

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

209176Orig1s000

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: 209176

Drug Name: Edaravone

Indication(s): Amyotrophic Lateral Sclerosis (ALS)

Applicant: Mitsubishi

Date(s): June 16, 2016

Review Priority: Standard

Biometrics Division: Division of Biometrics I

Statistical Reviewer: Tristan Massie, Ph.D.

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

Hsien Ming (James) Hung, Ph.D., Division Director, DB I

Medical Division: Division of Neurology

Clinical Team: Christopher Breder, M.D.

Project Manager: Dan Jack

Keywords:

Replication issue(Second Study in Post-hoc subgroup of Failed First Study)

Table of Contents

LIST OF TABLES	3
LIST OF FIGURES	4
1. EXECUTIVE SUMMARY	5
2. INTRODUCTION.....	6
2.1 OVERVIEW.....	6
2.2 DATA SOURCES	7
3. STATISTICAL EVALUATION	8
3.1. DATA AND ANALYSIS QUALITY.....	8
3.2. EVALUATION OF EFFICACY.....	8
3.2.1. STUDY 16	8
3.2.1.1. STUDY DESIGN AND ENDPOINTS	8
3.2.1.2. STATISTICAL METHODOLOGIES	9
3.2.1.3. PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS	9
3.2.1.4. RESULTS AND CONCLUSIONS	10
3.2.1.4.1. SPONSOR'S RESULTS	10
3.2.1.4.2. REVIEWER'S RESULTS	13
3.2.2. STUDY 19	20
3.2.2.1. STUDY DESIGN AND ENDPOINTS	20
3.2.2.2. STATISTICAL METHODOLOGIES	23
3.2.2.3. PATIENT DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS	25
3.2.2.4. RESULTS AND CONCLUSIONS	26
3.2.2.4.1. SPONSOR'S RESULTS	26
3.2.2.4.2. REVIEWER'S RESULTS	30
3.3. EVALUATION OF SAFETY	36
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS.....	36
4.1 GEOGRAPHIC REGION, GENDER, RACE AND AGE	36
4.2 GEOGRAPHIC REGION	37
4.3 OTHER SPECIAL/SUBGROUP POPULATIONS.....	38
5. SUMMARY AND CONCLUSIONS	39
5.1 STATISTICAL ISSUES	39
5.2 COLLECTIVE EVIDENCE	40
5.3 CONCLUSIONS AND RECOMMENDATIONS	41

LIST OF TABLES

Table 1 Phase 3 Efficacy Studies of Edaravone in ALS.....	6
Table 2 Difference in ALSFRS-R between Baseline in Cycle 1 and the End of.....	11
Table 3 Study 16: Death or certain disease progression events up to Cycle 6 (FAS)	12
Table 4 Survival analysis for death or certain disease progression up to Cycle 12 (FAS).....	13
Table 5 Study 19 Patient Disposition through Cycle 6 (Double blind placebo controlled period)	25
Table 6 Study 19 Baseline Demographics and Disease Characteristics	26
Table 7 Analysis of change in ALSFRS-R score from baseline in Cycle 1 to the end of Cycle6 (LOCF) (FAS).....	27
Table 8 Number of events involving death or certain disease progression (FAS) up to 6 months	28
Table 9 Analysis of change in %FVC from baseline in Cycle 1 to the end of Cycle 6(LOCF) (FAS).....	29
Table 10 Analysis of change in Limb Norris Scale score from baseline in Cycle 1 to the end of Cycle 6 (LOCF) (FAS)	29
Table 11 Analysis of change in modified Norris Scale score (total) from baseline in Cycle 1 to the end of Cycle 6 (LOCF) (FAS)	30
Table 12 Analysis of change in ALSAQ40 score from baseline in Cycle 1 to the end of Cycle6 (LOCF) (FAS).....	30
Table 13 ALSFRS Treatment Group Difference at the end of Cycle 6 by baseline ALSFRS subgroups (MMRM)	33
Table 14Total number of events involving death or certain disease progression through Cycle 12(FAS).....	35
Table 15 Study 19: Estimated Treatment Differences on ALSFRS-R at Cycle 6 by Gender	37
Table 16 Study 19: Estimated Treatment Differences on ALSFRS-R at Cycle 6 by Age Group	37
Table 17 Summary of Efficacy Results from Phase III studies in ALS	40
Table 18 Sponsor's Analyses in Subgroups.....	41

LIST OF FIGURES

Figure 1 Study 16: Outline of Study Design.....	8
Figure 2 Checking the Normality of ALSFRS-R changes at Cycle 6 in the EESP/Def/2y subgroup.....	15
Figure 3 Exploratory Analysis of Treatment Differences by Quintile of Baseline ALSFRS-R (ITT).....	17
Figure 4 ALSFRS-R Changes at Cycle 6 as a function of baseline ALSFRS-R by group and efficacy subgroup classification.....	18
Figure 5 Distribution of ALSFRS-R at Baseline by EESP ⁺ subgroup Status.....	19
Figure 6 Study 19: Study Design	21
Figure 7 Study 19 Dynamic Allocation Flow Chart	22
Figure 8 Study 19: ALSFRS-R score (mean ± SD) (FAS)	28
Figure 9 Study 19 Normality Check for ALSFRS-R Changes at Cycle 6.....	31
Figure 10 ALSFRS-R Change from Baseline to Cycle 6 as a function of baseline ALSFRS-R..	34
Figure 11 Study 19 Mean Treatment Difference in Change from Baseline in ALSFRSR by Site at Cycle 6	38

1. EXECUTIVE SUMMARY

The submission's first study, 16, failed on its primary endpoint and the second study was designed on the basis of a post-hoc subgroup result in the first study. One could always data dredge a failed study for a 'positive' subgroup and then showing that subgroup is positive in a new study does not erase the fact that the first study failed. It requires circular logic to consider the second study to validate the first so that the submission might approach the efficacy standard of two positive studies. Objectively, the evidence here is not the same amount of evidence as two positive studies, positive on their originally prespecified primary hypotheses. This reviewer has observed many cases where attempts to validate a post-hoc subgroup finding in a follow-up study designed based on that subgroup failed but "validation" would tend to happen by chance alone at the rate of 1 in 20 (two-sided). Note, also that the studies in this submission were done exclusively in Japanese subjects, so there is no data in American subjects. Furthermore, the studies' double blind treatment periods were 24 weeks which is relatively short for studies in ALS.

If the post hoc subgroup from the first study is truly the correct subgroup for efficacy of the drug then it suggests that the drug effect is limited to less severe ALSFRS-R at baseline subjects or at least that the drug effect diminishes with increasing severity and longer disease duration since the study failed overall. It may be important to better understand through experimentation when in the course of a patient's disease the drug may be no longer helpful for the sake of long term treatment of individual patients. If we consider this post hoc subgroup of the overall non-positive first study as partial evidence then it seems incumbent upon us to re-evaluate the question of diminishing drug effect by baseline severity in the follow-up trial as well since there is still uncertainty about whether the drug has an effect in more severe ALS, especially since no effect on survival was shown and yet mortality becomes notable in the course of the natural history of the disease. Overall, study 19 seems to support an effect of the drug on ALSFRS-R change from baseline but an exploration of efficacy by disease severity may also support the sponsor's conclusion for study 16 that it is only effective for certain less severe patients and exploratory analyses of study 19 suggest it could be a smaller subgroup than the sponsor hypothesized based on study 16. However, we acknowledge that study 19 was positive overall on the primary endpoint and it is reasonably consistent across the range of baseline severity in the study 19 population.

The "Minimization" randomization used would balance the treatment groups overall but if enrollment of the post-hoc subgroup was not uniform over time, for example, or since the minimization didn't take the post-hoc subgroup's balance into account then there could be imbalance between treatment groups within the post-hoc subgroup. For example, in the post-hoc subgroup 21.9% of Placebo were not treated with Riluzole as compared to just 7.5% in Edaravone. On the other hand, in the complementary subgroup these proportions were 7% for Placebo and 13% for Edaravone. Thus, between these two complementary subgroups the odds ratio of having taken Riluzole were somewhat inconsistent. In fact, an exploratory statistical test of this, a Breslow Day test for Homogeneity of Odds ratios, rejects homogeneity between these two complementary subgroups, $p=0.0347$. Also, in the Excluded subgroup 6% of placebo had

familial ALS as compared to 0% in the Edaravone group, p=0.08. These exploratory findings of possible imbalances in study 16 reinforce the difficulty of justifying a post-hoc analysis and of interpreting a post-hoc p-value.

Study 19 does not seem very persuasive on its own since many of the secondary analyses were not nominally significant (although the study was relatively small). Some alternative subjective interpretation (e.g., risk benefit considerations or orphan drug status) would seem necessary if one was to view this submission as sufficient for approval since it does not contain two studies that were each positive on the prespecified primary hypothesis.

2. INTRODUCTION

2.1 Overview

Edaravone (MCI-186) is a free radical-scavenger developed as a neuroprotectant by Mitsubishi Tanabe Pharma Corporation (the Sponsor). The IND number for ALS was 126396 but the efficacy studies were done in Japan and not under US IND. The first official communication between the Division of Neurology Products and the Sponsor started with a pre-investigational new drug (IND) application meeting on June 16, 2015.

Table 1 Phase 3 Efficacy Studies of Edaravone in ALS

Trial ID	Design*	Treatment/ Sample Size	Endpoint/Analysis	Preliminary Findings
19	DB (24 wks)	Drug / 69 Placebo/ 68	Primary: ALSFRS-R Change at 24 Weeks	P=0.0013 (0.0037)*
16	DB (24 wks)	Drug /100 (39) [#] Placebo/99 (29) [#]	Primary:ALSF RS-R Change at 24 Weeks	Overall: p=0.4108 Post-hoc subgroup [#] p=0.017 (.0147)

* MC: multi-center, R: randomized, DB: double-blind, PG: parallel group, PC: placebo controlled, AC: active controlled
24 week active extension

[#]Post hoc subgroup

2.2 Data Sources

The derived and raw datasets for the key efficacy studies 16 and 19 were located in the following directories at the time of review.

Study 16 data:

<\\cdsesub1\evsprod\nda209176\0001\m5\datasets\mci186-16\analysis\legacy\datasets\>

Note that study 16 Subject 0504 died according to the study report but no death event is recorded for this subject in the M17Event dataset. Note that this subject is not in the post hoc efficacy subset.

Study 19 data:

<\\cdsesub1\evsprod\nda209176\0001\m5\datasets\mci186-19\analysis\legacy\datasets\>

Four Subjects in study 19 are listed as deaths in the study report but not listed in the ADEVT dataset or the disposition dataset (DS).

SUBJID=M190102 SITEID=01 ARM=M(Edaravone) deathday=342
SUBJID=M190103 SITEID=01 ARM=P (Placebo) deathday=316
SUBJID=M191208 SITEID=12 ARM=P(Placebo) deathday=289
SUBJID=M192606 SITEID=26 ARM=P(Placebo) deathday=266

3. STATISTICAL EVALUATION

3.1. Data and Analysis Quality

The submitted data and analysis quality appear adequate.

3.2. Evaluation of Efficacy

3.2.1. Study 16

The study was initiated on May 8, 2006 and the last patient completed the double blind treatment period portion of the study on Sept 9, 2008. The active extension period of the study was completed on May 12, 2009.

3.2.1.1. Study Design and Endpoints

This was a multicenter, placebo-controlled, double-blind, parallel-group comparative study.

Allocation of investigational product

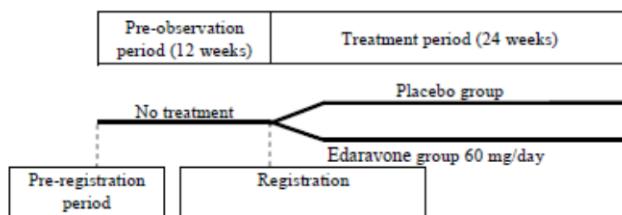
Patients were assigned to either the edaravone or placebo group in a 1:1 fashion. The investigational product was dynamically assigned by the minimization method with the following 3 factors.

Factor: 1)“ALSFRS-R score changes from the start to end (12 weeks later) of the pre-observation period,” 2)“initial symptom (bulbar or limbic)” and 3)“concomitant use of Riluzole”.

Each treatment cycle consisted of treatment and drug free periods, and this treatment cycle was repeated 6 times (approximately 24 weeks).

o shows the outline of the study design.

Figure 1 Study 16: Outline of Study Design



Pre-observation period: A 12-week observation period before the start of Treatment cycle 1 was designed.

Treatment cycle 1: The investigational product was administered for 14 consecutive days, followed by a 2-week drug free period.

Treatment cycles 2 to 6: The investigational product was administered for a total of 10 days per 2 weeks, followed by a 2-week drug free period.

3.2.1.2. Statistical Methodologies

The Primary Analysis according to the Dec 11, 2008 plan as well as the original protocol dated April 6, 2006 and the final protocol dated July 20, 2007 was to be as follows.

The efficacy of edaravone would be regarded as having been verified when either or both of the following analyses (i) and (ii) have demonstrated a significant difference between the edaravone group and the placebo group.

(i) The change from “baseline in Treatment Cycle 1” to “the end of Treatment Cycle 6 or at discontinuation (LOCF)” would be analyzed by using the factors in the dynamic allocation as covariates to perform treatment group comparisons.

