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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 21-436
Supplement #:	S-036
Drug Name:	Abilify (Aripiprazole) 2 – 15 mg daily
Indication(s):	Irritability Associated with Autistic Disorder
Applicant:	Bristol-Myers Squibb Company
Date of submission:	Prior Approval Supplement submitted on 03/15/2013; Amendment after Approval submitted on 05/24/2013.
Review Priority:	Standard
Biometrics Division:	Division of Biometrics I
Statistical Reviewer:	Andrejus Parfionovas, Ph.D.
Concurring Reviewers:	Peiling Yang, Ph.D., HM James Hung, Ph.D.
Medical Division:	Division of Psychiatry Products
Clinical Team:	Cara Alfaro, M.D., Reviewer
	Silvana Borges, M.D., Acting Team Leader
Project Manager:	Simran Parihar, Pharm. D.

Keywords: multi-center, clinical studies, NDA review, survival analysis, logrank test

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The sponsor's findings on aripiprazole (2—15mg/day) were confirmed by the reviewer to be statistically non-significantly different from placebo (log-rank test p-value 0.0965) in reducing the symptoms of irritability associated with autistic disorder during the 16 weeks of double-blind treatment of pediatric subjects who maintained a response for 12 weeks of treatment.

The effect on the non-white population may not be conclusive because of the small sample size and the possible confounding effects that were not accounted for in the study.

Although the study failed, it appears that the sponsor conducted it in accordance with the statistical analysis plan agreed upon by the Agency.

2 INTRODUCTION

2.1 Overview

This review provides statistical evaluation of the study CN138603, which was designed to assess whether pediatric subjects who maintain a response for 12 weeks of Abilify (aripiprazole) treatment for their symptoms of irritability associated with autistic disorder will experience a relapse significantly later when continuing therapy with aripiprazole than subjects treated with placebo.

Aripiprazole has been approved in the United States (US) in 2009 for the treatment of pediatric patients (aged 6-17 years) with irritability associated with autistic disorder. The current study was conducted as a post-marketing requirement as stated in our Agency letter dated 11/19/2009. The study provides long-term, placebo-controlled data in this patient population (see Table 1).

Table 1. List of the studies included in the analysis.

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
CN138603	Phase 4	16 weeks	None	Abilify: 41, Placebo: 44.	Pediatric patients with Irritability associated with Autistic Disorder

2.2 Data Sources

The sponsor submitted the clinical study report on the 03/15/2013 under the serial #037, available in \\Cdsesub1\evsprod\NDA021436\0037.

The sponsor provided the raw and derived datasets using SAS XPORT Transport Format on 08/30/2013. The data files are available in the following directory of the Electronic Document Room (EDR): \\CDSESUB1\evsprod\NDA021436\0046\m5\datasets\cn138603\

The listings of the SAS program codes for the derived variables and the statistical analysis were provided on 09/13/2013. The files are available in the following directory of the Electronic Document Room (EDR): \\CDSESUB1\evsprod\NDA021436\0047\m5\datasets\cn138603\

