CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

213535Orig1s000

STATISTICAL REVIEW(S)

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 #:	213535
Drug Name:	Risdiplam
Indication(s):	Spinal Muscular Atrophy
Applicant:	Genentech
Date(s):	September 14, 2019
Review Priority:	Standard
Biometrics Division:	I
Statistical Reviewer:	Tristan Massie, Ph.D.
Concurring Reviewers:	Kun Jin, Ph.D., Team Leader James (Hsien-Ming) Hung, Ph.D., Division Director
Medical Division:	Division of Neurology I
Clinical Team:	Rainer Paine, M.D. Teresa Burrachio, M.D., Team Leader
Project Manager:	Brenda Reggetz
Keywords:	Rare Disease

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The data from the placebo-controlled part of the Sunfish study seems supportive of the efficacy of Risdiplam in the rare disease of Spinal Muscular Atrophy, for which there are no approved orally administered treatments.

2 INTRODUCTION

2.1 Overview

The associated IND for the drug development was 128972. Risdiplam, also known as RO7034067, is an orally administered small molecule *SMN2* (survival of motor neuron 2) splicing modifier developed for the treatment of spinal muscular atrophy (SMA). The key studies intended to support efficacy are summarized in Table 1.

Table 1 Efficacy Study Characteristics

Study Name	Phase and Design	Treatment Period	# of Subjects per Arm	Study Population
39055 (Sunfish)	2/3 two-part seamless placebo controlled study	Part 1: exploratory dose finding minimum of 12 weeks placebo controlled then placebo switched to Risdiplam Part 2: Pivotal Dose 24 months -placebo switched at M12 (blinded)	Part 1: 51 patients Part 2: 180 : 60 placebo 120 Risdiplam	Later Onset (probable Type II or III SMA) ;aged 2-25 years 10 countries
39056 (Firefish)	2/3 Open label two-part study	Part 1: dose escalation 24 months treatment Part 2: single-arm 24 months treatment primary time point is Month 12	Part 1: 21 patients Part 2: 41	Infantile onset Type I SMA aged 1-7 months

2.2 Data Sources

The primary efficacy data for Sunfish part 2 were located in the following directory at the time of review.

\\cdsesub1\evsprod\nda213535\0015\m5\datasets\bp39055-part-2-ccod-06sep2019\tabulations\sdtm\za.xpt

The primary sitting data for Firefish part 1 were located in the following directory at the time of review.
\\cdsesub1\evsprod\nda213535\0002\m5\datasets\bp39056\analysis\adam\datasets\adbsid.xpt

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 Sunfish Study (BP39055)

BP39055, (Sunfish) A TWO-PART SEAMLESS, MULTI-CENTER RANDOMIZED, PLACEBO-CONTROLLED, DOUBLE-BLIND STUDY TO INVESTIGATE THE SAFETY, TOLERABILITY, PHARMACOKINETICS, PHARMACODYNAMICS AND EFFICACY OF RISDIPLAM IN TYPE 2 AND 3 SPINAL MUSCULAR ATROPHY PATIENTS
First Patient Enrolled in Part 1: 19 Oct 2016
Last Patient Enrolled in Part 1: 06 Jul 2017
First Patient Enrolled in Part 2: 09 Oct 2017
Last Patient Enrolled in Part 2: 04 Sep 2018

3.2.1.1 Study Design and Endpoints

Study BP39055 is a two-part, operationally seamless, multi-center, randomized, placebo-controlled, double-blind study designed to investigate the safety, tolerability, pharmacokinetics, pharmacodynamics, and efficacy of risdiplam in patients with later onset (Type 2 and Type 3) SMA.
Part 1 was designed in a dose-escalation manner to identify the appropriate dose of risdiplam for Part 2 of the study, and included both ambulant and non-ambulant patients

aged 2 to 25 years with Type 2 and 3 SMA.
Part 2 was designed to confirm the efficacy and safety of treatment with risdiplam at the dose level selected in Part 1 in patients aged 2 to 25 years with Type 2 and non-ambulant Type 3 SMA. The timing of the primary analysis was defined as the point at which the last patient in Part 2 completed 12 months of treatment.

3.2.1.2 Statistical Methodologies

The statistical analysis plan (SAP) version 3 is dated, 21 Oct 2019.

Determination of Sample Size

The purpose of the confirmatory Part 2 of this study is to estimate and test the treatment effect of risdiplam at the selected dose from Part 1 relative to placebo. The target sample size was 168 patients with 112 patients randomized to risdiplam and 56 patients randomized to placebo (2:1 randomization).
For the primary endpoint of the mean change from baseline in the total motor function measure (MFM) score at Month 12, the sample size of 168, allowing for a 10% dropout rate, provides at least 80% power at a two-sided 5% significance level for testing the null hypothesis that the true treatment difference is zero versus the alternative hypothesis, that the true treatment difference is 3 and assuming that the common standard deviation would be 6 (twice the value seen in Vuillerot et al. 2012). This corresponds to a hypothesized effect size of 0.5. The minimal detectable treatment difference is approximately 2.03.
The actual number of patients randomized to and enrolled in Part 2 is 180.

EFFICACY ANALYSIS

The ITT population for Part 2 will be the primary analysis population for all efficacy endpoints. The confirmatory efficacy analyses will only include data from patients randomized into Part 2 of the study. The primary efficacy estimand is based on a hypothetical treatment strategy assuming no prohibited medications intended for treatment of SMA are available and patients continue on their randomized treatment until the primary analysis timepoint. The prohibited medications are defined in the protocol. A treatment policy strategy will also be applied if applicable. For any patients who discontinue study treatment but continue in the study, all data will be included regardless of initialization on prohibited medications.
The baseline/original baseline for Part 2 of the study is defined as the last measurement prior to first dose of the study medication, either placebo or risdiplam. The adjusted baseline is defined as the last measurement prior to the first dose of risdiplam treatment. The adjusted baseline is the same as the original baseline for those patients initially randomized to and receive risdiplam treatment.