Analysis model: The change in scores from “baseline in Treatment Cycle 1” to “the end of Treatment Cycle 6 or at discontinuation (LOCF)” = treatment groups + dynamic allocation factors

(ii) With respect to the scores at each time point, a repeated measurements analysis of variance would be performed by using treatment groups, time points, and the interaction between treatment group and time point as factors and “baseline in Treatment Cycle 1,” and factors used for dynamic allocation as covariates, for comparison difference between the treatment groups. The following additional details were specified in the statistical analysis plan but not in the protocol. The structure of the covariance matrix used for this analysis is compound symmetry.

Analysis model: Scores at each time point = treatment groups + time points + treatment group \times time point + “baseline in Treatment Cycle 1” + dynamic allocation factors If there is an interaction between treatment group and time point, differences in the profiles would be investigated in detail. Especially, group comparison would be performed using the final point (the end of Treatment Cycle 6) as a primary evaluation point. The presence or absence of interaction effect would be determined with the significance level of 15% two-sided as a reference value.

Full Analysis Set (FAS)

The FAS was defined as an analysis set with exclusion of the following patients.

- Patients with diseases other than the target disease
- Patients with significant Good Clinical Practice violations
- Patients who are not treated with the investigational product
- Patients with no efficacy data available

The target sample size was set at 100 subjects in each group. With a difference in change of ALSFRS-R between treatment groups of 2.2 points based on the results of ALSFRS-R change in Study MCI186-12, this sample size would provide 95% power when the SD is 4.3, 85% power when the SD is 5.2, and 67% power when the SD is 6.5.

3.2.1.3. Patient Disposition, Demographic and Baseline Characteristics

The investigational product was administered to 206 patients (102 patients in the edaravone group and 104 patients in the placebo group). The Full Analysis Set (FAS) consisted of 205 patients (101 patients in the edaravone group and 104 patients in the placebo group). The Per Protocol Set (PPS) consisted of 195 patients (97 patients in the edaravone group and 98 patients in the placebo group).

3.2.1.4. Results and Conclusions

3.2.1.4.1. Sponsor's Results

The study failed to verify the efficacy of edaravone in the FAS. As the sponsor later presented this result to the Pharmaceutical and Medical Devices Agency of Japan (PMDA) on February 13, 2009 at a prior consultation, the PMDA advised as follows: “Although efficacy has not been disproved, it is not considered to have been clarified. Because investigation is insufficient, we suggest that additional analyses be performed to investigate in what population edaravone is effective.” Based on this advice from the Japanese regulatory authority, the sponsor performed additional analyses. Because the sponsor discovered a patient population in which edaravone may be effective, i.e., the Efficacy Expected Sub-population (EESP), the sponsor had a prior consultation with the PMDA on September 3, 2009, in which the PMDA advised as follows: “The result of additional analyses showing that edaravone is shown to be effective more clearly in patients with mild ALS is reasonable and is understandable from the viewpoint of the mechanism of action of edaravone. The results of the additional analyses of the confirmatory study and the extension study alone do not ensure application and approval. We think that a separate study must be conducted to verify the efficacy of edaravone.” The sponsor subsequently focused on diagnostic criteria. Because the sponsor thought that the difference in efficacy from placebo becomes larger, and the efficacy of edaravone can be shown more clearly with increasing accuracy of diagnosis, the sponsor had a face-to-face consultation with the PMDA on January 14, 2010 for the planning of the second confirmatory study. The sponsor prepared a second version of this clinical study report for study 16 because they had performed additional analyses of the study as described above. In the second version of the report, the sponsor described the results of analyses in the EESP, which was planned to be used for the verification of repeatability in the extension study, and the definite or probable/EESP/2y, which was investigated in confirmatory study 2, because the definite/EESP/2y was not investigated in the second confirmatory study in the end. (The first version of the clinical study report was prepared on September 17, 2009.) The EESP was added to the analysis plan as a subgroup of interest for the extension study on September 24, 2009.

The results of the analysis of the primary efficacy endpoint of change from Baseline to the end of Cycle 6 in ALSFRS-R are shown in

Table 2. While a beneficial trend favoring edaravone was observed in the FAS, the prespecified primary analyses did not statistically demonstrate the efficacy of edaravone in comparison to placebo.

In additional exploratory analyses to evaluate the beneficial trend observed with edaravone, the beneficial trend favoring edaravone was mainly driven by data from subjects who had functionality retained in most ADL domains with normal respiratory function. This population was described as the “EESP”. Note that this was not a prespecified subgroup.

Efficacy Expected Subpopulation (EESP) Definition: Subjects who met the following criteria

- Each individual item of the ALSFRS-R of 2 and better at Baseline (i.e., Functionality retained in most ADL domains)
- A %FVC of 80% and greater at Baseline (i.e., Normal respiratory function).

Subjects meeting the following two criteria as well as the EESP criteria were defined as “Definite or Probable/EESP/2y”.

Definite or probable/EESP/2y Definition: In addition to meeting EESP criteria,

- Definite or Probable ALS diagnosis based on the El Escorial and revised Airlie House diagnostic criteria at preregistration (To ensure diagnosis of ALS)
- Within 2 years of initial ALS symptom onset at preregistration (To exclude subjects who were stable for long-term with ALS).

The Post-hoc Expected Efficacy Subpopulation (EESP) consisted of 104 patients (54 patients in the edaravone group and 50 patients in the placebo group), and the definite or probable/EESP/2y consisted of 72 patients (40 patients in the edaravone group and 32 patients in the placebo group).

The Sponsor analyzed changes in ALSFRS-R scores in these study subgroups. The EESP and Definite or Probable/EESP/2y showed favorable trends with p-values less than 0.05 as shown in Table 2 but these were post-hoc.

Table 2 Difference in ALSFRS-R between Baseline in Cycle 1 and the End of Cycle 6 (LOCF) for Study MCI186-16 (FAS, EESP, Definite or Probable/EESP/2y)

Study number (Population)	Group	Number of subjects in LOCF analyses	Adjusted mean	Between-group differences in the adjusted mean	p-value
			LS mean±SE	LS mean±SE (95% CI)	
MCI186-16 (FAS)	P group	99	-6.35±0.84	0.65±0.78 (-0.90 , 2.19)	p=0.4108
	E group	100	-5.70±0.85		
MCI186-16 (EESP) ^a	P group	46	-7.06±1.13	2.20±1.03 (0.15 , 4.26)	p=0.0360
	E group	53	-4.85±1.24		
MCI186-16 (definite or probable/EESP/2y) ^a	P group	29	-7.59±1.34	3.01±1.33 (0.35 , 5.67)	p=0.0270
	E group	39	-4.58±1.55		
MCI186-16 (non-EESP) ^a	P group	53	-5.24±1.25	-1.42±1.16 (-3.73 , 0.89)	p=0.2251
	E group	47	-6.65±1.17		
MCI186-16 (non-“definite or probable/EESP/2y”) ^a	P group	70	-5.54±1.08	-0.57±1.00 (-2.55 , 1.41)	p=0.5711
	E group	61	-6.11±1.03		

Note: LOCF was applied to subjects who completed Cycle 3 (subjects who reached 81 days after treatment initiation). Subjects who dropped out before Day 81 were excluded.

^a EESP, Definite or Probable/EESP/2y, non-EESP, and non-“Definite or Probable EESP/2y” analyses were post-hoc.

Note: This table was copied from page 28 of the summary of clinical efficacy

Time to Death or Certain Disease Progression

In the examination of time to death or certain disease progression, the prespecified events of interest were death, disability of independent ambulation, loss of upper arm function, tracheotomy, use of respirator, and use of tube feeding. Table 3 shows the number of such events (FAS) by group. There were 38 events in the edaravone group and 37 events in the placebo group, with more frequent onset of "disability of independent ambulation" and "use of tube feeding" in the edaravone group. There were 2 events each of death in the edaravone and placebo groups. "Disability of independent ambulation" was the most common event, with 28 events in the edaravone group and 23 events in the placebo group.

Any of "death, disability of independent ambulation, loss of upper arm function, tracheotomy, use of respirator, and use of tube feeding" was defined as an event, and other cases as censored. The last day of observation was regarded as the censored day. In patients who experienced multiple events, the day of onset of the first event was defined as the day of event. Any of these events occurred in 32 patients ("−4, −3" change from screening to baseline: 12 patients; "−2, −1" Pre-change: 20 patients) in the edaravone group and in 27 patients ("−4, −3": 14 patients; "−2, −1": 13 patients) in the placebo group, showing no significant difference between the 2 groups ($P = 0.3814$, stratified log-rank test; $P = 0.3992$, stratified generalized Wilcoxon test).

Table 3 Study 16: Death or certain disease progression events up to Cycle 6 (FAS)

Event	Placebo	Edaravone	Total
Death	2	2	4
Disability of independent ambulation	23	28	51
Loss of upper limbs function	4	2	6
Tracheotomy	2	0	2
Use of respirator	3	1	4
Use of tube feeding	3	5	8

Note: This table was copied from page 114 of the sponsor's study 16 study report

In the examination of time to death or certain disease progression in the definite or probable/EESP/2y post-hoc subgroup there were 7 events in the edaravone group and 11 events in the placebo group thus showing no significant difference between the 2 groups ($P = .5782$, stratified log-rank test; $P = 0.6963$, stratified generalized Wilcoxon test).. "Disability of independent ambulation" was the most common event, with 7 events in the edaravone group and 8 events in the placebo group.

A survival analysis of time to disease progression was also performed for events through the extension phase (up to Cycle 12). The number of patients with an event in each group was 11 patients(4 patients with "−4, −3" and 7 patients with "−2, −1") in the edaravone-placebo group and 18 patients(7 patients with "−4, −3" and 11 patients with "−2, −1") in the edaravone-edaravone group. The results are shown in Table 4. Note that the Placebo-Edaravone group was not formally compared to the others (those who started with Edaravone).

Table 4 Survival analysis for death or certain disease progression up to Cycle 12 (FAS)

Change in ALSFRS-R score during the pre-observation period	Treatment group	No. of patients	Number of events*	Between-group comparison
-4, -3	Edaravone-placebo	11	4	$P=0.1540$ (Stratified log-rank test)
	Edaravone-edaravone	13	7	
-2, -1	Edaravone-placebo	33	7	$P=0.0684$ (Stratified generalized Wilcoxon test)
	Edaravone-edaravone	35	11	

*: Death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, and use of tube feeding.

For patients with more than one event, the onset date of the first event was defined as the date of event onset.

Events that occurred by the end of Cycle 12

Note: This table was copied from page 117 of sponsor's study 17 (extension of 16) study report

In the examination of time to death or certain disease progression in the definite or probable/EESP/2y subgroup through Cycle 12 there were 5 patients (1 patients with "-4, -3" and 4 patients with "-2, -1") in the edaravone-placebo group and 8 patients (3 patients with "-4, -3" and 5 patients with "-2, -1") in the edaravone-edaravone group showing no significant difference between the 2 groups (stratified log-rank test; $P = 0.3802$).

ALSFRS-R from Cycle 7 through 12 (Active Treatment Period)

According to the analysis plan the main comparison of interest was Edaravone-placebo to Edaravone-Edaravone. Analysis of the Placebo-Edaravone group was only descriptive.

The changes from "baseline in Cycle 7" to "the end of Cycle 12 (or discontinuation, LOCF)" were compared between the groups by an analysis using "change in ALSFRS-R score during the pre-observation period" as a covariate.

In the FAS, the LSMean \pm SE in each group was -5.58 ± 0.74 in the edaravone-placebo group and -4.42 ± 0.69 in the edaravone-edaravone group. The LSMean \pm SE of the between-group difference was 1.16 ± 0.93 with the 95% confidence interval of -0.70 to 3.01 . In the EESP, the LSMean \pm SE in each group was -5.86 ± 0.98 in the edaravone-placebo group and -4.01 ± 0.86 in the edaravone-edaravone group. The LSMean \pm SE of the between-group difference was 1.85 ± 1.14 with the 95% confidence interval of -0.45 to 4.15 . Note that the sample sizes for the Cycle 6 analysis were 29 and 39. In the EESP, the number of dropouts was 2 in the edaravone-placebo group, 3 in the edaravone-edaravone group, and 7 in the placebo-edaravone group.

3.2.1.4.2. Reviewer's Results

Missing data at the end of the first six cycles of treatment (double blind placebo controlled period) was moderately low $10/102=10\%$ for Edaravone and $14/103=13.6\%$ for Placebo. Based on a Wilcoxon test of the joint rank (combined ranking of ALSFRS-R measured function and survival time in case of death) the p-value for the all randomized population was 0.3909.

Based on an MMRM, repeated measures analysis, with an unstructured covariance assumption within patient (30 covariance parameters) the treatment difference on ALSFRS-R at the end of cycle 6 in the post-hoc subgroup was -3.41 ± 1.28 S.E, $p= 0.0100$. In the Overall ITT population the corresponding estimate was -0.8232 ± 0.8466 S.E, $p= 0.3320$.