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

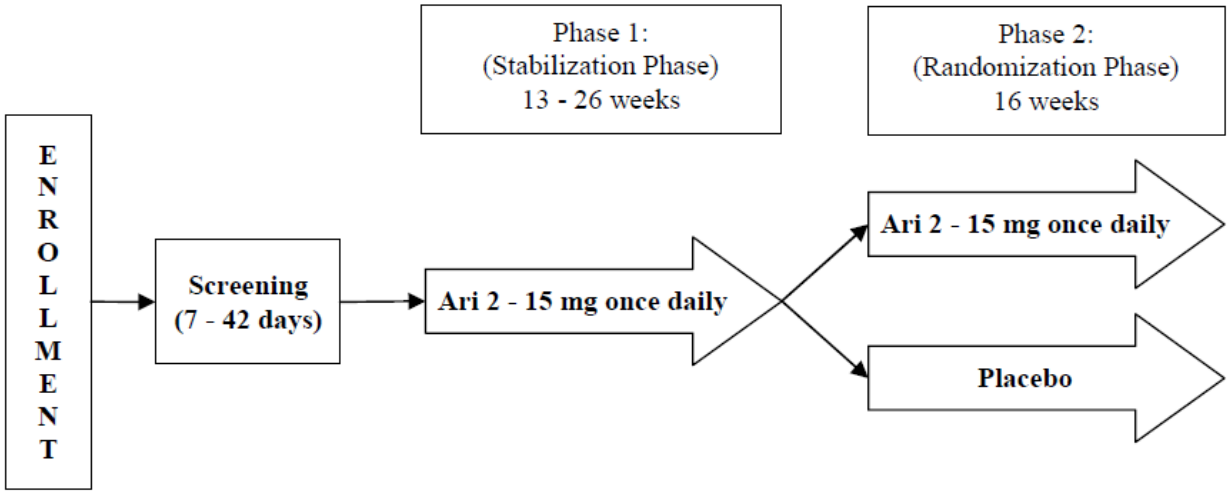
The reviewer finds the quality and integrity of the submitted data satisfying and acceptable for the reviewer's analysis. The reviewer was able to reproduce the primary analysis data from the submitted raw dataset, and to trace the derivation of the primary endpoint.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

This was a multicenter, US-only, double-blind, randomized, flexible-dose, placebo-controlled study with 2 parallel treatment groups designed to assess the safety and efficacy of aripiprazole in the long-term maintenance treatment of pediatric subjects with irritability associated with autistic disorder. The study consisted of 13–26 weeks *stabilization phase* (single-blind aripiprazole treatment for all the patients) and 16 weeks *randomization phase* (double-blind treatment with aripiprazole or placebo for the randomized patients). The graphical representation of the study design is shown on Figure 1.

Figure 1. Graphical design scheme of the study CN138603.



Source: Final Clinical study Report (pg. 24).

During the stabilization phase the optimal aripiprazole dose established for each patient, titrating from 2 mg/day but not exceeding 15 mg/day. During the randomization phase the subjects were starting at the dose received at the end of the randomization phase (2, 5, 10, or 15 mg/day), but the investigators were allowed to adjust the dose (within the range 2 –l5 mg/day) at their discretion based on clinical effects.

The *primary objective* of this study was to evaluate the efficacy of aripiprazole compared with placebo to prevent relapses in pediatric subjects who maintained a response for 12 weeks of

aripiprazole treatment for their symptoms of irritability associated with autistic disorder. The *primary efficacy endpoint* was defined as the time from randomization to relapse during the 16 weeks double-blind period. A patient was considered to have a *relapse* if one of the following occurred:

- ABC-I score increased ≥ 25% AND the CGI-I rating was either "much worse" or "very much worse" relative to the end of Phase 1 for two consecutive visits (one week apart).
- The patient drops out for reason of "lost-to follow-up" after a visit at which the ABC-I score increased ≥ 25% AND the CGI rating was either "much worse" or "very much worse" relative to the end of Phase 1.
- The patient starts a prohibited drug, regardless of prescriber, to treat worsening symptoms after a visit at which the ABC-I score increased ≥ 25% AND the CGI rating was either "much worse" or "very much worse" relative to the end of Phase 1.
- The patient discontinues the study due to hospitalization for worsening symptoms of irritability OR due to the lack of efficacy in the judgment of the investigator.

No key secondary endpoints were specified.

3.2.2 Statistical Methodologies

The primary efficacy outcome measure (time from randomization to relapse) was depicted by a Kaplan-Meier (KM) survival curves over the *randomized sample*. The primary analysis for the comparison was based on *log-rank test*, stratified by baseline body weight (dichotomized into 2 categories: ≥ 40 kg, and < 40 kg). The estimated hazard ratio and 95% confidence interval (CI) was obtained from the Cox regression model, with dichotomized baseline body weight as a stratification factor and with treatment group (aripiprazole or placebo) as a covariate.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

The clinical protocol specifies that the primary efficacy outcome measure will be evaluated by survival analysis using the *Randomized Sample*. In addition to the randomized sample the sponsor also defines the following samples:

- Phase 2 *Safety Sample* – randomized subjects who received at least one dose of double-blind medication.
- Phase 2 *Efficacy Sample* – all as above + had at least 1 efficacy evaluation after the randomization.