Primary Efficacy Endpoint

The primary endpoint in Part 2 is the change from (original) baseline in the total motor function measure 32 (MFM32) score at Month 12. The MFM32 is also to be assessed at Week 17 and Week 35.
The MFM (Bérard et al. 2005) is an ordinal scale constructed for use in patients with neuromuscular disorders. The scale comprises 32 items (MFM32) that evaluate physical function in three dimensions:

- D1 (13 items) evaluates functions related to standing and transfer
- D2 (12 items) evaluates axial and proximal function in supine and sitting position on

mat and chair
• D3 (7 items) evaluates distal motor function
The score of each task uses a 4-point Likert scale based on the patient's maximal abilities without assistance:
• 0: cannot initiate the task or maintain the starting position • 1: performs the task partially
• 2: performs the task incompletely or imperfectly (with compensatory/uncontrolled movements or slowness)
• 3: performs the task fully and "normally"
The MFM total score will be calculated according to the user manual. The 32 scores are summed and then transformed onto a 0 – 100 scale (i.e., sum of 32 items scores divided by 96 and multiplied by 100) to yield the MFM total score expressed as a percentage of the maximum score possible for the scale (the one obtained with no physical impairment). The lower the total score, the more severe the impairment is.
The full MFM32 will be administered to all patients across age groups.
For items that are recorded as "Not Done" in the eCRF, these items are considered as missing with missing item scores. If the MFM has been administered at a visit but item scores are missing, the following rule will be applied to handle missing items.
Input from the holder of the MFM confirmed that score calculation by domain is only possible as follows. For the score calculation by domain, D1, D2, and D3, scores will only be calculated if there is less than 15% of missing data; i.e., for domain D1 and D2, scores will only be calculated if there is a maximum of 2 items missing in each domain; and for domain D3, a maximum of 1 item missing. In addition, total scores will only be calculated where there is a calculated score in all domains D1, D2, and D3.
If there are only two missing items in either D1 or D2, and/or one missing item in D3, the missing items in D1, D2, and D3 will be imputed with "0" prior to the calculation of the total score. Missing MFM total scores will not be imputed. If possible, the same assessor should follow the patient throughout the study.
The hypothesis to be tested is that the difference in the mean change from baseline in the total MFM32 score at Month 12 between risdiplam and placebo (δ) is
$H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$
If the two-sided p-value is $\leq$5%, then the null hypothesis, of no difference in the mean change from in the total MFM32 score at Month 12 between risdiplam and placebo, will be rejected.
The MFM32 total score and the change from original baseline in the total MFM32 score will be summarized descriptively at each timepoint (baseline and each post-baseline scheduled assessment visit) for the ITT population 1) by treatment group and 2) by age groups of 2 to 5, 6 to 11, 12 to 17, 18 to 25 years at randomization for the placebo-controlled period. The MFM32 total score and the change from adjusted baseline in the total MFM32 score will also be summarized 1) by treatment group and 2) by age groups of 2–5, 6–11, 12–17, and 18–25 years old at each timepoint for the all exposure to risdiplam treatment period. The number and percentage of patients with a change from baseline/adjusted baseline MFM32 total score of $\geq$0, 1, 2, 3, and 4 will also be summarized similarly at each timepoint for the placebo –controlled period and the all exposure to risdiplam treatment period.
The Mixed Model Repeated Measure (MMRM) analysis will also be performed on the change from baseline in the total MFM32 score using all data collected in Part 2 up to 12 months.
The model can be expressed as the following:
$Y_i = X_i \beta + Z_i v_i + \varepsilon_i$
where

– Y_i is the n_i x 1 vector of responses for patient i of the dependent variable.
– X_i is the known n_i x p design matrix of fixed effects.
– β is a p x 1 vector of the unknown population parameters relate to the fixed effect.
– Z_i is the known n_i x r random effect design matrix.
– v_i is the r x 1 vector of the unknown parameters for the subject/patient –effect which is distributed as N(0, Σ_v)
– ε_i is the random error term for patient i which is a n_i x 1 vector of random residuals distributed independently as N(0, $\Sigma\varepsilon i$)
– v_i and ε_i are independent.

This is a mixed-effects model which contains components for fixed effects, random effect and the random error term. The dependent variable of this model is the absolute change from baseline total MFM32 score and the fixed effects of the model will include independent variables of the baseline total MFM32 score, treatment group (placebo or risdiplam), time (i.e., relative to the first dose of randomized study medication in weeks—categorical), treatment-by-time interaction, baseline-by-time interaction and the randomization stratification variable of age (categorical: 2 to 5, 6 to 11, 12 to 17, and 18 to 25 years at randomization). The random effect will include the subject/patient effect. Time will be treated as a repeated variable within a patient (random effects). Patient, treatment, and time will be treated as factor variables and baseline total MFM32 score as covariate.

An unstructured variance–covariance matrix will be applied to model the within-patient variability ($\Sigma\varepsilon i$) in the above model. The components of variance and covariance matrix will be estimated by the restricted maximum likelihood method. Denominator degrees of freedom will be estimated using the Kenward–Roger approximation (2009). If the model does not converge, a heterogeneous autoregressive variance-covariance matrix will then be applied to model the within-patient variability in the above model.

Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

Motor Function

- Change from baseline in total score of the Hammersmith Functional Motor Scale Expanded (HFMSE) at Month 12.
- Change from baseline in the total score of the revised upper limb module (RULM) at Month 12.
- Proportion of patients who achieve stabilization or improvement (i.e., a change from baseline ≥0) on the total MFM score at Month 12.
- Proportion of patients with a change from baseline MFM32 total score of 3 or more (≥3) at Month 12.
- Proportion of patients who achieve an improvement of at least one standard error of measurement (SEM; calculated at baseline) on the total MFM score at Month 12
- Change from baseline in the each of the MFM domain scores of D1, D2, D3, and the total combined score of (D1 + D2) and D2 + D3 at Month 12.
- Proportion of patients who achieve stabilization or improvement (i.e., a change from baseline ≥0) on the total HFMSE score at Month 12.
- Proportion of patients who achieve stabilization or improvement (i.e., a change from baseline ≥0) on the total RULM score at Month 12.
- Proportion of patients with a change from baseline HFMSE total score of 2 or more

(≥2) at Month 12.
• Proportion of patients with a change from baseline RULM total score of 2 or more (≥2) at Month 12.

Revised Upper Limb Module
The RULM is a scale that assesses specifically the motor performance of the upper limbs in SMA patients. It consists of twenty items that test proximal and distal motor functions of the arm in patients with SMA. The first entry item, used to determine study eligibility is scored from 0 (no useful function of hands) to 6 (can adduct both arms simultaneously in a full circle until they touch above the head). This item serves as a functional class identification but does not contribute to the total score.
Eighteen of the tasks in the RULM are scored, with
• 0: cannot complete task independently
• 1: modified method but can complete task independently
• 2: completes task without any assistance
The remaining task is scored as a can/cannot score with 1 as the highest score.
The scores for all tasks, except the first entry item, are summed and can range from 0 (no tasks completed) to 37 (all tasks independently completed) with lower scores indicating greater impairment.