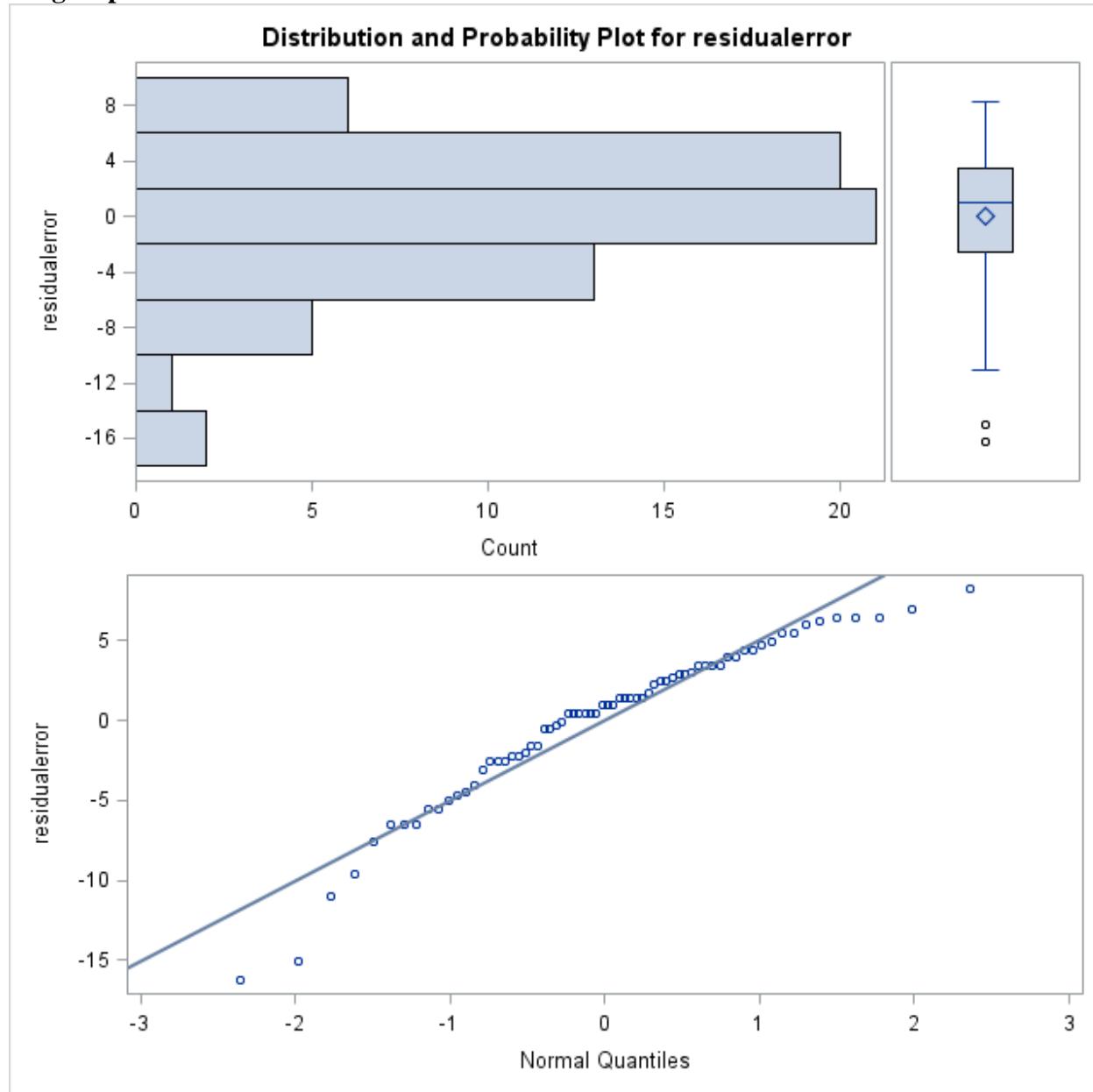
This reviewer notes that the sponsor had in effect specified two primary analyses either of which they believed could establish efficacy without adjustment for multiplicity which is of course problematic: i) LOCF ANCOVA and ii)a repeated measures analysis of variance with compound symmetric assumption for the structure of the correlation between repeated measures within patients over time. The latter analysis had an estimated difference of $-.98 \pm .51$ S.E., $p=0.0565$ in the overall randomized population including the insignificant visit*treatment interaction term ($p=0.883$) in the model which is usually done and required to estimate the difference at the last visit (excluding this interaction the test for treatment effect had an estimate of $-.67 \pm .43$ S.E., $p=0.126$). However, there is evidence that the compound symmetric covariance assumption (equal variance across visits and equal correlation between all visits within patient) does not fit the data well. For example, the Akaike's information criterion (AIC) measure of model fit was 4293 for the model with the unstructured covariance matrix assumption and 5363 for the compound symmetric covariance assumption. Smaller values of AIC suggest better model fit (note: AIC includes a penalty for having more parameters in the model). The large difference in AIC suggests that a likelihood ratio test (closely related to the difference in AIC) would reject the simpler compound symmetry covariance assumption as inadequate. Examining the estimated covariance matrices from the two models it seems that the variance is considerably larger at the final visit (and in fact it grows roughly linearly as Visit increases from Baseline). Also, although the model with compound symmetric covariance assumption has $p=0.0565$ which is somewhat close to 0.05 the size of the treatment effect estimated with it is small, much smaller than that in the post-hoc subgroup and that which we will see below was obtained for the same population in study 19. Therefore, the prespecified repeated measures ANOVA model is considered inadequate for these data because of the compound symmetric covariance assumption and so the overall p-value based on the MMRM analysis should be 0.332 as reported above.

This reviewer notes that in the complement of the post-hoc subgroup, i.e., Overall minus the post-hoc subgroup, the estimated effect on the change from baseline in ALSFRS-R at the end of Cycle 6 was $-.46 \pm 1.00$ favoring placebo numerically, $p=0.6487$.

The sponsor's pre-specified primary analysis did not take deaths into account. There were 2 deaths on drug and 2 deaths on placebo in the first 6 cycles of treatment (double blind treatment period). Based on a Wilcoxon test of the joint rank analysis of ALSFRS-R change from baseline and survival time in case of death the p-value for treatment difference in the post-hoc subgroup ($N=72$) was $p=0.0898$ and excluding those not reaching 3 months as specified in the protocol ($N=68$) $p=0.1094$.

This reviewer found that the normality of the residuals from the sponsor's ANCOVA LOCF model is questionable (Figure 2) in the post hoc subgroup EESP/Def/2y. In particular, they fail all of the standard tests of normality at the 0.05 significance level. This means that the sponsor's reported p-value of 0.027 may be incorrect. A nonparametric Wilcoxon rank sum test in this subgroup not relying on any distributional assumption for the ALSFRS-R changes gave a p-value of 0.1098 and adding adjustments for the stratification factors (in a RANK ANCOVA) the p-value was 0.199. In this subgroup a Wilcoxon test of the joint rank of ALSFRS-R change and survival gave a p-value of 0.1085. Therefore, the post-hoc subgroup result is not robust to apparent non-normality or proper handling of deaths in the analysis.

Figure 2 Checking the Normality of ALSFRS-R changes at Cycle 6 in the EESP/Def/2y subgroup



In an exploratory analysis suggested by the sponsor's post-hoc claim of efficacy in a subgroup this reviewer found that there is an interaction between EESP classification and baseline ALSFRSR value ($p=0.0003$) and a three way interaction between EESP, baseline ALSFRSR value and treatment assignment: $p=0.0804$. The implication of these interactions is as follows. Although the sponsor claimed that there is a treatment effect in lesser affected patients, paradoxically, within the lesser affected subgroup the effect when examined as a function of baseline severity seems to increase with increasing baseline severity (Figure 4). If the effect was truly bigger in less severe patients we would expect the opposite. Therefore, this post-hoc

subgroup effect could just be a random high driven by a few cases of extreme/outlying worsening from placebo patients with milder severity in the post-hoc efficacy subgroup.

This reviewer noticed that there was a quantitative interaction in the baseline ALSFRS-R score between treatment group and the EESP classification ($p=0.06$). In particular, the mean baseline ALSFRS-R score was smaller for Edaravone in the EESP positive subgroup (38.5 vs. 40.0) reaching nominal significance and nearly balanced in the EESP negative subgroup (42.4 vs. 42.5). The interaction is slightly more significant if we add the additional criteria characterizing the smaller post-hoc subgroup ($p=0.04$). This may partly explain the observed differences between the apparent efficacy in these subgroups. In the overall population the treatment difference was numerically biggest in favor of Edaravone in the lowest baseline score tertile which again calls the sponsor's post-hoc theory into question although the differences (shown below) were slight.

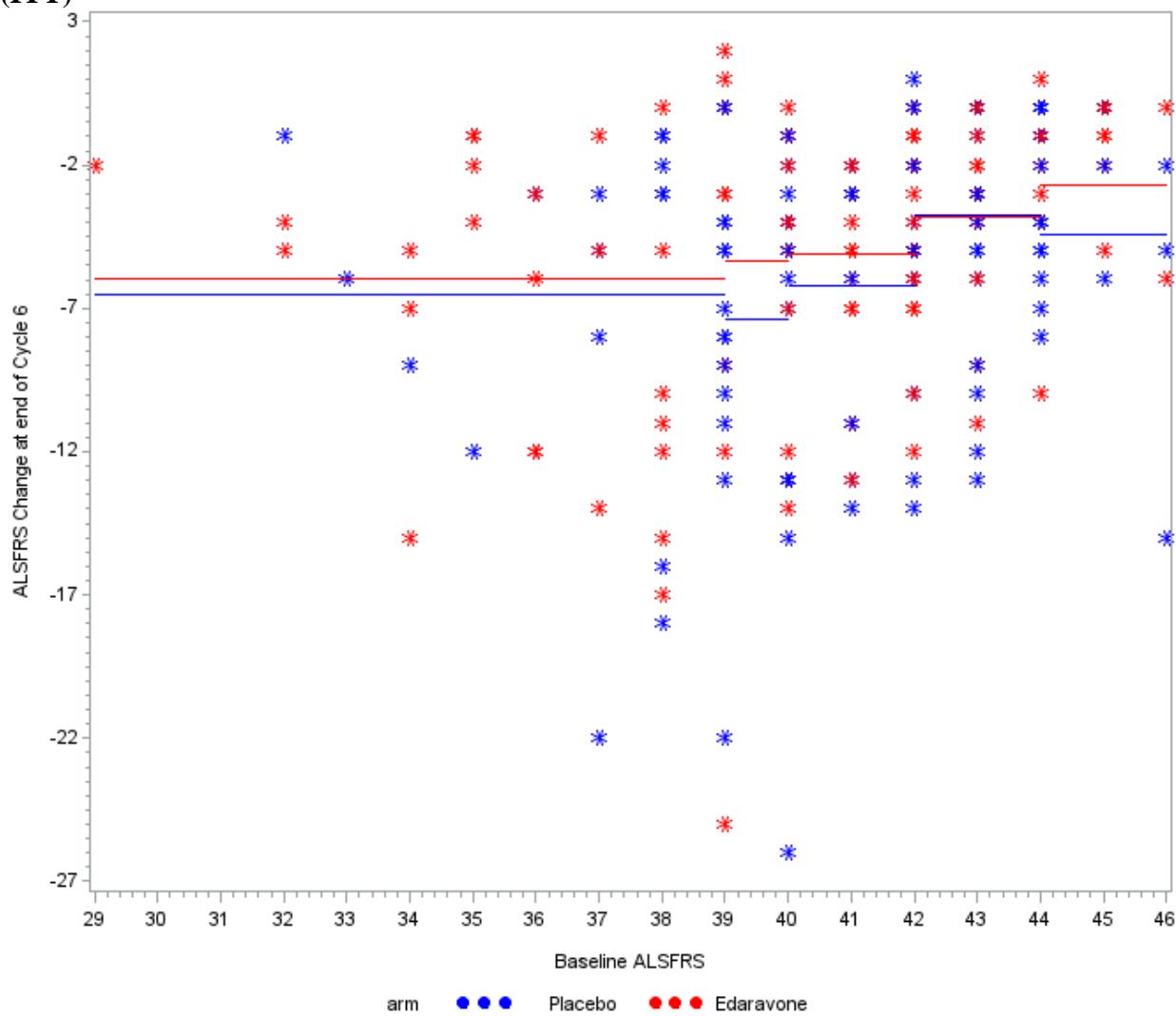
baseline ALSFRS-R tertile 1: 1.1 ± 1.2 S.E.;

baseline ALSFRS-R tertile 2: 0.9 ± 1.3 S.E.;

baseline ALSFRS-R tertile 3: 0.7 ± 1.7 S.E...

In the overall population treatment differences were also inconsistent across quintiles of the baseline ALSFRS-R score (Figure 3). In the second highest quintile the effect was the smallest among the quintile specific differences which seems to be somewhat contradictory to the sponsor's theory that the efficacy should be expected in less severe patients.

Figure 3 Exploratory Analysis of Treatment Differences by Quintile of Baseline ALSFRS-R (ITT)



Additional exploratory analyses performed by this reviewer in the overall population suggest that the change from baseline in ALSFRS-R may depend on the baseline ALSFRS-R in a non-linear way (i.e., quadratic or cubic terms in the baseline score may be needed) as seen in Figure 4. Cubic polynomials in the baseline ALSFRS-R were fit to the data within each possible combination of expected efficacy subgroup status and treatment group assignment which illustrate this.