Subjects who were randomized, but never treated and did not experience an event, were censored on their randomization date (one in placebo arm, and two in aripiprazole). Subjects, whose relapse occur more than 3 days after the last dosing date of double-blind medication were also treated as censored. Patients who do not relapse (including those patients who discontinue early for reasons other than relapse) will be censored on their date of last efficacy evaluation or their last dose of double-blind study medication during Phase 3, whichever is later.

There were 3 patients (IDs: 0029-00115, 0003-00143, and 0039-00090) who did not receive medication during the double-blind (DB) phase. Of them one (ID: 0003-00143) was listed in the

raw dataset as having no relapse (censored), the other two (ID: 0029-00115 and 0039-00090) had no data and were missing from the raw datasets when all these three should have been treated as censored (see SAP section 7.5.2). There was also one patient (ID: 0014-00151) who experienced a relapse more than 3 days after the end of DB phase, and thus should be treated as censored (see SAP section 8.3). For the statistical analysis to follow the SAP definition of the *Randomized Sample*, the derived censor/relapse variable (RELAPSE ≤ 3) must be used (see Table 2).

Table 2. Summary of patients whose definition of censorship required clarification.

Patient ID	First day of DB medication	Last day of DB medication	Date of Relapse	Date of Censoring	Relapse	Relapse criteria	Relapse ≤ 3 days after the last DB dose	Treatment arm
0014-00151	12/17/2012	01/06/2012	01/11/2012	01/06/2012	1	Lack of efficacy	0	Placebo
0029-00115	.	.	.	12/15/2011	.	.	0	Placebo
0003-00143	.	.	.	01/30/2012	0	.	0	Abilify
0039-00090	.	.	.	02/02/2012	.	.	0	Abilify

Source: computed by the reviewer.

The patients' disposition prior to randomization is summarized in Table 3.

Table 3. Patients' disposition during the single-blind stabilization period.

	N (%)
Total number screened/enrolled	215 (100.0)
Passed screening/entered stabilization phase	157 (73.0)
Discontinued during stabilization	72 (45.9)
Adverse event	12 (7.6)
Subject withdrew consent	7 (4.5)
Lost to follow-up	8 (5.1)
Administrative reason by sponsor	11 (7.0)
No longer meet study criteria	7 (4.5)
Lack of efficacy	25 (15.9)
Poor/non-compliance	2 (1.3)
Completed stabilization phase	85 (54.1)

Source: Final clinical study report CN138603, Table 4.2.1, pg. 31.

The summary of the patients' disposition between the treatment arms in the randomized sample is presented in Table 4.

Table 4. Patients' disposition during the double-blind period (randomized sample).

	Placebo N (%)	Abilify N (%)	Total N (%)
Randomized	44 (100.0)	41 (100.0)	85 (100.0)
Discontinued during the double-blind phase	25 (56.8)	19 (46.3)	44 (51.8)
Adverse event	1 (2.3)	0	1 (1.2)
Subject withdrew consent	0	5 (12.2)	5 (5.9)
Lost to follow-up	0	1 (2.4)	1 (1.2)
Lack of efficacy	23 (52.3)	13 (31.7)	36 (42.4)
Poor/non-compliance	1 (2.3)	0	1 (1.2)
Completed double blind phase	19 (43.2)	22 (53.7)	41 (48.2)

Source: Final clinical study report CN138603, Table 4.2.1, pg. 31.

Demographic characteristics of the patients in the randomized sample are presented in Table 5.

Table 5. Patients' demographic characteristics (randomized sample).

