, Schwam E, Payne-Parrish J. 2004. Cholinesterase inhibitors used in the treatment of Alzheimer's disease: The relationship between pharmacological effects and clinical efficacy. Drugs Aging 21:453–478.

Wilson BA, Ivani-Chalian R. 1995. Performance of adults with Down's syndrome on the children's version of the Rivermead Behavioural Memory Test: A brief report. Br J Clin Psychol 34:85–88.

Wilson B, Cockburn J, Baddeley A, Hiorns R. 1989. The development and validation of a test battery for detecting and monitoring everyday memory problems. J Clin Exp Neuropsychol 11:855–870.

Wilson B, Cockburn J, Baddley A. 1991. The Rivermead Behavioural Memory Test. Bury St. Edmunds, England: Thames Valley Test Company.

Winblad B, Engedal K, Soininen H, Verhey F, Waldemar G, Wimo A, Wetterholm AL, Zhang R, Haglund A, Subbiah P, the Donepezil Nordic Study Group. 2001. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. Neurology 57:489–495.

Winblad B, Kilander L, Eriksson S, Minthon L, Batsman S, Wetterholm AL, Jansson-Blixt C, Haglund A. 2006. Donepezil in patients with severe Alzheimer's disease: Double-blind, parallel-group, placebo-controlled study. Lancet 367:1057–1065.

Witts P, Elders S. 1998. The 'Severe Impairment Battery': Assessing cognitive ability in adults with Down syndrome. Br J Clin Psychol 37:213–216.

Yudkin PL, Stratton IM. 1996. How to deal with regression to the mean in intervention studies. Lancet 347:241–243.