Figure 4 ALSFRS-R Changes at Cycle 6 as a function of baseline ALSFRS-R by group and efficacy subgroup classification

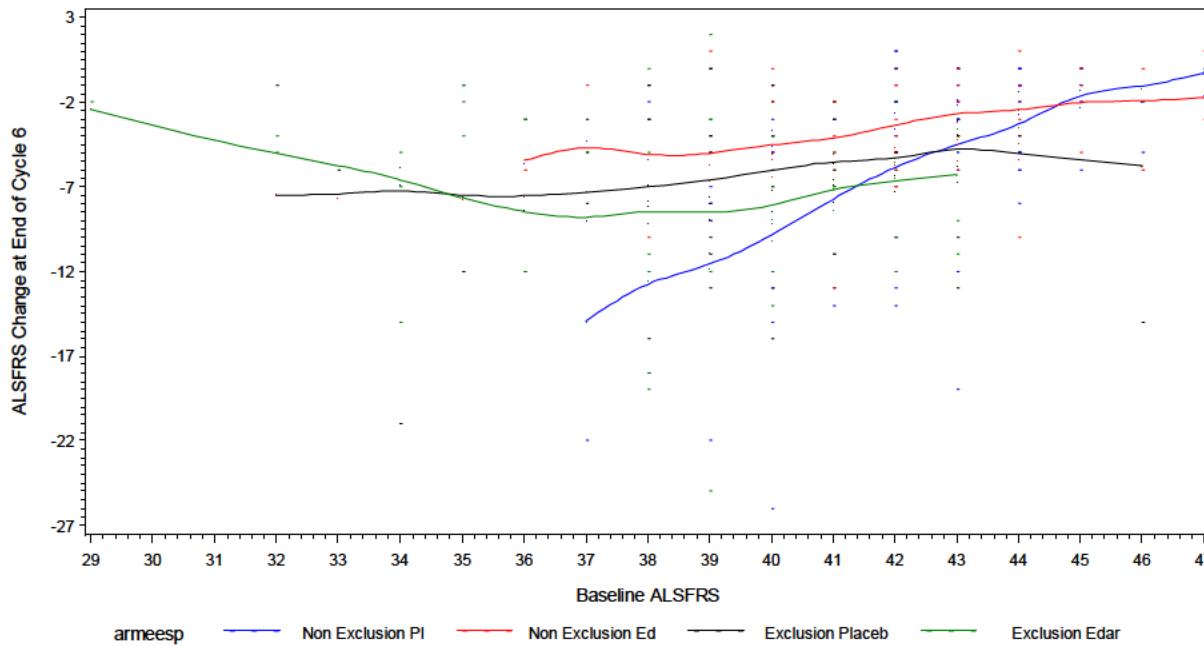
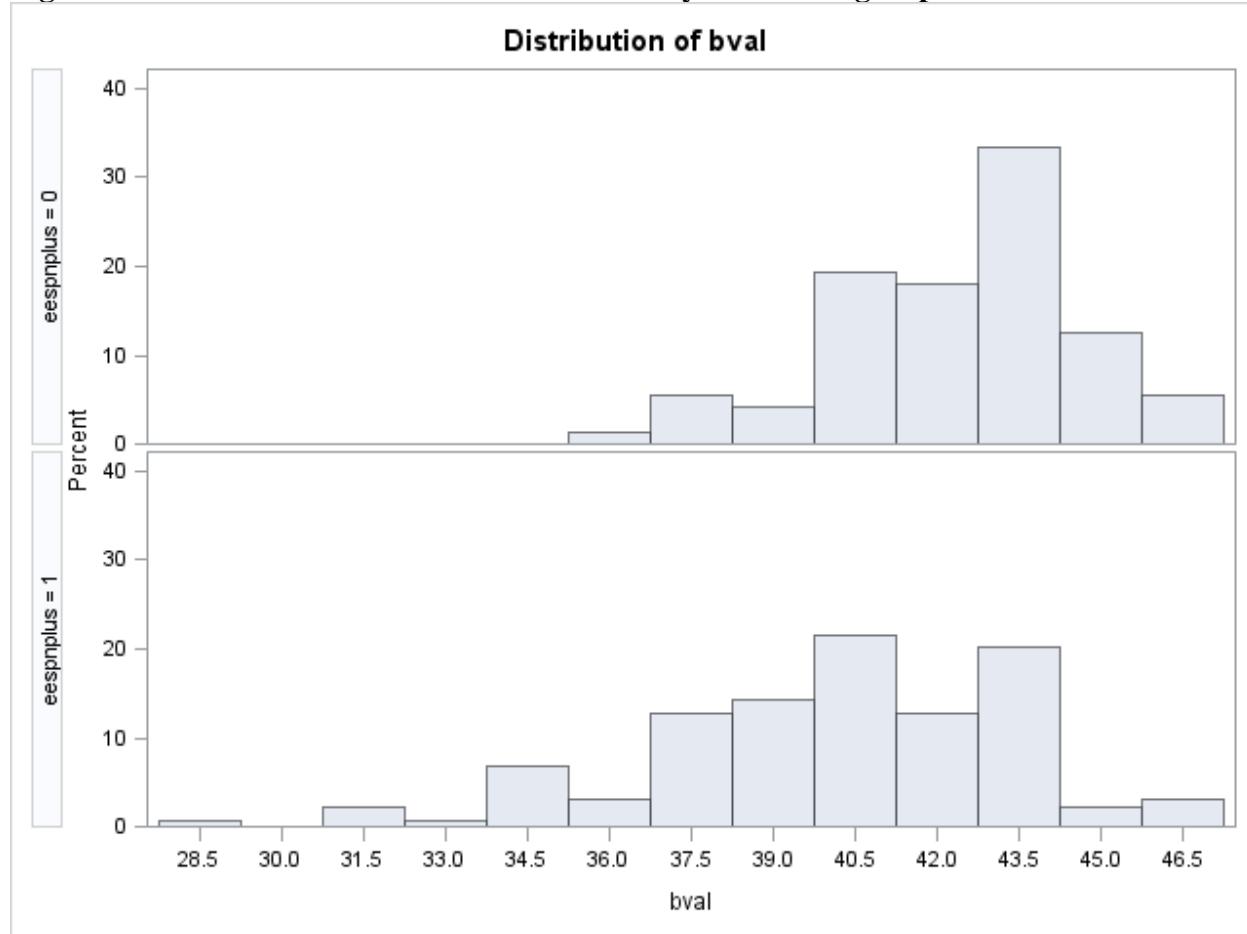


Figure 5 shows that the baseline ALSFRS-R tends to be higher in the non- EESP subgroup but at the same time it also shows that, despite this, there were still patients with high baseline scores in the non-“efficacy expected” subgroup.

The balance of the factors (BLPRECHG=change from pre-baseline to baseline, initial symptoms [bulbar vs. limbic], and Riluzole usage) was carefully maintained by means of utilizing the “Minimization” type of randomization. This tends to balance the groups over all baseline characteristics because whether or not the three chosen factors are balanced the allocation is random (when there is an imbalance during the process that exceeds a predetermined amount the next assignment is randomized to restore balance with higher probability). However, it would not necessarily lead to balance in a subpopulation. If enrollment of the post-hoc subgroup was not uniform over time, for example, or since the minimization didn't take this subgroup's balance into account then there could be imbalance between treatment groups within such post-hoc subgroups. In fact, within the post-hoc subgroup (Probable/Definite/EESP/2y) 21.9% of placebo were not treated with Riluzole as compared to just 7.5% in Edaravone within the same subgroup. On the other hand, in the complementary subgroup these proportions were 7% for Placebo and 13% for Edaravone. Between the two subgroups the odds ratio of having taken Riluzole were somewhat inconsistent statistically. In particular, a Breslow Day test for Homogeneity of Odds ratios of getting Riluzole between these two complementary subgroups gave a p-value of

$p=0.0347$, suggesting that the odds are different. Also, in the Complement of the post-hoc Efficacy subgroup, i.e., the rest of the randomized patients, 6% of placebo had familial ALS as compared to 0% in the Edaravone group ($p=0.08$) and there were nominally significant differences between the treatment groups in both the pre-baseline and the baseline ALSFRS mean scores (Mean (SD) was 40.8 (3.0) for Placebo and 39.3 (3.4) for Edaravone) in this complementary subgroup. This highlights that caution is necessary when attempting to interpret the treatment effect in a post-hoc subgroup because such imbalances may occur and may partially explain any observed difference in the primary endpoint between treatment groups.

Figure 5 Distribution of ALSFRS-R at Baseline by EESP⁺ subgroup Status



A test for an interaction between treatment and EESP⁺ status on ALSFRS-R change at the end of Cycle 6 yielded a nominally significant p-value of 0.0159. The estimated effects in the EESP⁺ and non-EESP⁺ subgroups suggest that this interaction is of a qualitative nature since the treatment difference is in favor of Edaravone -2.1 ± 0.97 in the EESP⁺ negative subgroup and numerically in favor of placebo $+1.2 \pm 0.97$ in the EESP⁺ positive subgroup. However, note that this test is exploratory; overall the treatment difference was not significant. This suggests exploring efficacy by other potentially important baseline disease characteristics. For example, a test for an interaction between continuous valued disease duration and treatment group had a non-significant p-value of 0.4774. However, test for an interaction between disease duration

categorized by years and treatment had a p-value less than 0.10: p=0.0797. More specifically, estimates of treatment differences within disease duration year subgroups suggested a qualitative interaction: -2.44 +/- 1.2 for <1 year, +1.02 +/- .96 for 1 -2 years and 0.07 +/- 1.66 for 2+ years. This reviewer also noted that there was no suggestion of an interaction between treatment difference and definite vs. probable ALS diagnosis, p=0.98.

In summary, this reviewer has noted baseline imbalances between treatment groups within the post-hoc subgroup and non-normality of the primary efficacy measure which are more tangible reasons for being wary of the sponsor's post-hoc analysis result.

3.2.2. Study 19

A Phase III, Double-blind, Parallel-group Study of Edaravone (MCI-186) for Treatment of Amyotrophic Lateral Sclerosis

November 28, 2011 (date of first patient enrollment)

September 3, 2014 (date of last patient observation [not including the follow-up for adverse events])

OBJECTIVES

The objective of this study was to investigate the efficacy of edaravone 60 mg versus placebo administered daily in a double-blind, parallel-group comparison design in patients with early amyotrophic lateral sclerosis (ALS) by comparing the changes in ALSFRS-S scores at 24 weeks of treatment. Another objective is to investigate the safety of edaravone in patients with ALS.

3.2.2.1. Study Design and Endpoints

Multicenter, placebo-controlled, double-blind, parallel-group design

Group composition

Cycle 1-6 (double-blind period)

Edaravone group: Edaravone injection 30 mg, 2 ampoules per dose, once daily, 60-minute intravenous infusion

Placebo group: Edaravone injection placebo, 2 ampoules per dose, once daily, 60-minute intravenous infusion

Patients who meet all of the following criteria and are capable of giving consent are included in the study.

(1) Patients who are categorized as either "Definite ALS" or "Probable ALS" in the El Escorial revised Airlie House diagnostic criteria

(2) Patients at Grade 1 or 2 in the Japan ALS severity classification

(3) Patients scoring ≥ 2 points on each single ALSFRS-R item ("4. Handwriting" and "5. Eating motion (1)" should be scored ≥ 2 points on each side.)

- (4) Patients with normal respiratory function (%FVC is $\geq 80\%$; to be assessed using the actual values.)
 - (5) Patients with ALS that occurred within 2 years at the time of providing written informed consent
 - (6) Patients aged 20 to 75 years at the time of providing written informed consent
 - (7) Patients who provided written informed consent to participate in the present study
- [At registration]
- Patients who meet the following criterion in addition to the inclusion criteria for pre-registration
- (8) Patients in whom change in ALSFRS-R score during the 12-week pre-observation are -1 to -4 points.

After the end of Cycle 6, patients who were willing to continue the study medication were to receive additional active treatment for 24 weeks. Patients who completed Cycle 6 were asked whether they wished to move on to Cycle 7 (into the active treatment period). Those who did wish to do so continued in the study to Cycle 12.

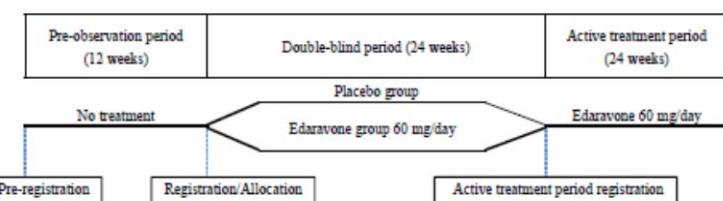
Cycle 7-12 (active treatment period)

Patients who were willing to continue the study medication: Edaravone injection 30 mg, 2 ampoules per dose, once daily, 60-minute intravenous infusion.

The study design is summarized in

Figure 6.

Figure 6 Study 19: Study Design



Pre-observation period: A 12-week observation period preceded Cycle 1 baseline.

Cycle 1: The investigational product was administered for 14 consecutive days. A 2-week drug-free period followed the end of treatment.

Cycles 2 to 12: The investigational product was administered on a total of 10 days out of the 2-week period. A 2-week drug-free period followed the end of treatment in each cycle.