	Placebo N = 44	Abilify N = 41	Total N = 85
Gender, N (%)			
Male	38 (86.4)	30 (73.2)	68 (80.0)
Female	6 (13.6)	11 (26.8)	17 (20.0)
Age (years)			
Mean (SD)	10.8 (2.77)	10.1 (2.80)	10.4 (2.79)
Min – Max	6 – 17	6 – 16	6 – 17
Weight (kg)			
Mean (SD)	50.6 (21.91)	51.7 (24.38)	51.1 (23.00)
Min – Max	19 – 110	21 – 117	19 – 117
Weight group, N (%)			
< 40 kg	15 (34.1)	17 (41.5)	32 (37.6)
≥ 40 kg	29 (65.9)	24 (58.5)	53 (62.4)
Height (cm)			
Mean (SD)	148.6 (18.24)	143.6 (14.24)	146.2 (16.53)
Min – Max	115 – 186	112 – 172	112 – 186
BMI (kg/m^2)			
Mean	21.9 (5.19)	24.0 (7.37)	22.9 (6.38)
Min – Max	14 – 38	15 – 43	14 – 43
Race, N (%)			
White/Caucasian	28 (63.6)	31 (75.6)	59 (69.4)
Black/African American	11 (25.0)	8 (19.5)	19 (22.4)
Asian	3 (6.8)	0	3 (3.5)
American Indian/Alaska Native	1 (2.3)	0	1 (1.2)
Other	1 (2.3)	2 (4.9)	3 (3.5)
Ethnicity, N (%)			
Hispanic or Latino	9 (20.5)	10 (24.4)	19 (22.4)
Not Hispanic or Latino	34 (77.3)	29 (70.7)	63 (74.1)

Source: Final clinical study report CN138603, Table 4.3.1, pp. 33-34.

Baseline disease characteristics, as measured by Aberrant Behavior Checklist (ABC) and Clinical Global Impression (CGI) of the randomized patients are summarized in Table 6.

Table 6. Patients disease baseline characteristics (randomized sample) .

	Placebo N = 44	Abilify N = 41	Total N = 85
ABC – Irritability			
Mean (SD)	8.2 (6.20)	9.5 (5.75)	8.8 (5.98)
Min – Max	0 – 22	0 – 22	0 – 22
ABC – Hyperactivity			
Mean (SD)	10.1 (9.36)	10.9 (7.16)	10.5 (8.33)
Min – Max	0 – 36	0 – 25	0 – 36
ABC – Stereotypy			
Mean (SD)	3.9 (3.74)	4.3 (3.42)	4.1 (3.57)
Min – Max	0 – 13	0 – 11	0 – 13
ABC – Social Withdrawal			
Mean (SD)	5.8 (6.60)	7.5 (6.03)	6.6 (6.35)
Min – Max	0 – 25	0 – 20	0 – 25

ABC – Inappropriate Speech Mean (SD) Min – Max	 2.1 (2.48) 0 – 9	 2.7 (2.96) 0 – 12	 2.4 (2.72) 0 – 12
CGI – Severity Mean (SD) Min - Max	 2.9 (1.13) 1 – 6	 3.0 (0.89) 1 – 5	 3.0 (1.02) 1 – 6

Source: Final clinical study report CN138603, Table 4.3.2, pp. 36-37.

3.2.4 Sponsor's Efficacy Results and Findings

The results of the primary efficacy analysis performed by the sponsor are summarized in Table 7.

