CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

216718Orig1s000

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#:	NDA216718 (SN 01 - 35)
Related IND#:	IND122349
Drug Name:	RTA 408 (Omaveloxolone capsule 50 mg strength at 3 capsules per day)
Indication(s):	FA (Friedreich's Ataxia)
Applicant:	Reata Pharmaceuticals, Inc.
Date(s):	Stamp date: 3/30/2022 (date complete package received by CDER)
	PDUFA date: 2/28/2023 (Extended from 11/30/2022 to 2/28/2023 on 8/9/2022 because of major amendments received after mid cycle)
	AC meeting: Canceled (10/4/2022 Changed to 1/19/2023 then canceled.)
Review Priority:	Priority
Biometrics Division:	Division of Biometrics 1
Statistical Reviewer:	(Joanne) Jinnan Liu, Ph.D., Statistical Reviewer
Concurring Reviewers:	John Lawrence, Ph.D., Secondary Statistical Reviewer
	Hsien Ming James Hung, Ph.D., Division Director, Tertiary Statistical Reviewer
Medical Division:	Division of Neurology I
Clinical Team:	Veneeta Tandon, Pharm.D., Clinical Reviewer
	Emily Freilich, M.D., Clinical Team Leader
	Teresa Buracchio, M.D., Division Director
Project Manager:	Brenda Reggettz

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

This NDA submission is to seek approval of RTA408 for Friedreich's Ataxia (FA) which is a rare disease. The efficacy evidence for RTA408 (Omaveloxolone oral dissolving capsule 50-mg strength at 3 capsules per day) is based on one Phase II study 408-C-1402-PT2. The study was a multi-center, randomized, double-blinded, parallel group and placebo-controlled study to evaluate the efficacy of RTA408 (Omaveloxolone) versus placebo for the treatment of FA. The study was completed.

This study provided statistical evidence to support that RTA408 is efficacious in patients with for FA based on change from baseline mFARS (modified Friedreich's Ataxia Rating Scale) score at week 48.

2 INTRODUCTION

One study 408-C-1402-PT2 was included in this statistical review for efficacy evaluation.

2.1 Overview

Friedreich's ataxia (FA) is a rare, genetic, neurodegenerative disorder. Currently there is no approved therapy. Majority of patients have disease onset before 15 years of age, with a mean duration until wheelchair use of 11 to 14 years after disease onset and a median survival of 34 to 37 years after disease onset.

Triterpenoids are a group of natural products derived from plant extracts (examples: rosemary, olive leaves, and soybeans) which have been used extensively in Asian medicine.

Omaveloxolone (also known as RTA 408), as one of a class of semisynthetic triterpenoids, is hypothesized to be therapeutic in patients with FA.

Table 1 List of All Studies included in Review

Trial ID	Design*	Treatment/ Sample Size	Endpoint/ Analysis	Preliminary Findings	mFARS Data Entry Error Correction Addendum
408-C-1402-pt2	MC, R, DB, PG, PC	Placebo / 42 Drug / 40	**Primary:** Change from Baseline in mFARS score at Week 48 **Two Key Secondary:**	-1.55 in drug group 0.85 in placebo -2.40 mean diff between drug and placebo with p-value 0.0141 3.89 in drug group 4.32 in placebo	-1.56 in drug group 0.85 in placebo -2.41 mean diff between drug and placebo with p-value 0.0138

			PGIC score at Week 48 CGIC score at Week 48	-0.43 mean diff between drug and placebo with p-value 0.1300 3.92 in drug group 4.06 in placebo group -0.13 mean diff between drug and placebo with p-value 0.5259	

* MC: multi-center, R: randomized, DB: double-blind, PG: parallel group, PC: placebo controlled, mFARS = modified Friedreich's Ataxia Rating Scale, PGIC = Patient Global Impression of Change, CGIC = Clinical Global Impression of Change. (Source: sponsor's study report replicated but **NOT** verified by the reviewer.)

2.2 Data Sources

All documents reviewed for this NDA submission are in electronic form.