For the RULM, a score will be collected for each item on both the left and right side; the highest score will be used in calculating the total RULM score. If 3 or fewer items are missing, the missing items will be imputed to be “0” (unable to perform the task) prior to the calculation of the total score of RULM. If more than 3 items are missing at an assessment timepoint, the total score of RULM at this assessment timepoint will not be calculated.

Adjustment for Multiple Testing
To control the Type I error rate due to multiple testing of risdiplam versus placebo for the primary and the six key secondary efficacy endpoints in the ITT population of Part 2 of the study, a gatekeeping approach will be applied to the seven null hypotheses which are grouped into six families. Hypotheses to be tested are ordered hierarchically and the truncated Hochberg procedure will be used in the family which contains more than one hypothesis. In this study, the truncation fraction is set to 0.95 to allow a relatively higher weight to Family 4 than to subsequent families (Family 5 and Family 6). The following shows the seven null hypotheses and the six families of the testing for Part 2.
• Family 1 includes the hypothesis for the primary endpoint on the change from baseline total MFM32 score at Month 12 comparing risdiplam versus placebo:
H_{11} (MFM32)
• Family 2 includes the hypothesis for the proportion of patients who achieve a change from baseline ≥3 on the total MFM32 score at Month 12 comparing risdiplam versus placebo:
H_{21} (Prop. MFM32 ≥3)
• Family 3 includes the hypothesis for the change from baseline total score of RULM at Month 12 comparing risdiplam versus placebo:
H_{31} (RULM)

• Family 4 includes the hypothesis for the change from baseline total score of HFMSE at Month 12 comparing risdiplam versus placebo and also the hypothesis for the change from baseline best percentage predicted value in FVC at Month 12 comparing risdiplam versus placebo:
H_{41} (HFMSE)
H_{42} (FVC)
• Family 5 includes the hypothesis for the change from baseline in total score of caregiver/parent reported SMAIS at Month 12 comparing risdiplam versus placebo:
H_{51} (SMAIS)
• Family 6 include the hypothesis for the proportion of patients rated by clinician as "Improved" in the CGI-C scale at Month 12 comparing risdiplam versus placebo:
H_{61} (CGI)

3.2.1.3 Patient Disposition, Demographic and Baseline Characteristics

Subject Accountability

A total of 180 patients were enrolled in Part 2 of the study. Of the 180 patients, 120 patients were randomized to treatment with risdiplam and 60 patients were randomized to placebo.
A total of 117 patients (97.5%) in the risdiplam arm and 59 patients (98.3%) in the placebo arm completed the placebo-controlled period (Table 2).
At the time of the CCOD, 4 patients had discontinued the study during the placebo-controlled period and switched to treatment with nusinersen (Spinraza); 3 patients (2.5%) in the risdiplam arm and 1 patient (1.7%) in the placebo arm. Reasons for discontinuation of patients in the risdiplam group were: changed to other treatment (n=1), move to nusinersen (Spinraza) treatment (n=1), and the patient's family requested discontinuation of the study to initiate the use of nusinersen (Spinraza) (n=1). The reason for discontinuation of the patient in the placebo arm was: access to nusinersen (Spinraza) treatment (n=1).

Table 2 Sunfish Part 2 Patient Disposition

	Risdiplam (N=120)	Placebo (N=60)
Safety population	120 (100%)	60 (100%)
ITT population	120 (100%)	60 (100%)
No. completed placebo-controlled period	117 (97.5%)	59 (98.3%)
No. discontinued early from placebo-controlled period	3 (2.5%)	1 (1.7%)
Reason for early discontinuation#		
OTHER : ACCESS TO SPINRAZA	0	1 (1.7%)
OTHER : CHANGED TO OTHER TREATMENT	1 (0.8%)	0
OTHER : MOVE TO SPINRAZA TREATMENT	1 (0.8%)	0
OTHER : THE PATIENT'S FAMILY REQUESTED DISCONTINUATION OF THE STUDY TO INITIATE THE USE OF NUSINERSEN	1 (0.8%)	0

Percentages are based on all randomized subjects.
Safety population includes all subjects randomized in the study who received least one dose of study drug during the placebo-controlled period.
Reason for withdraw for 'other: free text field'.
Clinical Cutoff Date: 06SEP2019

Note: This table was copied from page 14 of the sponsor's study report

Demography

The ITT population includes a wide age range covering childhood to adulthood. At screening, the median patient age was 9.0 years (range: 2–25 years) in the risdiplam arm and 9.0 years (range: 2–24 years) in the placebo arm. Overall, 68 patients (37.8%) were aged 12 years or older at screening.

In each arm, approximately 50% of patients were male and 50% were female. The majority of the population were White (RIS: 66.7%; PLB: 68.3%), and of non-Hispanic or Latino ethnicity (95.0% in each arm). The median height, weight, and BMI were similar across the risdiplam and placebo treatment arms (Table 3).