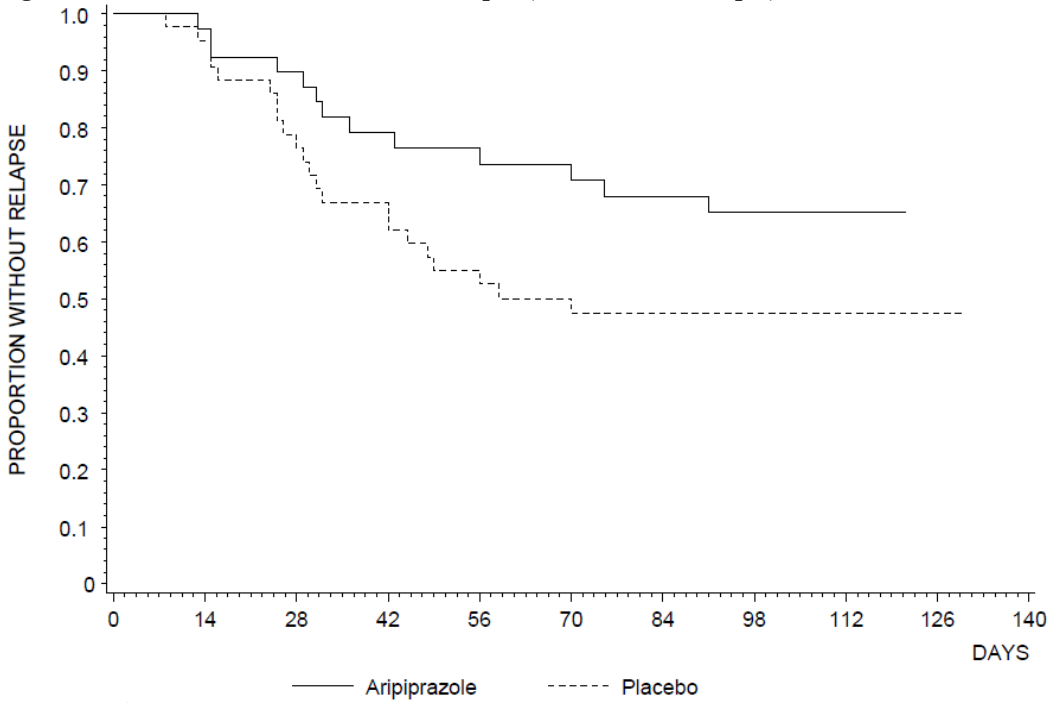
Table 7. Summary of primary efficacy analysis (Randomized Sample).

N of events/N of patients (%)		Treatment comparison using Cox-proportional hazard model: Hazard ratio (95% CI)	p-value from stratified log-rank test
Placebo	Abilify		
22/44 (50.0)	13/41 (31.7)	0.57 (0.28, 1.12)	0.097

Source: Final clinical study report CN138603, Table S.5.1, pg. 156.

The time from randomization to relapse analyzed using stratified log-rank test was found statistically not significant for the comparison of aripiprazole to placebo (p-value = 0.097). The Kaplan-Meier relapse rates at Week 16 were 32% and 50% for aripiprazole and placebo respectively (see Figure 2).

Figure 2. Time from randomization to relapse (Randomized Sample).



Source: Final clinical study report CN138603, Figure6.1, pg. 42.

The sponsor performed sensitivity analysis for the primary efficacy endpoint using permutation based exact log-rank test. The results of which were consistent with the primary efficacy analysis, i.e., the difference between placebo and treatment were found statistically not significant (p-value = 0.0871).

3.2.5 Statistical Reviewers' Findings and Comments

The statistical reviewer confirmed the sponsor's analysis results for the primary efficacy endpoint (time to relapse). The results were not found statistically significant (stratified log-rank test p-value 0.0965), showing no statistically significant difference between the Abilify and placebo treatments in time to relapse (see Table 8).

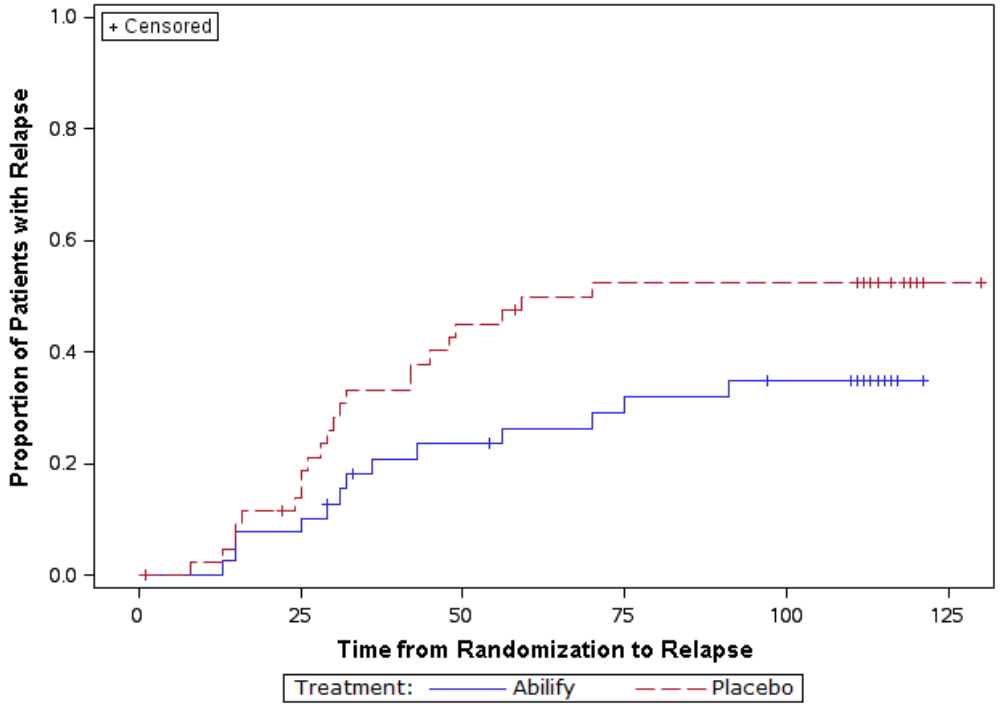
Table 8. Analysis of Maximum Likelihood Estimates and Log-Rank Test.

Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio (Abilify to Placebo)	95% Hazard Ratio Confidence Limits		Log-rank p-value
-0.57009	0.35058	2.6443	0.1039	0.565	0.284	1.124	0.0965

Source: computed by the reviewer.

The plot of the cumulative proportion of patients with relapse over time (derived from Kaplan-Meier estimates) is provided in Figure 3. The curves on the plot display the proportions of patients in each treatment arm who had a relapse by a given day after randomization. Although in these two plots the curves appear to separate from each other after a certain period of time, the findings are not statistically significant.

Figure 3. Kaplan-Meier Estimates of Proportion of Patients with Relapse over Time.



Source: computed by the reviewer.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

This section contains the reviewer's results of the exploratory analysis using Cox-proportional hazard model on the time from randomization to relapse during the DB period for different population subgroups using baseline weight category (≥ 40 kg, and < 40 kg) as a stratification factor (see Table 9). The data were grouped by gender and largest race groups (white/non-white). Grouping by other races or ethnicity produced too few patients per group. Grouping by country/region was not applicable since this was a US-only study. Grouping by age group (6-12 and 13-17) was not performed, since, as it is common for pediatric studies that the weight was strongly correlated with age and consequently there were no subjects with weight < 40 kg in the age group of 13-17 years old.

Table 9. Subgroup analysis of the time to relapse stratified by weight group (Randomized Sample).

Subgroup		N	Abilify	Placebo	Weight < 40 kg	Weight ≥ 40 kg	Abilify to Placebo Hazard Ratio (95% CI)
Randomized Sample	Relapsed	35	13	22	13	22	
	Censored	50	28	22	19	31	0.57
	Total	85	41	44	32	53	(0.28, 1.12)
Gender: Male	Relapse	28	10	18	10	18	
	Censored	40	20	20	12	28	0.69
	Total	68	30	38	22	46	(0.32, 1.50)
Gender: Female	Relapse	7	3	4	3	4	
	Censored	10	8	2	7	3	0.35
	Total	17	11	6	10	7	(0.08, 1.59)
Race: White	Relapse	25	8	17	9	16	
	Censored	34	23	11	11	23	0.34
	Total	59	31	28	20	39	(0.15, 0.79)
Race: Not white	Relapse	10	5	5	4	6	
	Censored	16	5	11	8	8	1.68
	Total	26	10	16	12	14	(0.48, 5.83)

Source: computed by the reviewer.

In a similar manner the reviewer performed the subgroup analysis for the Phase 2 *Safety Sample*, where the three observations (Patient ID: 0029-00115, 0003-00143, 0039-00090) that did not receive treatment during the double-blind phase were removed from the analysis. The results are summarized in Table 10 and appear to be consistent with the results based on the *Randomized Sample* (Table 9).

Table 10. Subgroup analysis of the time to relapse stratified by weight group (Phase 2 Safety Sample).