At the time of review, the following is the link to the EDR Location for

Study 1402 Part 2 CSR original version and Datasets:
\\CDSESUB1\evsprod\NDA216718\0001

Study 1402 Part 2 Corrected Datasets (Data Entry Error):
\\CDSESUB1\evsprod\NDA216718\0029

Study 1402 Part 2 CSR addendum (Data Entry Error):
\\CDSESUB1\evsprod\NDA216718\0035

3 STATISTICAL EVALUATION

One study (Study 408-C-1402-PT2) is included in this statistical review.

3.1 Data and Analysis Quality

There are some potential review issues we discovered at the time of filing review.

On face, this study was statistically positive on the primary endpoint with a p-value of 0.0141 and effect size of about -2.4 points. The two key secondary endpoints had p values 0.1300 and 0.5259, respectively. Almost none of the 5 other secondary endpoints showed a treatment effect, except the Activities of Daily Living score (the last one in the testing hierarchy) had a nominal p-value of 0.04 with a treatment difference of about 1.30 points between drug and placebo. The 4 functional measures, the 9-HPT, 25-foot timed walk test, frequency of falls, and peak work, were not statistically significant.

For the primary endpoint, 1 patient in the placebo group and 6 in the drug group had missing data. Such 'imbalance' in missingness is a potential issue for bias, since this is a relatively small study with about 40 patients randomized in each group to start with. We noticed such imbalance at the filing meeting and sent an IR. The applicant did more checking at the site level and responded with additional information and details about the "admin reasons" for 2 subjects. The two subjects had to leave early to catch flight. Refer to the section on missing data and sensitivity analyses of this review for more details.

The sponsor conducted several post-hoc analyses; two major ones were named as "delayed start study" and "baseline-controlled study". However, only a small subset (less than 50%) of patients staying at week 48 (end of DB period) continued to week 72 (in the extension study), so the randomization protection was non-existent in these two analyses. The results were hard to interpret. The delayed-start study provides analysis of patients who received omaveloxolone during the double-blind study 408-C-1402 Part 2 to patients initially randomized to placebo who subsequently began omaveloxolone in study 408-C-1402 Extension, comparing the difference in effect of omaveloxolone treatment on disease course. The baseline-controlled Study provides an analysis of treatment-naïve patients (from study 408-C-1402 Part 1 and Part 2) serving as their own controls to assess whether there is a significant difference in mFARS scores between the pre-treatment and treatment periods.

The sponsor sent in more data for OLE (Open Label Extension) after mid cycle, however, the number of subjects with data at both extension week 48 and 120, appears to be at most 50% to 60% of the randomized FAS population (n=82) at end of DB (n=75).

Data entry error regarding primary endpoint was discovered late in the review, although it did not alter the efficacy conclusion and the impact was minimal. Please see appendix for more details regarding this data entry error.

3.2 Evaluation of Efficacy Study – 408-C-1402-PT2

3.2.1 Study Design and Endpoints

Study 408-C-1402-PT2 was a Phase 2, multicenter, randomized, double-blind, placebo-control, parallel group study of subjects with FA.

Study 408-C-1402 was conducted in 2 parts. The first part of the study (Part 1) was a randomized, placebo-controlled, double-blind, dose-ranging study to evaluate the safety, efficacy, PK, and PD activity of omaveloxolone at 2.5 mg, 5 mg, 10 mg, 20 mg, 40 mg, 80 mg, 160 mg, and 300 mg in patients with FA. The second part of the study (Part 2) was a randomized, placebo-controlled, double-blind, parallel-group study to evaluate the safety, PK, and efficacy of omaveloxolone 150 mg in patients with FA.

The **primary endpoint** was the change from the baseline in the mFARS score at week 48 of the 48-week double-blind treatment period.