Table 3 Sunfish Part 2 Baseline Demographics

	Risdiplam (N=120)	Placebo (N=60)	Total (N=180)
Age [years] at screening			
n	120	60	180
Mean (SD)	9.9 (5.8)	10.3 (6.1)	10.0 (5.9)
Median	9.0	9.0	9.0
IQR	5 - 14	5 - 14	5 - 14
Min - Max	2 - 25	2 - 24	2 - 25
Age Group [years]			
n	120	60	180
2 - <6	37 (30.8%)	18 (30.0%)	55 (30.6%)
6 - 11	39 (32.5%)	18 (30.0%)	57 (31.7%)
12 - 17	30 (25.0%)	16 (26.7%)	46 (25.6%)
18 - 25	14 (11.7%)	8 (13.3%)	22 (12.2%)
Sex			
n	120	60	180
Male	59 (49.2%)	30 (50.0%)	89 (49.4%)
Female	61 (50.8%)	30 (50.0%)	91 (50.6%)
Race			
n	120	60	180
Asian	23 (19.2%)	12 (20.0%)	35 (19.4%)
Black or African American	2 (1.7%)	0	2 (1.1%)
White	80 (66.7%)	41 (68.3%)	121 (67.2%)
Multiple	1 (0.8%)	0	1 (0.6%)
Unknown	14 (11.7%)	7 (11.7%)	21 (11.7%)
Ethnicity			
n	120	60	180
Hispanic or Latino	5 (4.2%)	2 (3.3%)	7 (3.9%)
Not Hispanic or Latino	114 (95.0%)	57 (95.0%)	171 (95.0%)
Not Stated	0	1 (1.7%)	1 (0.6%)
Unknown	1 (0.8%)	0	1 (0.6%)
Baseline Weight (kg)			
n	120	60	180
Mean (SD)	32.26 (19.54)	32.95 (19.64)	32.49 (19.52)
Median	27.10	27.15	27.10
IQR	17.0 - 39.9	18.2 - 40.2	17.3 - 39.9
Min - Max	10.4 - 100.0	11.0 - 112.0	10.4 - 112.0
Baseline Height (cm)			
n	119	60	179
Mean (SD)	131.84 (25.80)	133.66 (25.11)	132.45 (25.51)
Median	133.82	131.29	133.18
IQR	109.5 - 154.0	112.8 - 156.4	111.5 - 154.3
Min - Max	85.9 - 180.0	91.0 - 183.6	85.9 - 183.6
Baseline Head Circumference (cm)			
n	25	10	35
Mean (SD)	50.21 (1.69)	50.83 (2.17)	50.39 (1.83)
Median	50.00	52.00	50.50
IQR	49.0 - 51.0	50.3 - 52.2	49.0 - 52.0
Min - Max	46.5 - 54.5	46.0 - 52.4	46.0 - 54.5
Baseline Body Mass Index (kg/m2)			
n	119	60	179
Mean (SD)	16.92 (5.20)	16.99 (5.19)	16.94 (5.18)
Median	15.69	15.79	15.73
IQR	13.3 - 19.8	13.6 - 19.6	13.3 - 19.8
Min - Max	7.8 - 34.1	9.8 - 34.6	7.8 - 34.6

Patient are grouped by initial treatment received.
Clinical Cutoff Date: 06SEP2019

Note: copied from page 16 of the sponsor's study report

Baseline Disease Characteristics

This study included patients with late-onset SMA; of the 180 patients in the ITT

population, 128 patients (71.1%) had Type 2 SMA and 52 patients (28.9%) had Type 3 SMA. The proportion of Type 2 SMA patients was similar across each treatment arm. The majority of patients (87.2%) had 3 copies of the *SMN2* gene (RIS: 89.2%; PLB: 83.3%). The median age of onset of initial SMA symptoms, as reported by the parents or patients, was 12.3 months (range: 0–57 months) in the risdiplam treatment arm and 12.8 months (range: 6–135) in the placebo arm. All but one patient were non-ambulatory. This patient was included because of confusion between the inclusion criteria for the two parts of the study by the site; the patient enrollment was reported as a major protocol deviation.
One hundred and twenty patients (66.7%) had scoliosis at screening, with a higher proportion in the placebo arm (73.3%) than in the risdiplam arm (63.3%). Similarly, a higher proportion of patients in the placebo arm had severe scoliosis with a curvature of >40 degrees (38.3%) than in the risdiplam arm (28.3%). The proportion of patients with scoliosis increased with age (age 2–5: 27.3%; age 18–25: 95.5%), reflecting the progressive character of the disease.

3.2.1.4 Results and Conclusions

3.2.1.4.1 Sponsor's Results

Motor Function Measure 32
Primary Analysis
The primary endpoint in Part 2 of Study BP39055 was the change from baseline in the MFM32 total score at Month 12. In the MMRM analysis of the MFM32 total score, the least square means (lsmeans) (SE) change from baseline at Month 12 was 1.36 (0.38) in patients receiving risdiplam and -0.19 (0.52) in patients receiving placebo (Table 4). This improvement in MFM32 total score with risdiplam treatment when compared to placebo was statistically significant (the least square difference [95% CI] for change from baseline in MFM32 at Month 12: 1.55 [0.30, 2.81]; p=0.0156).

Table 4 Sunfish Part 2, Primary Analysis

MFM 32 Total Score		(N=120)	(N=60)
Baseline	n	115	59
	Mean (SD)	45.48 (12.09)	47.35 (10.12)
MMRM Change from Baseline at Week 52	Lsmeans(SE)	1.36 (0.38)	-0.19 (0.52)
	95% CI	(0.61, 2.11)	(-1.22, 0.84)
MMRM Difference from Placebo	Estimate (SE)	1.55 (0.64)	
	95% CI	(0.30, 2.81)	
	p-Value	0.0156	
	adjusted p-Value	0.0156	

Baseline is the last measurement prior to patients first dose of risdiplam or Placebo.
Patient are grouped by initial treatment received.
Model: Change from baseline = baseline + treatment + visit + age group + treatment * visit + baseline * visit.
Clinical Cutoff Date: 06SEP2019

Program:
/root/clinical_studies/RO7034067/CDPT7916/BP39055/data_analysis/CSR_part2/prod/program/t_ef_mmrm_mfm32.sas
Output:
/root/clinical_studies/RO7034067/CDPT7916/BP39055/data_analysis/CSR_part2/prod/output/t_ef_mmrm_mfm32_DB_IT_P2_06SEP2019_39055.out
29NOV2019 18:05

Note: This table was copied from page 21 of the sponsor's interim study report

Sensitivity Analyses

A treatment policy strategy was applied to the MFM32 total score data as a sensitivity analysis. In the MMRM analysis of the MFM32 total score using the treatment policy estimand, the least squares mean (SE) change from baseline at Month 12 was 1.34 (0.38) in patients receiving risdiplam and -0.19 (0.52) in patients receiving placebo. The difference from placebo in mean (95% CI) change from baseline in MFM32 at Month 12 was 1.52 (0.27, 2.78) (p=0.0178). There was only one patient in Study BP39055 Part 2 who received prohibited treatment (nusinersen, Spinraza) prior to withdrawal of the study. The results based on the treatment policy estimand, which had included all data up to study withdrawal, are consistent with the results of the primary analysis based on the hypothetical estimand.

A tipping point analysis was performed for the primary hypothetical efficacy estimand to determine the delta required to be applied for those patients (n=3) in the risdiplam arm with missing MFM32 total scores, due to withdrawal from the study, in order to overturn the primary analysis results.
A series of analyses were performed with an increasing value of delta until the analysis conclusion of a statistically significant treatment effect no longer held. A delta of -7.75 (representing the tipping point) was required to overturn the primary results, suggesting even a large decline in MFM32 total score would not change the conclusion of the primary analysis. The tipping point analysis was also performed for the treatment policy estimand.