Methods of Allocation

Patients who have been found to be eligible (according to the inclusion and exclusion criteria) were to be assigned to 2 treatment groups (edaravone group, placebo group).

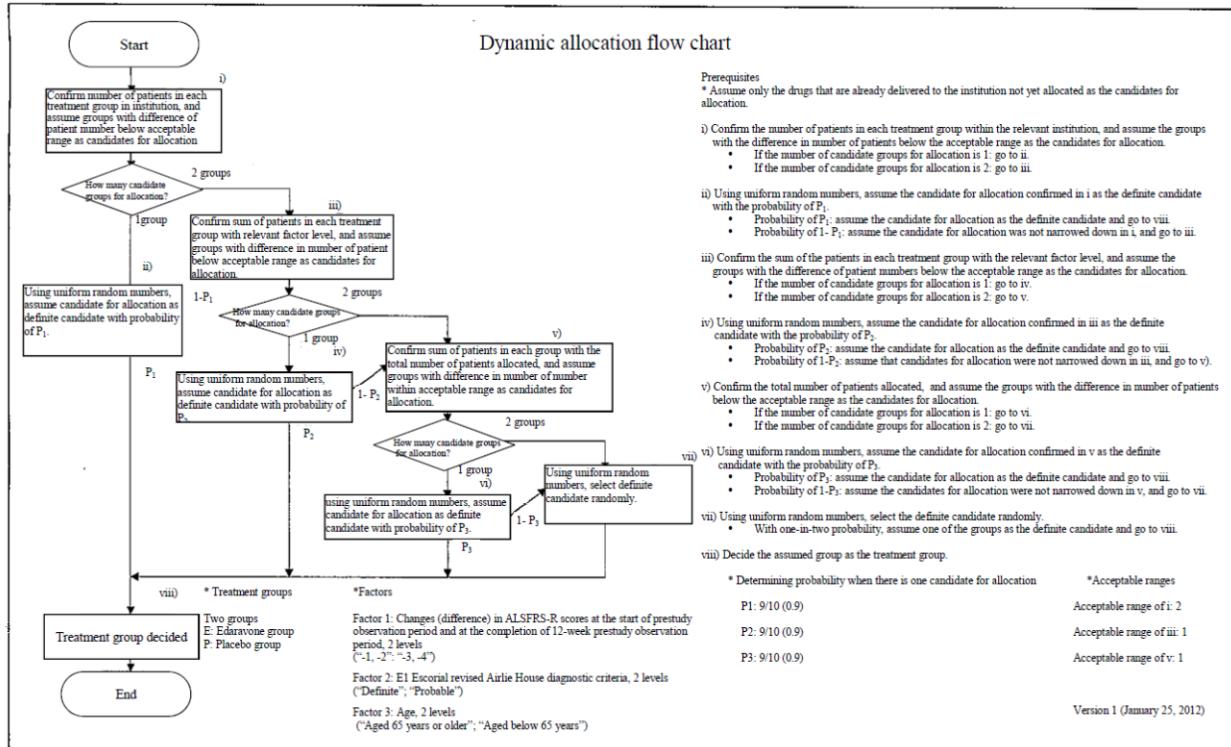
For this procedure, dynamic allocation by minimization method (Figure 7) was to be used, using the following factors that are considered to affect drug evaluation: change (difference) between the

Appears this way on original

ALSFRS-R score at the start of the pre-observation period and the ALSFRS-R score at the end of the pre-observation period (12 weeks later) (-1, -2/-3, -4); the El Escorial revised Airlie House diagnostic criteria (Definite/Probable); and age (≥ 65 years and <65 years).

Assignment was to be performed by the edaravone ALS registration center.

Figure 7 Study 19 Dynamic Allocation Flow Chart



Note: copied from page 76 of randomizations scheme document.

Efficacy Endpoints

(1) Primary: ALSFRS-R score

(2) Secondary

Time to death or certain disease progression (death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, loss of useful speech), %FVC, Modified Norris Scale score, Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ40) score, grip strength, pinch grip strength, ALS severity classification.

Planned Number of Patients

A total of 128 patients (64 patients per group) were planned as the primary analysis set for the primary efficacy endpoints.

Analysis of the subgroup results of the previous study suggested that the difference (adjusted value) in ALSFRS-R between the placebo group and the edaravone group may be 3.0, and the standard deviation 6. The sample size calculated based on the significance level of 5% (two-sided) and the power of 80% was 128 patients in total.

Thus, the decision was made to use the number of patients of 128 (64 per group) as the planned number of patients of the primary analysis set for the primary endpoint.

3.2.2.2. Statistical Methodologies

1) Primary endpoint

With respect to ALSFRS-R score, which is the primary endpoint, summary statistics by group and time points (mean, standard deviation, minimum, median, maximum) and summary statistics were to be calculated for the differences from “at baseline in Cycle 1.”

(a) Primary analysis

The difference between “at baseline in Cycle 1” and “the end of Cycle 6 or at discontinuation” was to be analyzed by using the factors in the dynamic allocation as covariates to perform group comparisons. For patients whose data “at the end of Cycle 6” are missing, data was to be complemented by the last observation carried forward (LOCF). The 95% C.I. of the difference between the groups was to be calculated as reference for the interpretation of results.

2) Secondary endpoints

- Time to death or certain disease progression (death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, loss of useful speech)
- %FVC
- Modified Norris Scale score
- Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ40) score
- grip strength
- pinch grip strength
- ALS severity classification

In order to investigate time to death or certain disease progression, death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, and loss of useful speech were defined as events, and other discontinuations were defined as censored cases for the purposes of survival analysis. %FVC, Modified Norris scale score, ALSAQ40 score, grip strength, pinch grip strength were to be analyzed in the same method used in the ALSFRS-R score. With respect to the ALS severity classification, a shift table of the changes from “at baseline in Cycle 1” to “the end of Cycle 6” was to be created. The end of Cycle 12 or at discontinuation” was to be shown by group.

As stated in the final analysis plan dated June 6, 2014, no multiplicity adjustment was to be performed because all of secondary analyses of the primary endpoint and analyses of the secondary endpoints were performed as exploratory analysis.

Secondary analyses

With respect to the scores at each time point, a repeated measurements analysis of variance was to be performed by using treatment groups, time points, treatment group by time point interaction as factors and “at baseline in Cycle 1,” and factors used for dynamic allocation as covariates, for comparison difference between the groups.

With respect to the difference between “at baseline in Cycle 1” and each time point, the mean was to be calculated for each patient and used as a summary measure, and analysis of

covariance was to be performed by using the factors used in dynamic allocation as covariates to compare between the groups.

- After simple regression analysis is performed for each patient, the slope of time-dependent change was to be calculated and used as a summary measure, and an analysis of covariance was to be performed by using the factors used in the dynamic allocation as covariates to compare between the groups.
- Analysis using statistical modeling taking into consideration changes over time, such as analysis by a mixed effect model, was to be performed.
- A survival analysis based on events defined by ALSFRS-R score was to be performed.

In order to investigate time to death or certain disease progression, any of the following conditions will be defined as events, and the other cases will be defined as censored cases to show a Kaplan-Meier curve. With respect to the patients in which multiple events occurred, the date when the first event occurs will be used as the date of the onset of the main event. The day of censoring will be the last observation date for each patient. For the Full Analysis Set (denoted FAS as defined below) during the double-blind period,* events occurring from Cycle 1 through Cycle

12 will be analyzed (count of the number of events and a Kaplan-Meier curve).

- Deaths
- Disability of independent ambulation
- Loss of upper limbs function
- Tracheotomy
- Use of respirator
- Use of tube feeding
- Loss of useful speech

Analysis Sets

The analysis sets are defined as follows. In this study, data was to be locked twice, i.e., at the end of the double-blind period and at the end of the active treatment period. The details of handling of patients in the double-blind period was to be determined by the sponsor by the time of code breaking for the double-blind period, and the details of handling of patients in the active treatment period was to be determined by the sponsor before data lock for the active treatment period.

Efficacy was to be analyzed in the full analysis set (FAS). The FAS is an analysis set consisting of all patients completing the investigational product of the double blind period except the following patients:

- Patients who have been found to have no ALS(Reviewers Comment: as it turned out there were none of these)
- Patients who have never administrated the investigational product
- Patients with no efficacy data after treatment with the investigational products

Interim analysis

A prior consultation with the Pharmaceuticals and Medical Devices Agency of Japan was carried out on June 19, 2014, and it was determined that the application would be submitted with the data for the double-blind period and that the data for the active treatment period would be submitted as additional data. Consequently, an interim analysis was performed using the data for the double-blind period. Prior to the analysis, the case handling in the double-blind period was determined before the data-base for this period was locked, and a separate statistical analysis plan (double-blind period) was prepared and documented.

Therefore, an interim analysis was performed with the data from the double-blind period, and the clinical study report version 1 was prepared using data obtained by the database lock date (June 10, 2014) of the double-blind period.

Thereafter, data for the entire period, including the active treatment period, were locked (November 10, 2014), and the clinical study report version 2 was prepared with additional information from the active treatment period.

Reviewer's Comment: This is only an interim analysis in the sense that the active phase was ongoing when the placebo controlled phase was analyzed; it is not an interim analysis in the usual sense that the primary hypothesis data had only been collected on a portion of patients at the time of the interim analysis.

The Data and Safety Monitoring Board (DSMB) convened when the number of patients who completed or discontinued investigational product treatment exceeded 25%, 50%, and 75% of the target number of patients, or if new serious adverse events occurred in 10 patients after the start of administration or after the DSMB was held.

3.2.2.3. Patient Disposition, Demographic and Baseline Characteristics

The sponsor found no imbalance in the inclusion or exclusion of patients and the presence or absence of discontinuation of treatment in all of the analysis sets.

Table 5 Study 19 Patient Disposition through Cycle 6 (Double blind placebo controlled period)

Number of patients who completed the double-blind period		Number of patients who discontinued from the study	
		Total	10
Total	127		
Placebo	60		
Edaravone	67		
		Placebo	8
		Edaravone	2

Note: This table copied from page 84 of the sponsor's study report

The homogeneity of the groups was examined by the sponsor with a significance level of 15%. The results showed no imbalances between the groups for any of the demographic or baseline variables in the FAS.

Table 6 Study 19 Baseline Demographics and Disease Characteristics

Baseline or Demographic Variable	Statistic or Category as applicable	Placebo	Edaravone	Overall
Age	Mean (SD)	60.1 (9.6)	60.5 (10.1)	60.3 (9.9)
Age Group	>=65	22 (32.4)	23 (33.3)	45 (32.8)
Age Group	<65	46 (67.6)	46 (66.7)	92 (67.2)
Diagnosis	Definite	27 (39.7)	28 (40.6)	55 (40.1)
Diagnosis	Probable	41 (60.3)	41 (59.4)	82 (59.9)
Heredity	Familial	2 (2.9)	1 (1.4)	3 (2.2)
Heredity	Sporadic	66 (97.1)	68 (98.6)	134 (97.8)
Pre ALSFRS-R	Mean (SD)	43.5 (2.2)	43.6 (2.2)	43.6 (2.2)
Baseline ALSFRS-R	Mean (SD)	41.8 (2.2)	41.9 (2.4)	41.8 (2.3)
Baseline FVC	Mean (SD)	97.4 (13.6)	100.2 (15.1)	98.8 (14.4)
Disease duration	Mean (SD)	1.1 (0.5)	1.1 (0.5)	1.1 (0.5)
Japan ALS category	1	16 (23.5)	22 (31.9)	38 (27.7)
Japan ALS category	2	52 (76.5)	47 (68.1)	99 (72.3)
Height	Mean (SD)	162.5 (8.4)	161.8 (9.5)	162.1 (9.0)
Sex	Female	27 (39.7)	31 (44.9)	58 (42.3)
Sex	Male	41 (60.3)	38 (55.1)	79 (57.7)
Weight	Mean (SD)	57.8 (9.3)	57.9 (12.9)	57.8 (11.2)
Riluzole	No	6 (8.8)	6 (8.7)	12 (8.8)
Riluzole	Yes	62 (91.2)	63 (91.3)	125 (91.2)
Initial Symptom	Bulbar	14(20.6)	16(23.2)	30(21.9)
Initial Symptom	Limbic	54(79.1)	53(76.8)	107(78.1)

3.2.2.4. Results and Conclusions

3.2.2.4.1. Sponsor's Results

Primary endpoint: ALSFRS-R score

The change (mean \pm SD) from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was -4.4 ± 3.8 in the edaravone group and -6.8 ± 4.9 in the placebo group. For reference purposes, changes from baseline in Cycle 1 to each time point was compared between groups by time point using the two-sample *t*-test. In addition, a paired *t*-test was performed on the changes from "baseline in Cycle 1" for each group by time point.

Changes from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)," were compared between groups with the following 3 covariates, which were the factors used in dynamic allocation: "change in ALSFRS-R score from the beginning to the end of the pre-observation period (12 weeks after pre-registration)"; "El Escorial revised Airlie House diagnostic criteria"; and "age". Results (FAS) are shown in Table 7. The least square mean (LSMean) \pm standard error (SE) for each treatment group was -5.01 ± 0.64 for the edaravone group and -7.50 ± 0.66 for the placebo group. Thus, the LSMean \pm SE of the difference between the groups (edaravone group – placebo group, the same applies hereinafter) and the 95% confidence interval of this mean was 2.49 ± 0.76 (0.99 to 3.98), and the difference between the groups was significant ($P=0.0013$).