The full neurologic FARS (maximum score 125) is divided into 5 sections: bulbar (score 0 to 11), upper limb coordination (score 0 to 36), lower limb coordination (score 0 to 16), peripheral nervous system (score 0 to 26), and upright stability (score 0 to 36). The mFARS (score 0 to 99) is not a separate assessment but rather calculated as the full neurologic FARS assessment minus section D (peripheral nervous system). A higher score on the FARS or mFARS examination signifies more severe physical impairment, and a reduction in FARS or mFARS score signifies improvement in functioning.

Two Key Secondary endpoints
PGIC (Patient Global Impression of Change) score at Week 48
CGIC (Clinical Global Impression of Change) score at Week 48

The PGIC is a 7-point scale that requires patients to assess how much their illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy, 1976). The PGIC is assessed by completing the following statement “since I began trial treatment, my overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), and very much worse (7).” Patients completed this assessment following completion of the FARS neurological assessment or the CGIC.

The CGIC is a 7-point scale that requires the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy, 1976). The CGIC is assessed by completing the following statement “Compared to the patient’s condition at the start of the trial, this patient’s overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), and very much worse (7).” Clinicians could take prior results of study assessments into account to influence their clinical judgment of the CGIC. Clinicians completed this assessment following the patient’s completion of the FARS neurological assessment or the PGIC.

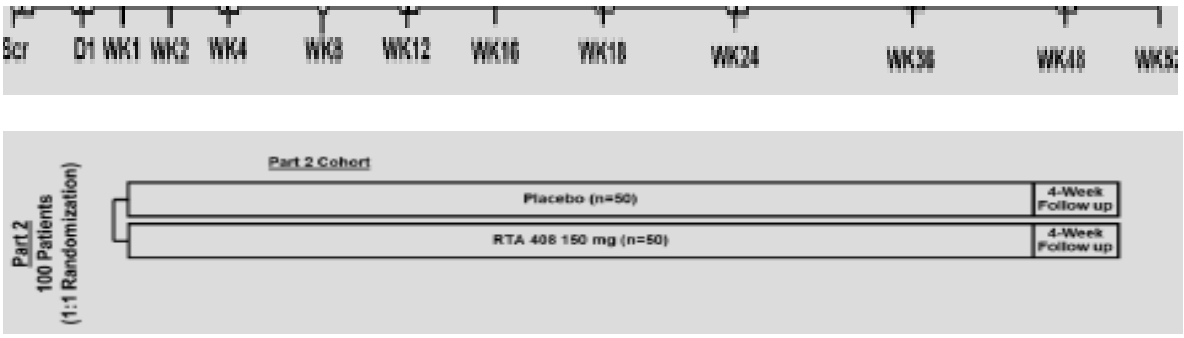
Five Other Secondary endpoints

1. Change in performance on a 9-hole peg test (9-HPT) at Week 48
2. Change in performance on a timed 25-foot timed walk test (T25-FWT) at Week 48
3. Frequency of falls over 48 weeks
4. Change in peak work during maximal exercise testing at Week 48
5. Change in the Activities of Daily Living (ADL) score at Week 48

Please refer to the clinical review for more details on how these 5 secondary endpoints were defined and assessed.

The study consisted of a 4-week screening and a 48-week double blind treatment phase.

Figure 1 Trial Design – Study 408-C-1402-PT2



(Source: adapted from Sponsor's study protocol)

The study population included male or female, 16 to 40 years of age, genetically confirmed FA (Friedreich's Ataxia), had a mFARS score (average of Screening and Day 1, the two scores must be within 4.5 points of each other) 20 to 80, able to swallow capsules, had no changes to their exercise regimen within 30 days prior to Study Day 1 and be willing to remain on the same exercise regimen during the study period, had the ability to complete maximal exercise testing, and had not used antioxidant supplements, including but not limited to idebenone, coenzyme Q10, nicotinamide, and vitamin E above the recommended daily allowance, within 14 days prior to Study Day 1, or plan to take any of these supplements during the time of study participation.

During the Randomized-Controlled Period, patients in the Omaveloxolone or placebo groups were to self-administer study treatment once daily for 48 weeks.