Subgroup		N	Abilify	Placebo	Weight < 40 kg	Weight ≥ 40 kg	Abilify to Placebo Hazard Ratio (95% CI)
Randomized Sample	Relapsed	35	13	22	13	22	
	Censored	47	26	21	18	29	0.57
	Total	82	39	43	31	51	(0.28, 1.12)
Gender: Male	Relapse	28	10	18	10	18	
	Censored	37	18	19	11	26	0.69
	Total	65	28	37	21	44	(0.32, 1.50)

Gender:	Relapse	7	3	4	3	4	
Female	Censored	10	8	2	7	3	0.35
	Total	17	11	6	10	7	(0.08, 1.59)
Race:	Relapse	25	8	17	9	16	
White	Censored	32	21	11	10	22	0.34
	Total	57	29	28	19	38	(0.15, 0.79)
Race:	Relapse	10	5	10	4	6	
Not white	Censored	15	5	5	8	7	1.68
	Total	25	10	15	12	13	(0.48, 5.83)

Source: computed by the reviewer.

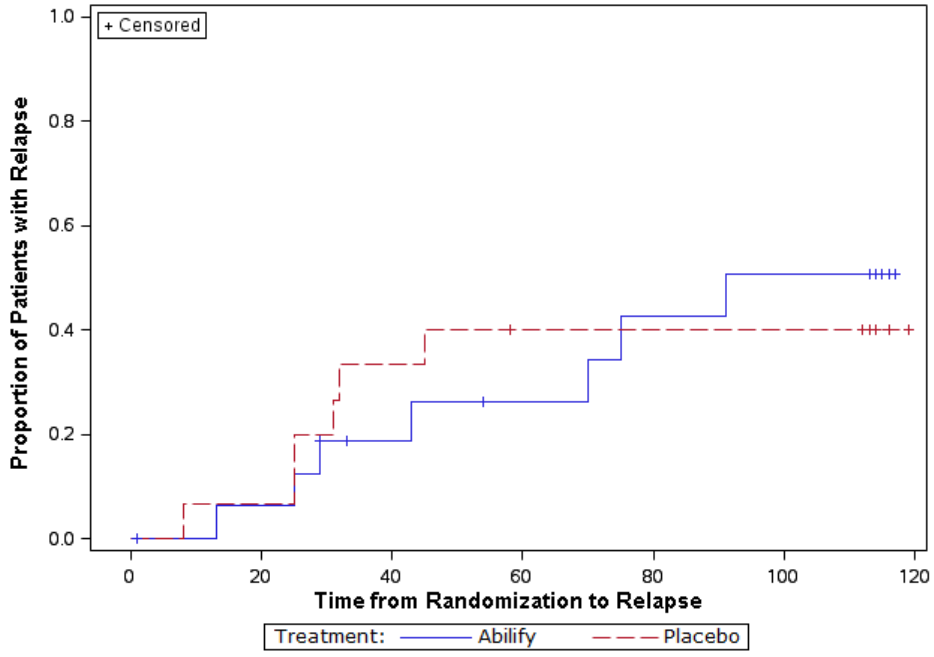
The majority of patients were male. In terms of race, the majority were white. The results of the subgroup analyses suggest somewhat an inconsistent trend between the white and the non-white, as also noted by the sponsor. However, taking into account that this exploratory analysis is post-hoc and that this trial was relatively small, particularly in the non-white subgroup, we cannot conclude that the observed disparity is real. Furthermore, the race subgroups might be confounded (for example) with the social-economic background, which was not accounted in the current trial and may require additional studies.

Although the sponsor concluded a treatment-by-race interaction (white vs. non-white) by comparing the relapse rates at Week 16 for white subjects (25.8% and 60.7% for aripiprazole and placebo, respectively), and for non-white subjects (50.0% and 31.3% for aripiprazole and placebo, respectively), the significance of this disparity finding is limited by the sample size of the study and the lack of overall significance of the treatment difference in the primary analysis result.