Key Secondary Endpoints

MFM Total ≥ 3 Responder

There was a greater proportion of responders (MFM32 total score ≥3) in the risdiplam arm (38.3%) than in the placebo arm (23.7%) at Month 12; the difference between the risdiplam arm and the placebo arm was statistically significant (odds ratio [95% CI]: 2.35 [1.01, 5.44]; unadjusted p=0.0469; adjusted p=0.0469).

Revised Upper Limb Module

The change from baseline in the RULM at Month 12 was a secondary endpoint in Part 2 of Study BP39055.

In the MMRM analysis of the RULM total score, the least squares mean (SE) change from baseline at Month 12 was 1.61 (0.31) in patients receiving risdiplam and 0.02 (0.43) in patients receiving placebo. This improvement in RULM total score with risdiplam treatment when compared to placebo was statistically significant (difference from placebo in mean [95% CI] change from baseline in RULM at Month 12: 1.59 [0.55, 2.62]; p-value=0.0028; adjusted p=0.0469).

HFMSE

In the MMRM analysis of the HFMSE total score, the mean (SE) change from baseline at Month 12 was 0.95 (0.33) in patients receiving risdiplam and 0.37 (0.46) in patients receiving placebo. This improvement in HFMSE total score with risdiplam treatment when compared to placebo was not statistically significant (least squares difference from placebo in mean [95% CI] change from baseline in HFMSE at Month 12: 0.58 [-0.53, 1.69]; unadjusted p=0.3015; adjusted p=0.3902).

Forced Vital Capacity (FVC)

In the MMRM analysis of FVC (best percentage predicted value), the least squares mean (SE) change from baseline at Month 12 was -5.16% (1.40%) in patients receiving risdiplam and -3.11% (1.94%) in patients receiving placebo. This difference in FVC with risdiplam treatment when compared to placebo was not statistically significant or clinically relevant (difference from placebo in mean [95% CI] change from baseline in FVC at Month 12: -2.05% [-6.67%, 2.56%]; unadjusted p=0.3804; adjusted p=0.3902).

Reviewer's Comment: Both of the two endpoints in family 4 of the key endpoint hierarchy, HFMSE's and FVC's, p-values exceeded 0.05, so the truncated Hochberg for rejecting either hypothesis in this family was not beneficial in this case and testing of subsequent secondary endpoints in the overall hierarchy had to stop.

3.2.1.4.2 Reviewer's Results

The reviewer verified the sponsor's primary and key secondary analyses. Missing data was very limited. There was just one patient affected by the primary use of the hypothetical estimand. This Risdiplam patient switched to Nusinersen and had an early termination assessment. If the patient's early termination visit is included (a MFM change from baseline of -1.04 at week 17), i.e., the treatment policy estimand is used, the resulting estimated treatment difference at Week 52 is 1.52 (Std Err= 0.6363), p=0.0178. The analysis was not stratified by site or region and there was a nominally significant interaction between country and treatment effect on the primary endpoint, MFM32 (see page 22).

3.2.2 Firefish Study (BP39056)

First Patient Enrolled in Part 1: 23 Dec 2016
Last Patient Enrolled in Part 1: 21 Feb 2018

First Patient Enrolled in Part 2: 13 Mar 2018
Last Patient Enrolled in Part 2: 19 Nov 2018
Data cut-off: 27 Feb 2019
There were two parts to the study, both with two year duration but the primary analysis timepoint for the confirmatory part 2 was 12 months. No formal hypothesis testing was planned for Part 1 (dose finding) of the study.
Part 2 of Study BP39056 was designed to test whether the proportion of patients sitting without support after 12 months of treatment would be higher than a performance criterion set at 5%. This 5% threshold was chosen to allow sufficient confidence that any effect seen in treated patients is greater than what could be expected from the natural history of the disease. A statistically significant result would be achieved when a minimum of 6 out of 41 infants are sitting without support for 5 seconds after 12 months of treatment, based on an exact binomial test with a one-sided 5% significance level. If, in Part 2 of the study, 6 out of 41 patients are sitting without support at Month 12, the lower limit of the two-sided 90% Clopper-Pearson (exact) confidence interval (CI) would be above 5%. A similar level of confidence would be achieved in Part 1 if at least 4 out of 21 infants (19.0% [6.8%-38.4%; 90% CI]) are sitting without support after 12 months of treatment. Based on this number, the lower limit of the 90% CI would be above 5%.

Intent-to-Treat Population
The intent-to-treat (ITT) population was defined as all patients enrolled in Part 1 of the study, regardless of whether they received treatment or not. The ITT population is the primary analysis population for all efficacy analyses.

Subject Accountability
A total of 26 patients were screened for enrollment in Part 1 of the study, of whom 5 patients were screen failures. Reasons for screen failure are provided; the most common reason was consent withdrawn (2 patients).
A total of 21 patients were enrolled in Part 1 across 7 different sites in 5 countries (Belgium, France, Italy, Switzerland, and the United States). The first patient was enrolled on 23 December 2016. The last patient was enrolled on 21 February 2018 and completed the Month 12 assessments on 27 February 2019. All patients received treatment with risdiplam.
At the time of the CCOD for this report, 18 of 21 patients (85.7%) in Part 1 were still on study. Three patients (14.3%) died due to SMA-related respiratory complications; of these, two died (b) (6) and one died (b) (6) respectively). As of the CCOD, 1 patient had completed the 24-month treatment period of Part 1 and had entered the OLE phase.

Cohort 1 includes the first 3 patients enrolled in the study who received Dose Level 1 for at least 12 months and the patient enrolled at Dose Level 1 who discontinued from the study on Study Day 19. Cohort 2 includes the infants enrolled at Dose Level 2 and the patient whose dose was escalated to Dose Level 2 on Study Day 83.

Demographic and Baseline Disease Characteristics
Of the 21 patients enrolled in Part 1, 15 were female (71.4%). The majority of patients were White (17/21; 81.0%) and 4 (19.0%) were Asian; none were of Hispanic or Latino

ethnicity. Overall, patients in Part 1 were on the older end of the permitted age range, with a median age of 6.7 months (range: 3.3-6.9 months) at enrollment. The mean age of patients was higher in Cohort 1 (6.8 months) than Cohort 2 (5.6 months), which reflects the staggered approach to enrollment during the dose-selection period and the inclusion criterion that the first 3 patients enrolled in Part 1 had to be at least 5 months old.