Approximately 100 subjects were planned to be randomized with a ratio of 1:1 to the drug, or the placebo group, stratified by pes cavus status (pes cavus vs. no pes cavus). Patients with pes cavus, a musculoskeletal foot deformity characterized by high arch of the foot that does not flatten with weight-bearing, would not comprise more than 20% of patients enrolled in Part 2

Roughly 45 subjects per group (after about 10% dropout) were expected to be included in the efficacy data set. Assuming RTA408 provides roughly a 2 points further reduction over placebo

on the primary endpoint, and a common standard deviation of 3.5, the study was planned to have roughly 80% power on the primary endpoint.

Efficacy assessments for the primary endpoint of mFARS score were planned at 6 timepoints after dosing: week 4, 12, 18, 24, 36, and 48.

Figure 2 Schedule of Efficacy Assessments – Study 408-C-1402-PT2

Visit Number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13	Visit 14
Study Day/Week	Screening	Day 1	Week 1 (telephone)	Week 2	Week 4	Week 8 (telephone)	Week 12	Week 18	Week 24	Week 30 (telephone)	Week 36	Week 42 (telephone)	Week 48 End of Treatment	Week 52[a] End of Study/ 4-week follow-up
Day Relative to First Dose	-60 to -1[b]	1[c]	7 (±3 days)	14 (±3 days)	28 (±3 days)	56 (±3 days)	84 (±3 days)	126 (±3 days)	168 (±3 days)	210 (±3 days)	252 (±3 days)	294 (±3 days)	336 (±3 days)	192 (±3 days)
Maximal exercise test[l]	X	X			X		X	X	X		X		X	
Neurologic FARS (practice)	X[m]													
Neurologic FARS[n]	X[o]	X[o]			X		X	X	X		X		X	
9-hole peg test	X	X							X				X	
25-foot timed walk test	X	X[p]							X[p]				X[p]	
Video Collection of Normal Walking [q]		X							X				X	
SF-36 Health Survey Update		X							X				X	
Activities of Daily Living	X								X		X		X	
Patient Global Impression of Change							X[r]		X[r]		X[r]		X[r]	
Clinical Global Impression of Change							X[r]		X[r]		X[r]		X[r]	
Falls Diary	X[s]	←---------- X ----------→												
PK analysis				X[t]			X[u]		X[u]				X[u]	

(Source: adapted from Sponsor's study protocol)

3.2.2 Statistical Methodologies

The primary endpoint was analyzed using a mixed effect repeated measures (MMRM) model that included treatment group, mFARS score at baseline, site, visit, baseline-by-visit and the treatment-by-visit interactions, using analysis visits 4, 12, 18, 24, 36, and 48, and assuming an unstructured covariance matrix.

The study SAP was changed 3 times close to the end of study, and site was added to be a covariate in the model in the last version of SAP. The model produced very similar result whether including site as covariate or not. Please refer to study results section in this review for more details.

Efficacy analyses were based on the FAS (Full Analysis) population. Subjects were grouped based on the assigned treatment.

FAS population was defined as randomized patients without pes cavus who were treated and had at least 1 post-baseline efficacy assessment.

Type 1 error was controlled by hierarchical testing. The primary endpoint was analyzed at the 0.05 level. If the primary endpoint was significant, the 2 key secondary and 5 secondary endpoints were to be analyzed at the 0.05 level in the following pre-specified order.
1. PGIC at Week 48 (key secondary)
2. CGIC at Week 48 (key secondary)
3. Change in 9-HPT (reciprocal time measure of the non-dominant hand) at Week 48
4. Change in T25-FWT (reciprocal time measure) at Week 48
5. Frequency of falls over 48 weeks
6. Change in peak work during maximal exercise testing at Week 48
7. Change in ADL score at Week 48

The primary analysis method and multiplicity adjustment plan were prespecified and appropriate.

3.2.3 Patient Disposition, Demographics and Baseline Characteristics

This study had its first subject enrolled on 20 October 2017, and last patient completed on 31 October 2019. The study report was created on the date of 05 November 2020.