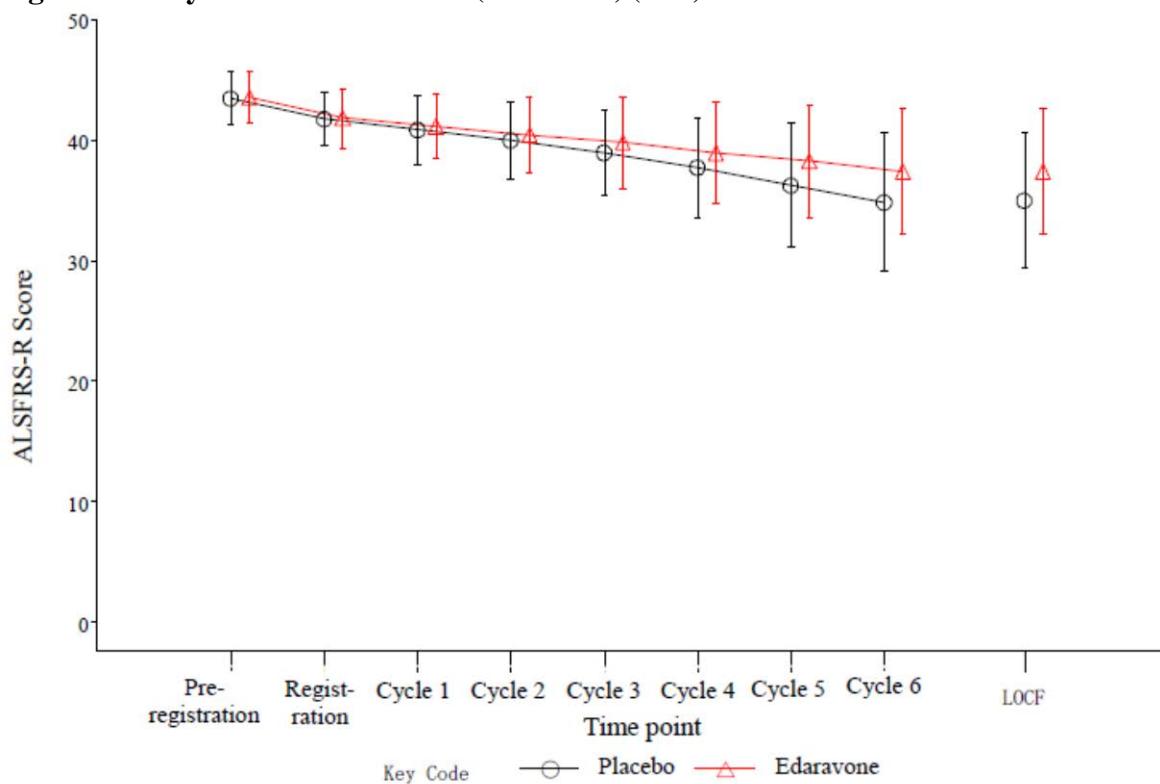
Table 7 Analysis of change in ALSFRS-R score from baseline in Cycle 1 to the end of Cycle6 (LOCF) (FAS)

Treatment group	Adjusted mean	Between-group difference in adjusted mean		
	LSMean \pm SE (95% CI)		t-value	P-value
Placebo	-7.50 ± 0.66	2.49 ± 0.76 (0.99, 3.98)	$t=3.29$	$P=0.0013$
Edaravone	-5.01 ± 0.64			

Note: This table was copied from page 107 of the Sponsor's study report for study 19

Figure 8 shows the observed mean ALSFRS-R scores over time in the double-blind placebo controlled period.

Figure 8 Study 19: ALSFRS-R score (mean ± SD) (FAS)



Note: This figure was copied from page 105 of the sponsor's study report

To examine the time to death or certain disease progression, the number of events (FAS) of each of the following was tabulated by treatment group (Table 8) death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, and loss of useful speech.

Table 8 Number of events involving death or certain disease progression (FAS) up to 6 months

Event	Placebo	Edaravone	Total
Death	0	0	0
Disability of independent ambulation	2	0	2
Loss of upper limbs function	0	0	0
Tracheotomy	0	1	1
Use of respirator	0	0	0
Use of tube feeding	1	0	1
Loss of useful speech	3	1	4

Population analyzed: FAS in the double-blind period. Events that occurred during Cycle 1 to Cycle 6 were analyzed.

Note: This table was copied from page 119 of the sponsor's study report

A Kaplan-Meier curve was constructed, with any of "death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, and loss of useful speech" defined as an event, and other cases defined as censored data. A log-rank test and generalized Wilcoxon test were performed. In this case, the censoring date was the day when the last observation was performed. For patients who completed the double-blind period, this was the end of Cycle 6. For patients who discontinued treatment, it was 2 weeks after the last dose. For patients with multiple events, the day of onset of the first event was considered the event onset day. The number of events was 2 in the edaravone group and 6 in the placebo group. Although there were fewer events in the edaravone group as compared with the placebo group, the difference was not statistically significant ($P=0.1284$ [log-rank test], $P=0.1415$ [generalized Wilcoxon test]).

(1) The change in %FVC from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was analyzed with the 3 factors used in dynamic allocation as covariates. The groups were compared, and the results (FAS) are shown in Table 9.

Table 9 Analysis of change in %FVC from baseline in Cycle 1 to the end of Cycle 6(LOCF) (FAS)

Treatment group	Adjusted mean LSMean ± SE	Between-group difference in adjusted mean		
		LSMean ± SE (95% CI)	t-value	P-Value
Placebo	-20.40±2.48	4.78±2.84	$t=1.69$	$P=0.0942$
Edaravone	-15.61±2.41	(-0.83, 10.40)		

Note: This table was copied from page 132 of the sponsor's study report

The modified Norris Scale score was examined with the Limb Norris Scale score, Norris Bulbar Scale score, and the sum of their scores. The change in the Limb Norris Scale score from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was analyzed with the 3 factors used in dynamic allocation as covariates.

The groups were compared, and the results (FAS) are shown in Table 10.

Table 10 Analysis of change in Limb Norris Scale score from baseline in Cycle 1 to the end of Cycle 6 (LOCF) (FAS)

Treatment group	Adjusted mean LSMean ± SE	Between-group difference in adjusted mean		
		LSMean ± SE (95% CI)	t-value	P-Value
Placebo	-14.91±1.68	3.44±1.92	$t=1.79$	$P=0.0757$
Edaravone	-11.47±1.61	(-0.36, 7.24)		

Note: This table was copied from page 135 of the sponsor's study report

The change in the modified Norris Scale score (bulbar component) from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was not nominally significant with an estimated treatment difference of 1.46 ± 0.90 , $p=0.1092$.

The change in the modified Norris Scale score (total) from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was analyzed with the 3 factors used in dynamic allocation as covariates. The groups were compared, and the results (FAS) are shown in Table 11.

Table 11 Analysis of change in modified Norris Scale score (total) from baseline in Cycle 1 to the end of Cycle 6 (LOCF) (FAS)

Treatment group	Adjusted mean	Between-group difference in adjusted mean		
	LSMean \pm SE (95% CI)		t-value	P-Value
Placebo	-20.80 \pm 2.06	4.89 \pm 2.35	$t=2.08$	$P=0.0393$
Edaravone	-15.91 \pm 1.97	(0.24, 9.54)		

The change in the ALSAQ40 score from "baseline in Cycle 1" to "the end of Cycle 6 (or discontinuation, LOCF)" was analyzed with the 3 factors used in dynamic allocation as covariates. The groups were compared, and the results (FAS) are shown in Table 12.4.1.1-33. The LSMean \pm SE for each treatment group was 17.25 ± 3.39 in the edaravone group and 26.04 ± 3.53 in the placebo group. Thus, the LSMean \pm SE of the difference between the treatment groups and the 95% confidence interval of the difference was -8.79 ± 4.03 (-16.76 to -0.82), and the difference was significant ($P=0.0309$). Note that ALSAQ40 was only collected at baseline and the end of Cycle 6.

Table 12 Analysis of change in ALSAQ40 score from baseline in Cycle 1 to the end of Cycle6 (LOCF) (FAS)

Treatment group	Adjusted mean	Between-group difference in adjusted mean		
	LSMean \pm SE (95% CI)		t-value	P-Value
Placebo	26.04 \pm 3.53	-8.79 \pm 4.03	$t=-2.18$	$P=0.0309$
Edaravone	17.25 \pm 3.39	(-16.76, -0.82)		

Note: This table was copied from page 143 of the sponsor's study report

3.2.2.4.2. Reviewer's Results

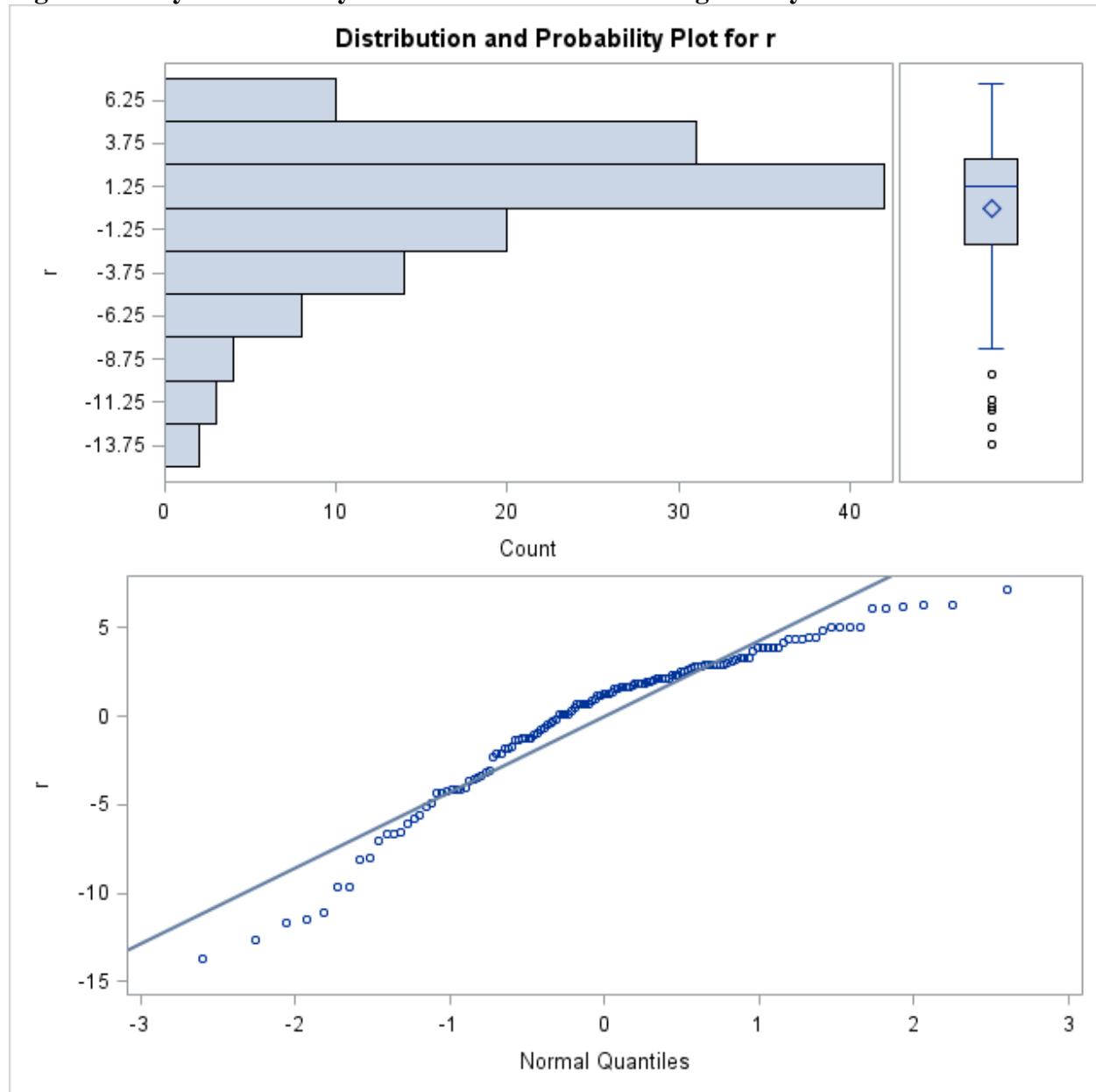
Dropouts in the first phase (placebo controlled phase) were reasonably low: 2/69 =3% for Edaravone and 8/68=12% for placebo. The treatment difference at the end of cycle 6 based on Mixed Model Repeated Measures (MMRM) analysis was $2.83 \pm .76$, $p=0.0003$.

There were four deaths (2 placebo and 2 drug) but none of these occurred in the first phase (up to Cycle 6). Based on a Wilcoxon test of the joint rank of combined function and survival the p-value is 0.0009. Note that the residuals from the sponsor's primary ANCOVA model showed

signs of significant non-normality (Figure 9). However, a Wilcoxon rank sum test of the Cycle 6 ALSFRS-R changes (or LOCF as applicable) had a p-value of 0.0022. The placebo and drug unadjusted means at Cycle 6 were -6.85 and -4.41 whereas the respective medians were -5.00 and -4.00. The Hodge Lehmann estimate of the median of all differences between treatment groups was -2.0.