The baseline characteristics of patients in Part 1 were typical of a Type 1 SMA population, with a median age of 2.0 months (range: 0.9-3.0 months) at symptom onset. The median disease duration (time between onset of symptoms and first treatment) was 4.0 months (range: 2.0-5.8 months).
All patients had an *SMN2* copy number of 2, as required by the study inclusion criteria. Median baseline scores for CHOP-INTEND (24.0), BSID-III (2.0), HINE-2 (1.0), and CMAP amplitude (0.2 mV) were low, as expected for this symptomatic patient population. No patients were sitting without support at baseline.

Twenty of 21 patients (95.2%) had at least one prior medication that was taken and completed prior to starting risdiplam treatment.
A total of 21 patients (100%) received prior and ongoing medications (i.e., that were started prior to onset of risdiplam treatment and that were ongoing at the date of first dose). The most common medications, by medication class, were antispasmodics and anticholinergics (16 patients, 76.2%); adrenergics/sympathomimetics (14 patients, 66.7%); and herbal, homeopathic, and dietary supplements (6 patients, 28.6%).
Twenty patients (95.2%) received concomitant medications for AEs on or after the date of the first dose of risdiplam.
Eleven patients (52.4%) received concomitant medications for prophylaxis (defined as those within the medication classes of Vitamins and Minerals or Vaccines, Toxoids, and Serologic Agents).
Twelve patients (57.1%) received other concomitant medications (i.e., medications not taken for AEs or prophylaxis) on or after the date of the first dose of risdiplam. The most common medications, by medication class, were antispasmodics and anticholinergics (6 patients, 28.6%) and herbal, homeopathic, and dietary supplements (5 patients, 23.8%).

Two patients had undergone a total of 3 SMA-related surgeries or procedures prior to treatment with risdiplam (by MedDRA preferred term, these were decreased oxygen saturation, orthosis user, and tracheostomy). The previous history of tracheostomy in 1 patient was closed prior to study participation. No patient had a tracheostomy at the time of study enrollment.
Ten patients (47.6%) underwent SMA-related surgeries or procedures after the date of first risdiplam dose; those performed in more than 1 patient, by MedDRA preferred term, were orthosis user (5 [23.8%]), central venous catheterization (3 [14.3%]), and sleep studies and gastrointestinal tube insertion (each 2 [9.5%]).

3.2.2.1.1 Sponsor's Results

Clinical efficacy data from at least 12 months of risdiplam treatment for patients in Part 1 of the study indicate that 33.3% (7/21) of patients achieved sitting without

support for 5 seconds or more, as assessed by Item 22 of the BSID-III gross motor scale. The 90% confidence interval for this proportion is (16.8, 53.6). All of the patients who reached this milestone at 12 months were in Cohort 2 (7/17, 41.2% in Cohort 2) and received risdiplam at the target exposure of mean $AUC_{0\text{-}24h,ss} \leq 2000$ ng/h/mL (the pivotal dose). This is a significant deviation from the well-established expected natural history for patients with Type 1 SMA, who will never sit without support (Finkel et al. 2014a, 2015; Faravelli et al. 2015; De Sanctis et al. 2016; Kolb et al. 2017).

- Motor milestone development was also further confirmed by the HINE-2, as a second independent assessment; at Month 12, 66.7% of patients (14/21) were classified as motor milestone responders (defined as having more milestones that showed improvement than showed worsening) as assessed by the HINE-2.
- Treatment was also associated with an improvement in motor function. At Month 12, 52.4% of patients achieved a CHOP-INTEND total score of 40 or higher, and 85.7% of patients (including all patients who were alive at Month 12 except for 1 patient who was uncooperative at that visit) achieved an increase of at least 4 points in their CHOP-INTEND score from baseline. According to the sponsor these results are markedly different compared to natural history.

In general, patients in Cohort 2, who received the selected pivotal dose (at a higher target exposure than Cohort 1) of risdiplam within the first 12 months of treatment, had better efficacy outcomes compared with patients in Cohort 1.
As of the CCOD, 3 patients (14.3%) in Part 1 had died and the remaining 18 patients (85.7%) were alive without permanent ventilation.

3.2.2.1.2 Reviewer's Results

Part A of the study was initially designated as exploratory but because the study is uncontrolled and open label and the sponsor began to see favorable outcomes they decided to consider part A as part of the potentially confirmatory evidence. This creates a multiplicity issue. The confidence interval of the binomial distribution probability depends on the number of trials, subjects treated in this case, which changed when the sponsor designated part A as confirmatory. Part B of Firefish, which was to be confirmatory, is still ongoing or at least not yet submitted to the NDA. It would be important to obtain the confidence interval for the probability of sitting from the part B when it is available. An exploratory analysis that might be helpful is a negative binomial design analysis, i.e., how many patients would need to be treated before the goal of 4 successes was reached if the true chance of sitting at Month 12 was 5%. Reaching 4 successes in the first 18 subjects or less would equate to significance at the one-sided 2.5% level. Similarly, reaching 7 successes in 50 or fewer subjects would suggest significance at the one-sided 2.5% level under a hypothetical negative binomial design. This assumes subjects are independent and also is not the true design used.

Forty one patients were enrolled in Part 2 of Firefish the 12 month primary endpoints have not been submitted as Part 2 was ongoing at the time of NDA submission. The sponsor made a press release that part 2 met its primary endpoint in January, 2020 but the press release provided no

details on the observed proportion sitting at 12 months, nor has this information been submitted to the NDA.

3.3 Evaluation of Safety

Safety in general is not addressed in this review. Please see the Clinical safety review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

4.1.1 Gender, Race, and Age

In Sunfish Part 2, 51% of randomized patients were female, 67% were White, 19% were Asian and the remainder were Unknown race12%, Black or African American 1%, and Multiple Races 1%. Thirty one percent were aged 2-<6, thirty two percent were 6-11, twenty six percent were 12-17, and twelve percent were 18-25.

The difference in MFM32 at Week 52 was bigger in females than males (1.74 [S.E. .91] vs. 1.39 [S.E. .91]) the interaction term for sex*Visit had a p-value of 0.0679 but the interaction term for sex*arm*Visit had a p-value of 0.6660. An F test for the significance of all terms involving sex (6 numerator degrees of freedom) had a p-value of .1949. The treatment difference estimate was in the same direction for both sexes, so the smaller effect in males might be a matter of sample size.