Of the 156 screened patients, 53 patients were screen failures. In total, 103 patients were randomized and treated (placebo: N = 52; Omaveloxolone: N = 51). Of the 103, 83 patients were without pes cavus (placebo: N = 42; Omaveloxolone: N = 41).

Out of the 83 randomized patients without pes cavus, 82 were included in the FAS (Full Analysis) population. One patient was not included in FAS due to not having any post-baseline mFARS score ("Week 4 visit was not completed due to flight delay/missed flight. Patient withdrew from study."), although randomized and treated. This patient was included in the SAF (safety population).

Table 2 Subject Dispositions - Study 408-C-1402-PT2

Patient Counts	RTA408	Placebo	Total
All Randomized/Safety Population	41	42	83*
FAS (Full Analysis) Population	40	42	82
Per-Protocol Population	29	32	61

*1 patient (b) (6) was not included in FAS due to not having any post-baseline mFARS score.
(Source: Table 7 in the sponsor's study report, verified by the statistical reviewer.)

Demographic variables such as age and race were similar across the drug and placebo groups. We note that the rate of female was higher at 60% in the RTA408 arm comparing to 33.3% in the placebo arm.

Table 3 Demographics– Study 408-C-1402-PT2– FAS Population

	RTA 408 (N=40)	Placebo (N=42)	All Subjects (N=82)
Age (years), n			
Mean (SD)	23.6 (7.79)	24.2 (6.48)	23.9 (7.14)
Median	21.0	23.0	21.0
Minimum	16	16	16
Maximum	40	39	40
Gender, n (%)			
Female	24 (60.0%)	14 (33.3%)	38 (46.3%)
Male	16 (40.0%)	28 (66.7%)	44 (53.7%)
Race, n (%)			
White	40 (100.0%)	40 (95.2%)	80 (97.5%)
Other	0	2 (4.8%)	2 (2.5%)
BMI (kg/m2), n			
Mean (SD)	23.6(5.4)	22.9 (5.1)	23.3 (5.2)
Median	23.3	21.9	22.8
Minimum	17.4	15.6	15.6
Maximum	43.0	40.5	43.0

BMI = body mass index (Source: Table 8 from study report, verified by the statistical reviewer.)

At baseline, disease characteristics were consistent with a FA population and similar across the treatment groups.

Table 4 Baseline Disease Characteristics – Study 408-C-1402-PT2– FAS Population

	RTA 408 (N=40)	Placebo (N=42)
Age at FA Onset (Years)		
Mean (SD)	15.9 (5.74)	15.1 (5.34)
Median	15.0	15.0
Min, Max	7, 36	5, 30
Years Since FA Onset (Years)		
Mean (SD)	4.8 (3.98)	4.7 (4.70)
Median	4.0	3.0
Min, Max	0, 16	0, 17
mFARS		
Mean (SD)	40.94 (10.39)	38.77 (11.03)
Median	39.15	35.65
Min, Max	24.3, 59.3	19.8, 63.0
Peak Work (watt/kg)		
Mean (SD)	1.08 (0.54)	1.18 (0.61)
Median	1.02	1.06
Min, Max	0.16, 2.05	0.23, 2.65
ADL		
Mean (SD)	10.74 (4.77)	9.87 (4.83)
Median	11.50	10.00
Min, Max	2.0, 21.0	1.0, 20.5
GAA1 repeat length	31	36
Mean (SD)	739.2 (214.88)	693.8 (277.19)
Median	700.0	684.5
Min, Max	230, 1160	170, 1270
GAA1 repeat length >= 675, n (%)	21 (52.5%)	18 (42.9%)
Ambulatory status, n (%)		
Ambulatory	37 (92.5%)	39 (92.9%)
Non-ambulatory	3 (7.5%)	3 (7.1%)
History of cardiomyopathy, n (%)	19 (47.5%)	12 (28.6%)
History of areflexia, n (%)	36 (90.0%)	41 (97.6%)
History of scoliosis, n (%)	29 (72.5%)	32 (76.2%)
History of scoliosis surgery, n (%)	12 (30.0%)	7 (16.7%)

mFARS Data Entry Error Correction Addendum

mFARS	RTA 408 (N=40)	Placebo (N=42)
Mean (SD)	40.95 (10.39)	38.78 (11.03)
Median	39.15	35.70
Min, Max	24.3, 59.3	19.8, 63.0