Because the randomization was a non-standard Minimization randomization a re-randomization test was performed as a sensitivity analysis and it was found to support the primary analysis ($p=.0018$).

Figure 9 Study 19 Normality Check for ALSFRS-R Changes at Cycle 6



Note: r stands for residual error here

Some of those who died had Cycle 6 ALSFRS changes from baseline that were not bad compared to the rest of the randomized patients and even better than some survivors. One placebo patient had a change of -1 and two drug patients had changes of -6 and -10 respectively as compared to the minimum which was -20. Therefore, this highlights that it is important to use the joint rank analysis since it reflects the final outcome better by assigning the worse rank to deaths based on survival time.

The change from baseline in ALSFRS-R may depend on the baseline ALSFRS-R in a non-linear way (i.e., quadratic or cubic terms in the baseline score may be needed) as seen in Figure 10. A test for a non-linear cubic dependence on baseline value of ALSFRS-R of the same form for both treatment arms gives a p-value of 0.0486 suggesting that the effect of the baseline is significantly more complicated than linear in the baseline ALSFRS-R score, i.e., the treatment difference likely varies with the baseline score. However, the estimated treatment difference on ALSFRS-R change at the end of Cycle 6 based on this cubic baseline ALSFRS-R model (of the same form for both treatment groups) is $2.32 \pm .74$ S.E., $p=0.0022$ which is similar to the prespecified model.

We may consider exploring the approach of letting the baseline score be a class variable to account for the lack of linearity displayed in the pattern of LS Means as a function of baseline ALSFRS. Since we are not sure of the true function form of non-linearity it is convenient to let the effect of baseline vary with each baseline value by entering baseline into the model as a class variable (although it is acknowledged that if the number of parameters approaches the number of observations then the model estimates may not be consistent but this is an exploratory analysis anyway). When we compare the linear baseline model to the class baseline model we find that Akaike's information (AIC) is 766.2 for the model having baseline linear and 720.6 for the class formulation; note that smaller AIC suggests better model fit so that even after accounting for the penalty for its additional parameters the class formulation is better. Therefore, if we average the baseline specific treatment difference estimates from the model with baseline ALSFRS-R as a class (discrete valued) variable also incorporating an interaction between treatment group and baseline ALSFRS-R we get an average of 2.378 ± 1.085 S.E., $p=0.0306$ (LOCF analysis). Therefore, once again the estimated treatment difference is similar.

Average baseline ALSFRS-R score was 41.9 for both groups. Average change from pre-baseline to baseline in ALSFRS-R was -1.7. An additional Preval (screening ALSFRS-R value) covariate effect was nominally significant even after the difference between preval and bval (baseline ALSFRS-R) was included in the ANCOVA model for ALSFRS-R change at the end of cycle 6 $p=0.022$. This reviewer focused on analyses adjusted by the baseline value instead of by the prespecified difference between pre-baseline and baseline values since the former is the usual standard in clinical trials in Neurology. A model including both terms suggests that the former is more significant (having a p-value of 0.02 while the pre-baseline difference has a p-value of 0.25). This suggests if we were to pick just one of these for inclusion in the model the baseline value would be the better choice. Note that the sponsor did not treat the change from screening to baseline as a continuous variable in the analysis model but instead grouped the changes into the

classes -4 to -3 and -2 to -1 that were also used in the randomization. However, adding the baseline ALSFRS-R as an additional covariate beyond the stratification factors has essentially no effect on the analysis result (estimate or p-value, e.g., the p-value decreases by .0002).

Analysis of Observed cases (MMRM) showed sizable difference in effect size around subgroups defined by being above or below the median baseline ALSFRS value of 42. AIC 762.5 with baseline linear and no interaction 762.0 with baseline linear plus interaction; 717.5 for model with categorical baseline and no interaction; 666.4 for model with categorical baseline plus interaction

Forty percent had baseline ALSFRS-R scores less than or equal to 41 and 59% had baseline ALSFRS-R scores less than or equal to 42. The treatment difference seems to decrease in the lower half of the baseline scores as shown in Table 13.

The Sponsor's premise from the study 16 results was that efficacy is expected in less severe patients. This raises the question, could it be ineffective for some severity? This would be important to know for the long term management of individual patients.

The Median baseline ALSFRS-R is 42 but that is actually the 59th percentile also, because the distribution is discrete. Therefore, because there is considerable uncertainty in the median baseline ALSFRS-R I explore efficacy by several baseline severity subgroups.

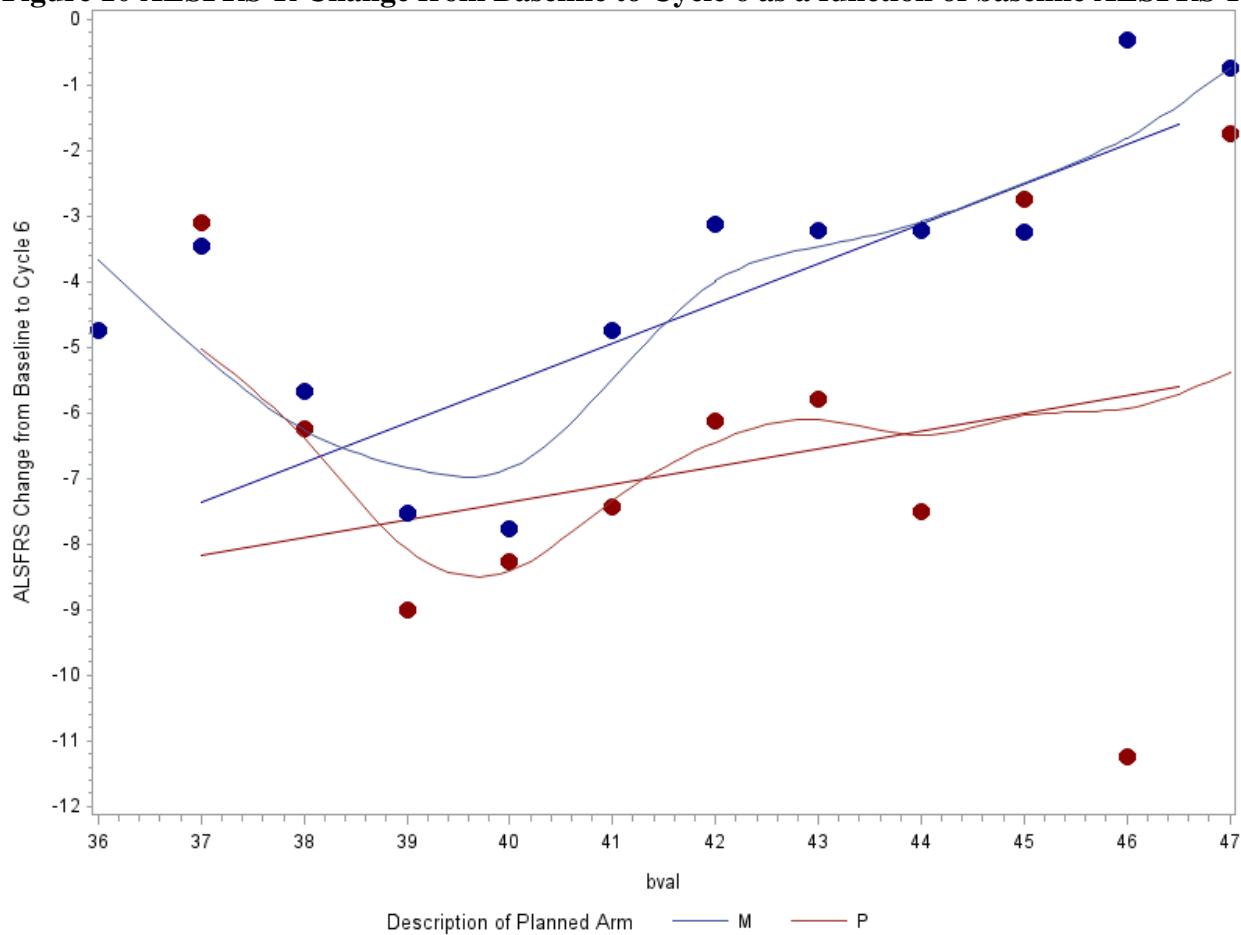
Table 13 ALSFRS Treatment Group Difference at the end of Cycle 6 by baseline ALSFRS subgroups (MMRM)

Baseline ALSFRS-R Cutoff	Baseline ALSFRS-R Subgroup	Est. Treat Diff	StdErr	p-value
th 42=59 percentile	<=42 d6 DB	2.25	1.02	0.0289
	>42 d6 DB	2.99	1.22	0.0159
st 41=41 percentile	<=41 d6 DB	1.96	1.16	0.0943
	>41 d6 DB	3.32	0.96	0.0007
nd 40=32 percentile	<=40 d6 DB	1.67	1.32	0.2053
	>40 d6 DB	3.36	0.90	0.0003

An exploratory likelihood ratio test of whether the linear baseline ALSFRS-R term (slope of the change as a function of baseline score) depends on treatment group gives a p-value <0.0001 (-2*log likelihood = 2364.4) and in fact the relationship between change from baseline in ALSFRS and baseline score may be more complicated than linear. In particular, if we perform a likelihood ratio test comparing a cubic model with quadratic and cubic terms for the baseline ALSFRS

(common to both treatment groups) to the standard model with only a linear term the cubic terms are significant, $-2\log$ likelihoods are 2384 and 2318, $p<0.0001$. A likelihood test of whether group specific quadratic and cubic terms are significant was also significant, $-2\log$ likelihood= 2282, $p<0.0001$. However, we must acknowledge that the prespecified primary analysis with the usual linear term for baseline score was statistically significant and the exploratory polynomial models seem to support efficacy over most of the observed range of the baseline ALSFRS-R scores. Figure 10 shows the exploratory polynomial model predicted change from baseline in ALSFRS-R by group at each baseline ALSFRS-R score.

Figure 10 ALSFRS-R Change from Baseline to Cycle 6 as a function of baseline ALSFRS-R



Note there were 6 M (Edaravone) and 5 P (placebo) patients with $bval=45$; 3 M and 2 P with $bval=46$; and 1 and 1 with $bval=47$

In summary, although the treatment effect may not be exactly consistent across baseline ALSFRS-R, the various sensitivity analyses support the significance of the sponsor's primary analysis of ALSFRS-R change overall at the end of Cycle 6 in Study 19.

Secondary Endpoints

The secondary analyses of the limbic and bulbar components of the modified Norris scale were not significant at the end of cycle 6 as reported by the sponsor. The mean total score for those who discontinued in the treatment group was worse than for the completers. The reported

analysis of the modified Norris total score at the end of Cycle 6 was not nominally significant when data from discontinuation visits was included ($p=0.0520$).

Analysis of Time to Disease Progression

An exact log rank test of the sponsor's preferred and prespecified progression events (Table 14) occurring up until the end of Cycle 12 (active extension period) had a p-value of 0.0217 (exact Logrank test).

Loss of useful speech was added as a new type of progression event in this study, while the other types of progression events in this study were the same as in the Riluzole program as noted by the sponsor. Excluding loss of useful speech from progression events makes time to event at end of Cycle 12 lose nominal significance, $p=0.0710$ for Log-rank and 0.1176 for Wilcoxon. Also, note that this analysis restricted to events in the first 6 cycles (double blind placebo controlled period) was not significant in this study ($p=0.128$) or in study 16 ($P = 0.3814$ in a slightly more severe population in which more events might be expected and note that loss of useful speech was not counted as a progression event in study 16 or 17 either). There was also no multiplicity adjustment plan for secondary endpoints proposed in the analysis plan or protocol.

Table 14Total number of events involving death or certain disease progression through Cycle 12(FAS)

Treatment group	No. of patients	No. of events*
Placebo-Edaravone	68	19
Edaravone-Edaravone	69	10

*Death, disability of independent ambulation, loss of upper limbs function, tracheotomy, use of respirator, use of tube feeding, and loss of useful speech

Note: This table was copied from page 159 of the sponsor's study report

Active Extension Phase

The proportion of dropouts was substantial for the active extension phase and somewhat different between treatment groups: 24% for Edaravone and 45% for placebo. The sponsor reported a nominally significant difference at the end of the active extension (Cycle 12) based on LOCF analysis. However, this assumes that there would be no change from the last assessment after dropout which is unlikely in a degenerative disease and it may treat deaths as a good outcome. Therefore, this reviewer performed a joint rank analysis of ALSFRS-R score and survival time in the case of death.

Joint rank analysis of survival and ALSFRS change based on a Wilcoxon test and using all events ranking from worst (death) to least bad(loss of useful speech) followed by ranking no events with non-missing ALSFRS-R score gives $p=0.0520$ for the difference at the end of Cycle 12.

In addition, an exploratory logistic regression analysis of the event of dropping out before the end of the open label extension (cycle 12) found nominally significant associations with ALSFRS-R pre-value at screening, Disease duration, baseline Forced Vital Capacity, weight at baseline, definite vs. probable ALS, and Japanese ALS classification. This suggests that the missing at random assumption used with the MMRM analysis may be unlikely to hold in this case and certainly it means that the missing completely at random assumption used with the

LOCF analysis at the end of Cycle 12 is highly suspect as illustrated by the p-value from the joint rank analysis which exceeds 0.05. Therefore, the reported maintenance of effect on the ALSFRS-R change through the end of Cycle 12 (end of the active treatment period) may not be robust.