4.2 Other Special/Subgroup Populations

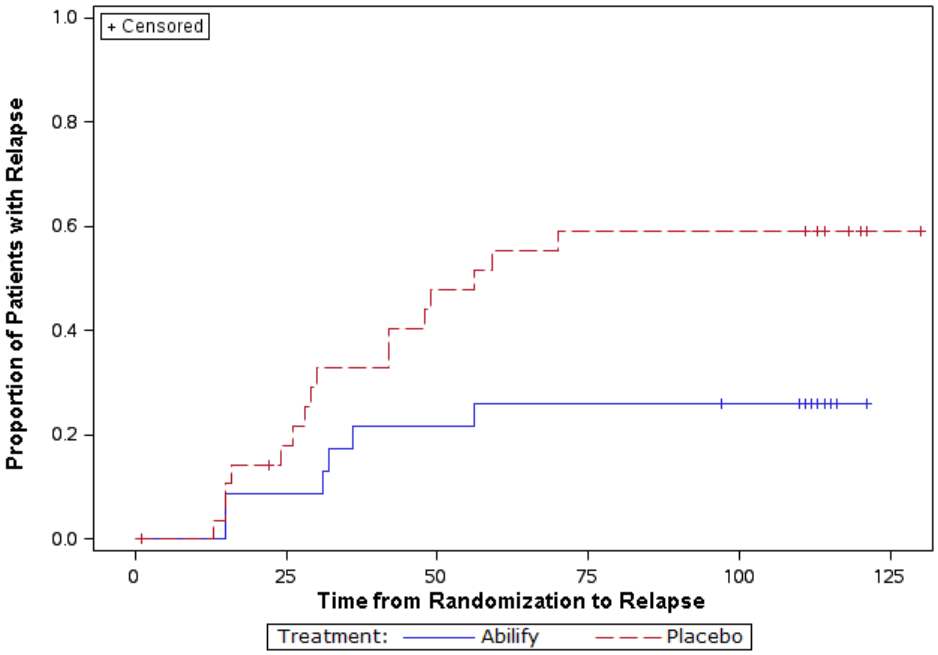
The primary efficacy analysis (log-rank test) was stratified by the baseline body weight group. The reviewer estimated the proportion of patients with relapse over time separately for each of the subgroup. The appropriate Kaplan-Meier curves are presented on Figure 4 (baseline weight ≥ 40 kg) and Figure 5 (baseline weight < 40 kg).

Figure 4. Kaplan-Meier Estimates of Proportion of Patients with Baseline Weight <40 kg Relapse over Time.



Source: computed by the reviewer.

Figure 5. Kaplan-Meier Estimates of Proportion of Patients with Baseline Weight ≥40 kg Relapse over Time.



Source: computed by the reviewer.

The difference between the aripiprazole and placebo (as measured by the relapse proportion) seems to be much large for patients with the DB phase baseline weight ≥ 40 kg. Some possible explanations for that may include the positive correlation between the patient's weight/age and

the titrated dose of the drug (e.g., heavier/older patients may have received higher doses of the drug, while younger/lighter patient may be more prone to adverse reaction, etc.). The summary of the relapsed/censored patients in each subgroup is presented in Table 11.

Table 11. Summary of the relapsed/censored patients in each weight subgroup.

	Baseline Weight < 40 kg			Baseline Weight ≥ 40 kg		
	Abilify	**Placebo**	**Total**	**Abilify**	**Placebo**	**Total**
Relapsed	7	6	13	6	16	22
Censored	10	9	18	18	13	31
Total	17	15	32	24	29	53

Source: computed by the reviewer.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The sponsor's findings on aripiprazole (2—15mg/day) were confirmed by the reviewer to be statistically non-significantly different from placebo (log-rank test p-value 0.0965) in reducing the symptoms of irritability associated with autistic disorder during the 16 weeks of double-blind treatment of pediatric subjects who maintained a response for 12 weeks of treatment.

The results of the exploratory subgroup analysis did not reveal inconsistencies between subgroups with respect to the age and gender. Although there seems to be an inconsistent trend between the white and the non-white, the finding is inconclusive because the majority of the population was white and the possible confounding effects were not accounted for in the trial.

5.2 Conclusions and Recommendations

Although the study failed, it appears that the sponsor conducted it in accordance with the statistical analysis plan agreed upon by the Agency.

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

/s/

ANDREJUS PARFIONOVAS
02/21/2014

PEILING YANG
02/23/2014

HSIEN MING J HUNG
02/24/2014