(Source: Table 9 from study report in SN 1, Page 7 of "408-c-1402-pt2-report-body.pdf" in SN 35)

The baseline mFARS score, and the mean of GAA1 repeat length, were slightly higher in the RTA408 compared to the placebo arm. The percentage of history of cardiomyopathy was also higher in the RTA408 arm at 47.5%, comparing to 28.6% in the placebo arm. Such slight imbalance is not uncommon for a small size trial. Please refer to the clinical review for more discussion.

3.2.4 Results and Conclusions

3.2.4.1 Efficacy Results of the Primary Endpoints

The results reported by the sponsor were confirmed.

The primary analysis of mFARS included 82 patients. There was a statistically significant greater mean reduction in the mFARS score from baseline to week 48 of the 48-week double-blind treatment phase for RTA408 compared with placebo. The difference in least squares mean (LSM) (95% CI) was −2.4 (−4.3, −0.5) for RTA408 vs placebo (p = 0.014).

Table 5 Primary Endpoint Analyses of mFARS – Study 408-C-1402-PT2– FAS population

			mFARS Data Entry Error Correction Addendum	
	RTA408 (N=40)	Placebo (N=42)	RTA408 (N=40)	Placebo (N=42)
Baseline				
N	40	42	40	42
Mean (SD)	40.94 (10.39)	38.77 (11.03)	40.95 (10.39)	38.78 (11.03)
Median	39.15	35.65	39.15	35.70
Minimum, Maximum	24.3, 59.3	19.8, 63.0	24.3, 59.3	19.8, 63.0
Change from baseline at Week 48				
N	34	41	34	41
Mean (SD)	-1.45 (4.21)	0.90 (4.30)	-1.46 (4.21)	0.90 (4.30)
Median	-1.55	1.00	-1.55	1.00
Minimum, Maximum	-11.4, 10.0	-6.3, 10.0	-11.4, 10.0	-6.3, 10.0
Adjusted analysis[a]				
LSM estimate	-1.55	0.85	-1.56	0.85
Difference in LSM	-2.40	---	-2.41	---
(95% CI of difference)	(-4.31, -0.50)	---	(-4.32, -0.51)	---
p-value	0.0141	---	0.0138	---

(Source: reviewer's own analysis and Table 10 from study report, verified by reviewer.)

Adjusted analysis utilized a form of MMRM (Mixed Model Repeated Measures) which includes treatment (drug or placebo), visit (Week 4, 12, 18, 24, 36, and 48), treatment by visit interaction and baseline mFARS by visit interaction as fixed effects, baseline value of mFARS and site as covariates, subject as random effect, and assumed an unstructured covariance structure.

Study SAP was changed 3 times close to study ending date and site was added as covariate in the last version. However, the result is very similar: the model without site as covariate produces a slightly larger p value of 0.0196 (or 0.0193 after the data entry error correction), still under .05. Therefore, we consider the results as reported by the applicant confirmed.

3.2.4.2 Missing Data and Sensitivity Analyses

The overall missing rate was relatively low in this study. The missingness was imbalanced between the two treatment groups: 1 from placebo and 6 from treated group.

Table 6 Patient Counts at Baseline and Week 4,12,18,24,36 and 48 – Study 408-C-1402-PT2– FAS population

	Week 4	Week 12	Week 18	Week 24	Week 36	Week 48	Missing N	Missing %
drug	40	40	38	36	35	34	6	15.00%
placebo	42	42	41	41	41	41	1	2.38%
total	82	82	79	77	76	75	7	8.54%

(Source: Reviewer's own analysis.)