Table 5 Gender Subgroup Analysis of MFM32 at Week 52

Subgroup	Estimate Treatment Effect	Std. Error	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Female	1.7393	0.9068	-0.05110	3.5297
Male	1.3904	0.9010	-0.3885	3.1693

There was a significant Age group main effect on the change from baseline in MFM 32 as well as an interaction between Age group and Visit (p=0.0387). An F test for the significance of all terms involving sex (6 numerator degrees of freedom) had a p-value of 0.0593. The estimated treatment difference in the 18-25 age group was in the wrong direction (numerically favored placebo -.65 +/- 1.71), but this might be unreliable due to the small sample size in this subgroup.

Table 6 Age Group Subgroup Analysis of MFM32 at Week 52

Subgroup	Estimate Treatment Effect	Std. Error	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Age 2-<6	3.1371	1.1783	0.8105	5.4636
Age 6-11	1.5795	1.0943	-0.5812	3.7402
Age 12-17	1.0405	1.1892	-1.3075	3.3886
Age 18-25	-0.6451	1.7145	-4.0303	2.7402

Black or African American, Multiple, Other, and Unknown race categories were combined by the reviewer because of small numbers, prior to the investigation of treatment differences by Race. There was a nominally significant interaction between Race and Treatment group (p=0.0099) although the Race main effect was not significant (p=0.4148). An F test for the significance of any effects involving Race (12 numerator degrees of freedom) had a p-value of .0283. The estimated treatment difference in Asians was in the wrong direction (-2.05 +/- 1.38), numerically favoring placebo, as compared to 2.50 +/- 0.76 in Whites and 2.58 +/- 1.75 in Unknown/Others, both favoring Risdiplam.

Table 7Race Subgroup Analysis of MFM32 at Week 52

Subgroup	Estimated Treatment Effect	Standard Error	Lower 95% Conf Limit	Upper 95% Conf Limit
Asian	-2.0452	1.3803	-4.7708	0.6803
Others	2.5804	1.7541	-0.8833	6.0442
White	2.4957	0.7603	0.9943	3.9970

Part 1 of the Firefish study only had 21 total patients, so analysis of subgroups in Firefish are unlikely to be reliable.

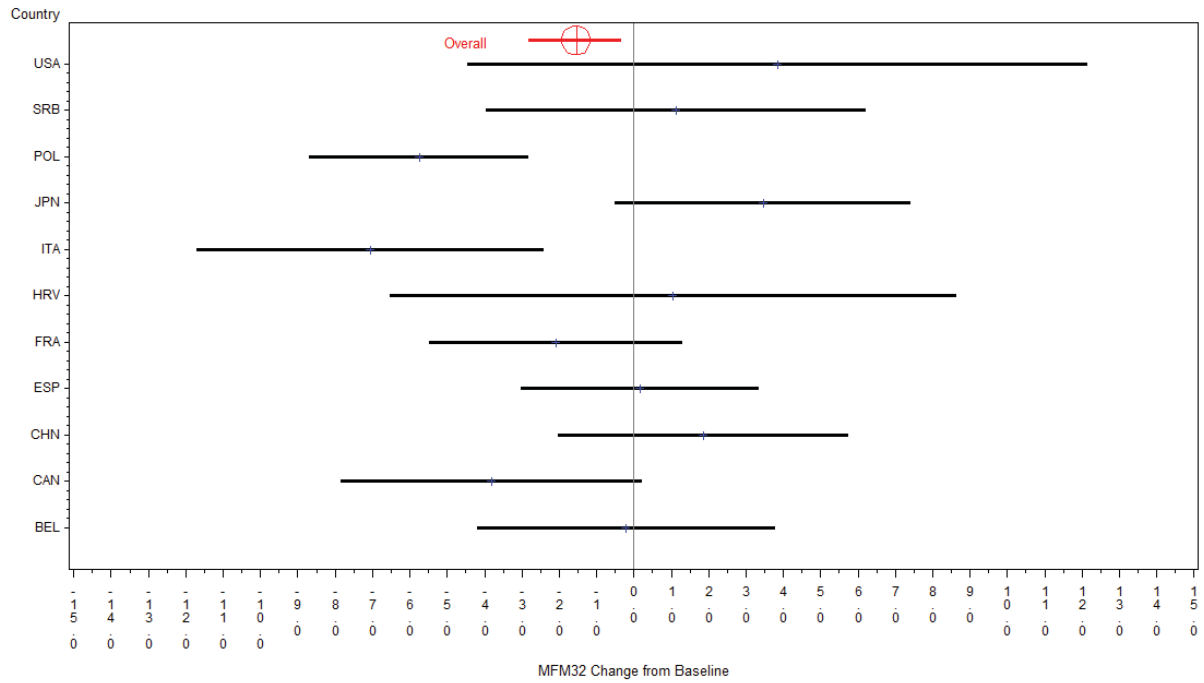
4.1.2 Geographic Region

There were 13 countries in Sunfish Part 2 from Western Europe, Eastern Europe, South America, North America (Canada), and Asia. There were 42 sites among the 13 countries in Sunfish Part 2 (N=180). Randomization was 2:1 and stratified by age group but not by site or country. Twenty four of the 42 sites had at least one randomized patient in both treatment groups at Week 52. In the US there were a total of just 4 patients between two sites or 2% of the randomized population. Poland accounted for 19% (N=32), Spain and France accounted for 12% each, Canada and Japan both accounted for 9%, and Italy accounted for 6%. The rest of the countries had smaller contributions.

There was a significant country main effect (p=0.0364) on the Change from baseline in MFM32 when country was added to the primary model to check for consistency across countries. A likelihood ratio test of the primary analysis model augmented with country effects, country by visit, country by treatment, and country by treatment by visit effects versus the primary model which did not adjust for any of these yielded a Chi square p-value of 0.0122, suggesting that there was statistically significant variation (lack of consistency) in the treatment effect across countries.

A forest plot of treatment effects on MFM32 at Week 52 by Country follows. Higher MFM32 scores are better and the difference is presented as Placebo – Risdiplam so that negative differences favor Risdiplam.