3.3. Evaluation of Safety

Safety is not evaluated in this review. Please see the medical officer's review for the evaluation of safety of the product.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Geographic Region, Gender, Race and Age

The Study was done in Japan only so no investigation of race differences is possible with the submitted efficacy data.

Of the randomized population in study 19 40% of Placebo and 45% of Edavarone were Female and 32% and 33% respectively were ≥ 65 years of age (the average age was 60 and S.D.=10). Study 19 Gender specific treatment differences in ALSFRS-R at end of cycle 6 appear consistent as seen in Table 15 (a test for interaction between age group and treatment has a p-value of 0.95). The same is true for age group specific differences as seen in Table 16 (a test for interaction between age group and treatment has a p-value of 0.83).

Table 15 Study 19: Estimated Treatment Differences on ALSFRS-R at Cycle 6 by Gender

Gender	Estimated Trt Diff	Std Error	p-value
Female	-2.7572	1.1804	0.0210
Male	-2.8516	1.0097	0.0055

Table 16 Study 19: Estimated Treatment Differences on ALSFRS-R at Cycle 6 by Age Group

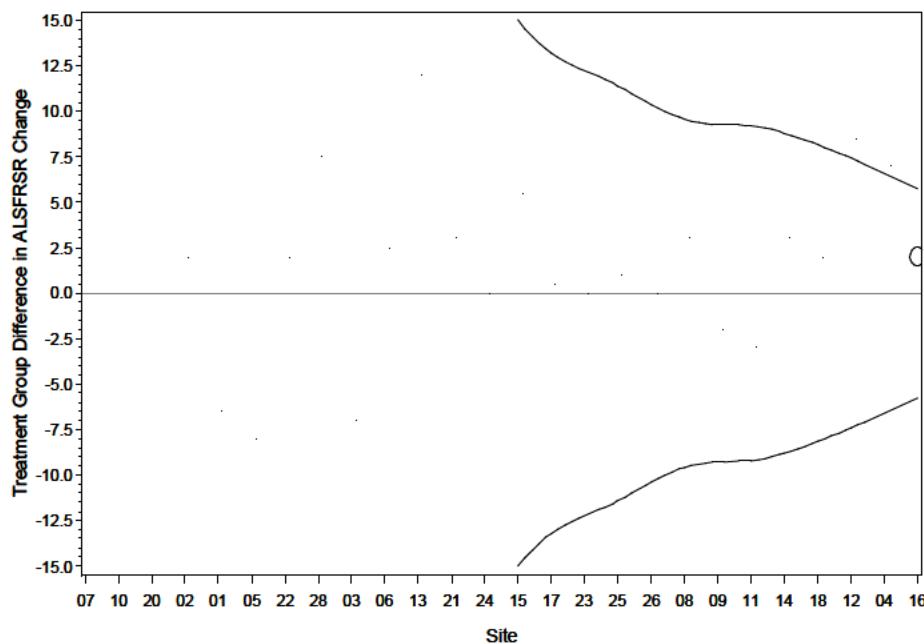
Age Group	Estimated Trt Diff	Std Error	p-value
≥ 65	-3.0004	1.3397	0.0268
<65	-2.7436	0.9355	0.0040

4.2 Geographic Region

Study 19 was done completely in Japan as were all of the efficacy studies in this application. There were 26 study sites in study 19 and the number of patients per site ranged from 1 to 12.

Figure 11 shows estimated treatment differences in ALSFRS-R change at the end of Cycle 6 (24 weeks) within individual sites in study 19. In the figure, the sites are ordered by size increasing from left to right, the size of the plotting symbol is proportional to the number of patients randomized at the site and the curves on the graph delineate the positive and negative 2*standard error limits as a function of sample size. The maximum sample size of subjects with post-baseline efficacy data for any site in study 19 is 12. Estimated treatment differences within each of Sites 04, 06, 08, 12, and 18 reached nominal significance on ALSFRS-R change at the end of Cycle 6. Site 11 was nominally significant in favor of placebo. Fifteen out of 23 sites with at least one patient in each group numerically favored the drug arm in terms of the ALSFRS-R change from baseline at the end of Cycle 6 (5 favored placebo and 3 had no treatment difference and 3 had zero patients in one group).

Figure 11 Study 19 Mean Treatment Difference in Change from Baseline in ALSFRSR by Site at Cycle 6



For study 16/17 the site with the biggest impact on primary efficacy (ALSFRS-R change) was site 3 (N=7) [2301465 National Hospital Organization Miyagi]. The maximum site size of post-baseline efficacy subjects among all sites in study 16/17 was N=14 [at site #13]. Removing site 13 to explore it's impact on overall efficacy we find that the overall p-value decreases to 0.1182 (-1.32 +/- 0.84), so this site may have been somewhat influential overall in a negative way. One drug patient in the post-hoc subgroup within this site had a decrease of 9 points from baseline to cycle 6. The baseline difference reached nominal significance in this site (Edaravone 39.1 range 32-42; Placebo 42.6 range 40-45) as well. Similarly, there may be a placebo outlier in site 03 as evidenced by a comparison of the means in site 03 vs. overall (n=3 -16.3 vs. n=99 -5.8 overall). The end of cycle 6 ALSFRS-R p-value increases to 0.0497 when site 03 is excluded (estimated treatment difference is 2.31) up from 0.0100 with all sites (estimated treatment difference is 3.41) in the efficacy subgroup Definite/Probable/EESP/2y.

4.3 Other Special/Subgroup Populations

The Japanese ALS severity scale ranges from 1 through 5 with 1 ("able to work or perform housework") being the least severe.

In study 19, Sixteen (23.5%) placebo patients and 22 (31.9%) Edaravone patients had Japanese ALS severity classification of 1 at baseline in study 19 and the remaining patients were classified as Japanese ALS severity 2 (“Independent living but unable to work”).

In study 19, the treatment difference on ALSFRS-R change from baseline at the end of Cycle 6 was numerically bigger in the category 1 group but the difference between categories (interaction between treatment and Japanese ALS severity rating) was not significant

(Category1 -2.9 +/- 1.4 S.E.
Category 2: -2.5 +/- 0.9 S.E.).

APPEARS THIS WAY ON ORIGINAL

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

There were 2 different analyses designated as primary for study 16 in the final analysis plan and the sponsor stated that they would consider the study positive if either was significant at 0.05 two-sided. This creates a multiplicity problem. The type I error would be inflated with this plan because there are two chances to win at 0.05, two-sided. However, either way the study failed in the overall population and there were imbalances found in the post-hoc subgroup that might explain the apparent significance of the post-hoc subgroup result as possibly due to something other than treatment assignment alone.

This reviewer noted significant non-normality of the primary efficacy measure, change from baseline in ALSFRS-R and also a trend for variances increasing with time in both studies which might be relevant for planning the analysis in future studies in ALS.

The study 19 p-value for the ALSFRS-R change at the end of Cycle 6 was significant. However, it requires circular logic to consider the second study to validate the first (study 16) so that the submission might approach the efficacy standard of two positive studies. It is not the same amount of evidence as two positive studies, positive on their originally prespecified primary hypotheses. This reviewer has observed many cases where attempts to validate a post-hoc subgroup finding in a follow-up study designed based on that subgroup failed but “validation” would tend to happen by chance alone at the rate of 1 in 20. Note, also that the studies in this submission were done exclusively in Japanese subjects, so there is no data in American subjects.

5.2 Collective Evidence

Collective evidence analysis is of questionable reliability in this case since the first study, study 16, failed to meet its primary objective and there appeared to be some baseline imbalances within the subpopulation of interest in that study. However, the effect sizes in terms of change from baseline in ALSFRS-R at the end of Cycle 6 (6 months) were fairly consistent between similar populations (note: ITT of study 19 was similar to post-hoc subgroup of study 16: def/prob/EESP/2y).

Table 17 Summary of Efficacy Results from Phase III studies in ALS

Study/Subgroup	Treatment Group	Sample Size	Baseline Mean	LS Mean Change in ALSFRS	LS Mean Diff	p-value
16 ITT	Placebo	99	41.2	-6.4	0.7 +/- 0.8	0.4108
	Edaravone	100	40.6	-5.7		
16 complement of def/prob/EESP/2y	Placebo	70	40.7	-5.6	-0.5 +/- 1.0	0.6487
	Edaravone	62	39.3	-6.0		
16 EESP	Placebo	46	42.5	-7.1	2.2 +/- 1.0	0.036
	Edaravone	53	42.4	-4.9		
16 definite or probable/EESP/2y	Placebo	29	42.1	-7.6	3.0 +/- 1.3	0.027
	Edaravone	39	42.5	-4.6		
19 ITT	Placebo	68	41.8	-7.5	2.5 +/- 0.8	0.0013
	Edaravone	69	41.9	-5.0		

Most of the secondary endpoints in study 19 that were not based on ALSFRS-R did not reach the level of nominal significance at the end of the placebo controlled double blind period (6 months). The sample size of study 19 was moderately small and some secondary endpoints had p-values >0.05 but < 0.10 so the power may have been suboptimal for the secondaries but regardless the secondaries were to be considered exploratory. ALSAQ40 was listed as nominally significant

with $p=0.03$ as was the Modified Norris Scale Total (Limbic + Bulbar) but time to disease progression ($p=0.128$) and change in Forced Vital Capacity ($p=0.094$), were listed before it in the list of secondary endpoints and were not significant (although no hierarchy or multiplicity adjustment plan was technically specified). Nor were change in grip strength or change in pinch grip strength nominally significant. However, there was no explicit multiplicity adjustment plan for secondary endpoints as they were considered exploratory.

Table 18 shows the sponsor's subgroup analyses of ALSFRS-R change for study 19 and 16 in the "efficacy" subgroup.

Table 18 Sponsor's Analyses in Subgroups

Category	Sub-group	Treatment group	Number of subjects	Study MCI186-19 FAS		Study MCI186-16 Definite or Probable/EESP/2y		
				Between-group differences in the adjusted mean ^a		Number of subjects	Between-group differences in the adjusted mean ^b	
				LS mean±SE	95% CI		LS mean±SE	95% CI
Sex	Male	P group	39	2.54±1.06	(0.42, 4.66)	18	0.42±1.56	(-2.75, 3.58)
		E group	38			26		
	Female	P group	27	2.36±1.11	(0.14, 4.58)	11	5.97±2.10	(1.57, 10.36)
		E group	30			13		
Age (years)	< 65	P group	44	2.31±1.00	(0.33, 4.30)	22	2.24±1.30	(-0.36, 4.85)
		E group	46			33		
	≥ 65	P group	22	2.73±1.13	(0.46, 5.01)	7	6.72±5.27	(-5.44, 18.88)
		E group	22			6		
Body Weight ^c (kg)	< Median	P group	31	3.21±1.15	(0.90, 5.51)	14	3.98±1.99	(-0.12, 8.07)
		E group	34			17		
	≥ Median	P group	35	2.05±1.07	(-0.08, 4.18)	15	1.58±1.69	(-1.85, 5.02)
		E group	34			22		
BMI ^c (kg/m ²)	< Median	P group	32	2.97±1.15	(0.66, 5.28)	15	2.09±2.07	(-2.14, 6.32)
		E group	34			18		
	≥ Median	P group	34	2.28±1.05	(0.18, 4.38)	14	3.74±1.77	(0.13, 7.35)
		E group	34			21		
Duration of illness (years)	<1 year	P group	32	2.56±1.17	(0.22, 4.90)	15	4.86±2.41	(-0.12, 9.84)
		E group	27			13		
	≥1 year	P group	34	2.22±1.03	(0.17, 4.28)	14	1.30±1.29	(-1.32, 3.92)
		E group	41			26		
ALS Diagnosis	Sporadic	P group	64	2.41±0.76	(0.90, 3.92)	29	2.98±1.35	(0.28, 5.68)
		E group	67			38		
	Familial	P group	2	-	-	0	-	-
		E group	1			1		
Initial Symptom	Bulbar symptoms	P group	14	2.42±1.46	(-0.60, 5.43)	7	5.65±4.33	(-4.33, 15.63)
		E group	15			4		
	Limb symptoms	P group	52	2.44±0.89	(0.68, 4.21)	22	2.51±1.35	(-0.20, 5.23)
		E group	53			35		
El Escorial revised Airlie House Diagnostic Criteria	Definite ALS	P group	26	2.13±1.19	(-0.25, 4.51)	9	3.69±2.42	(-1.35, 8.73)
		E group	28			17		
	Probable ALS	P group	40	2.85±0.99	(0.88, 4.82)	20	2.08±1.58	(-1.12, 5.28)
		E group	40			22		

Note: This table was copied from page 78 of the sponsor's summary of clinical efficacy

5.3 Conclusions and Recommendations

Study 19 does not seem very persuasive on its own since many of the secondary analyses were not nominally significant although the study was small. Some alternative subjective interpretation (e.g., risk benefit considerations or orphan drug status) would seem necessary if one was to view this submission as sufficient for approval since it does not contain two studies that were each positive on the prespecified primary hypothesis.

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

01/23/2017

KUN JIN

01/23/2017

I concur with the review.

HSIEN MING J HUNG

01/23/2017