Figure 1 Forest Plot of Change in MFM32 at Month 12 by Country

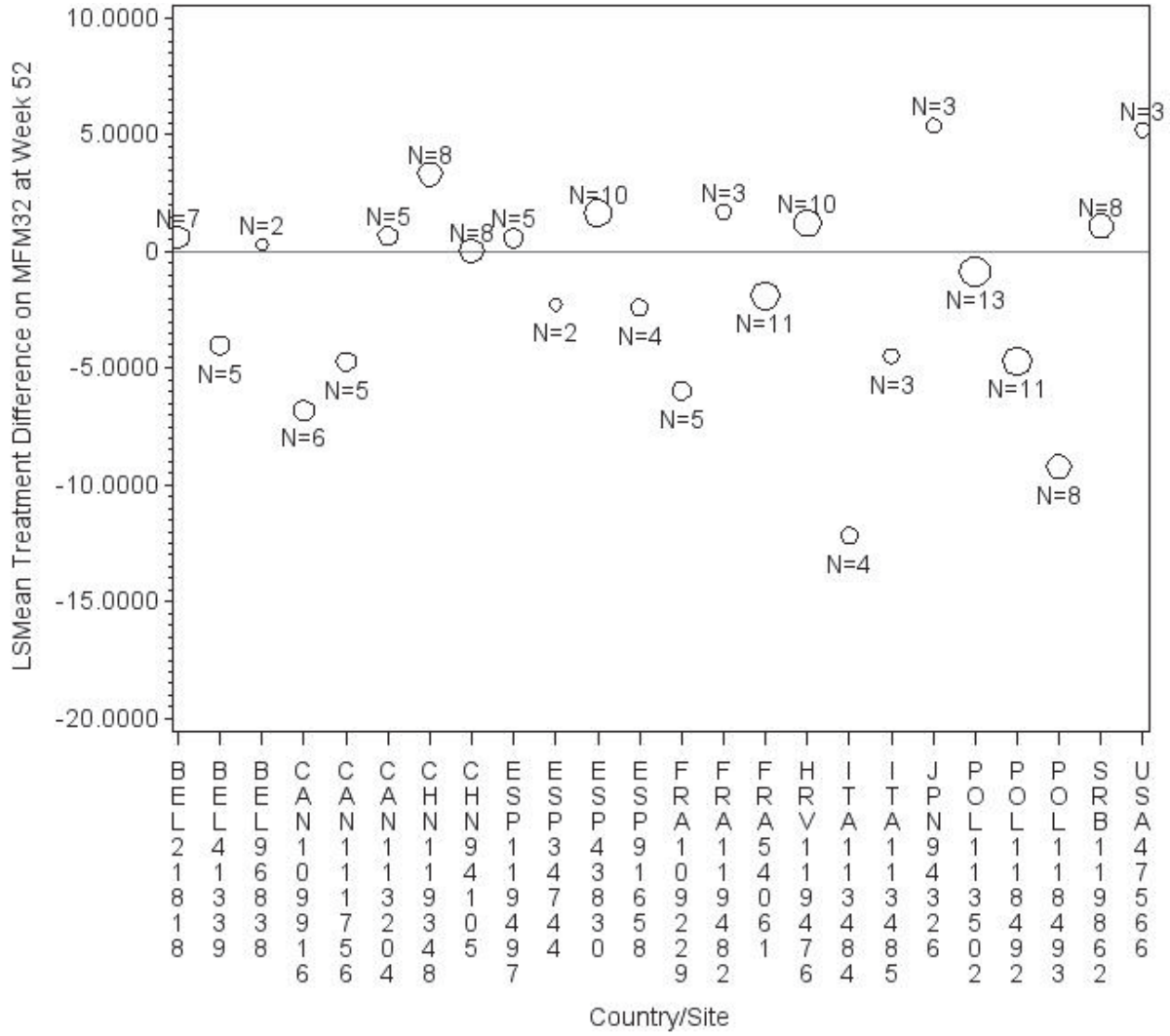


4.1.2.1 Individual Sites

There were 42 sites among the 13 countries in Sunfish Part 2. Twenty four of the 42 sites had at least one randomized patient in both treatment groups at Week 52. In the US there were a total of just 4 patients between two sites or 2% of the randomized population. Poland accounted for 19%, Spain and France accounted for 12% each, Canada and Japan both accounted for 9%, and Italy accounted for 6%. The rest of the countries had smaller contributions.
The F test for the significance of the country effects alone (i.e., not considering interactions, but only whether countries directly affected the primary outcome, MFM32) had a p-value of 0.0916. In order to investigate the real question of interest related to regions, i.e., whether the treatment effect is consistent across countries we have to introduce interaction effects between Country and Treatment and Visit to the model. An F test for the significance of Country and all Country interactions with Treatment and Visit had a p-value of 0.0534.
Exploratory analyses investigating the sensitivity of the primary analysis result to the exclusion of individual sites suggested that the exclusion of the Polish site 118493 or of site 118492 could increase the overall p-value to greater than 0.05 (without 118493[N=8] Week 52 MFM32 treatment effect p=0.1005 or without 118492 [N=11] p=0.0567).

The following figure (Figure 2) shows treatment effect at Week 52 on MFM32 by all individual sites. The sample size is shown next to the plotted symbol. The plotted symbol is larger for sites with larger sample sizes./

Figure 2 Subgroup Analysis of MFM32 by Site



In Firefish Part 1, one site (#107431) had 4/7 patients sit without support at 1 year who could not at baseline which was the highest response rate across sites.

4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

The key statistical issues are as follows.

- There was some inconsistency in effects across countries for the primary endpoint. However, some variability is to be expected and overall, the primary analysis was statistically significant. Inspection of study sites had to be cancelled due to the COVID-19 pandemic.
- There being only one controlled study for evidence of efficacy, but this is a rare and very serious disease.
- Part 1 of the open label single arm study, Firefish, in the more severe Type 1 SMA was submitted for evidence of efficacy even though it was originally designated as exploratory. This creates a multiplicity issue because knowledge of the results influenced the change in designation. j However, the key source of evidence for efficacy is the placebo controlled study, Sunfish part 2, because the evidence from Firefish, though impressive on face compared to the reported natural history, is not well controlled.

5.2 Collective Evidence

Collective evidence is not considered in this review since there was only one double-blind, controlled trial.

5.3 Conclusions and Recommendations

The data from the placebo-controlled part of the Sunfish study seems supportive of the efficacy of Risdiplam in the rare disease of Spinal Muscular Atrophy, for which there are no approved orally administered treatments.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

TRISTAN S MASSIE
05/13/2020 05:08:57 PM

KUN JIN
05/13/2020 05:51:30 PM
I concur with the review.

HSIEN MING J HUNG
05/14/2020 08:03:17 AM