Most missingness (1 from placebo, 3 from drug) was due to AEs. Please refer to clinical review for more details on reasons behind the missing or discontinuations.

Table 7 Patients Missing Week 48 mFARS score – Study 408-C-1402-PT2– FAS population

Trt	Patient	D/C early?	Exposure (wks)	Base line	Change from Baseline [a]						D/C reason	**mFARS Data Entry Error Correction Addendum**
					Wk 4	Wk 12	Wk 18	Wk 24	Wk 36	Wk 48		
Pbo	(b) (6)	Yes	8.1	44.5	-4.5	-4.8*	-	-	-	-	AE	
Omav		Yes	9.1	58	-12.5	-2.3*	-	-	-	-	WITHDRAWAL	Site 1875 affected by data error
Omav		Yes	9.1	58	-13	-2.3*	-	-	-	-	WITHDRAWAL	
		Yes	12.1	30.4	-5.4	-4.7	-	-	-	-	AE	
		Yes	2.4	27.3	-1.1	-5*	-1.6*	-	-	-	AE	
		Yes	20.1	54.7	3.3	7.3	2.3	-	-	-	WITHDRAWAL	
		Yes	18.7	40.3	-2.8	-3.5	-5	-6.1*	-	-	AE	
		No	47.9	46.1	-3.8	2.9	0.2	-8.1	0.9	-	Flight**	

a Change from baseline mFARS available in analysis dataset ADFAIN
* = post-treatment mFARS assessed after drug washout (more than 14 days after date of last dose)
** This subject had to leave to catch flight.
(Source: adapted from Applicant's table and ADAM datasets in SN8, SN 29, verified by reviewer.)

In the sensitivity analyses performed by the sponsor, missing data was imputed using a control-based approach, and treatment-based approach. The primary endpoint was analyzed using the same generalized linear mixed effect model after the imputations. The results of the sensitivity analyses were consistent with those of the primary analysis.

Table 8 Sensitivity Analysis of mFARS – Study 408-C-1402-PT2

Change from baseline mFARS at Week 48	Control-based multiple Imputation		Treatment-based multiple imputation	
	RTA408 (N=40)	Placebo (N=42)	RTA408 (N=40)	Placebo (N=42)
LSM estimate	-1.34	0.80	-1.58	0.79
Difference (95% CI)	-2.14 (-4.18, -0.10)	--- ---	-2.38 (-4.41, -0.35)	--- ---
p-value	0.0395	---	0.0218	---
mFARS Data Entry Error Correction Addendum				
LSM estimate	-1.35	0.80	-1.60	0.79
Difference (95% CI)	-2.15 (-4.18, -0.11)	--- ---	-2.38 (-4.41, -0.35)	--- ---
p-value	0.0386		0.0214	

Source: reviewer's own analysis and Table 14.2.7, Table 14.2.8, Table 14.2.7-cc, Table 14.2.8-cc, verified by the reviewer.

Tipping Point (Let shift parameter go step by step up in small increment to see when the p value will become larger than .05) analysis was also conducted, the 'shift point' was identified at 2 points to change from statistical significance to non-signifcance. "Shift 0" is based on observed data from the randomized treatment group and assumes that data were MAR. The tipping point analysis was summarized in the table below.

Table 9 Tipping Point Analysis – Study 408-C-1402-PT2

Week 48 Shift of Missing Values in Tipping Point Analysis	Reported[a]			Corrected Data Case		
	Week 48 Mean ± SE Change from Baseline (p-value vs zero)		Week 48 Mean ± SE Placebo-Corrected Change (p-value vs Placebo)	Week 48 Mean ± SE Change from Baseline (p-value vs zero)		Week 48 Mean ± SE Placebo-Corrected Change (p-value vs Placebo)
	Placebo (n = 42)	Omaveloxolone (n = 40)		Placebo (n = 42)	Omaveloxolone (n = 40)	
Shift 0	0.79 ± 0.703 (p = 0.2588)	-1.53 ± 0.773 (p = 0.0472)	-2.33 ± 1.057 (p = 0.0276)	0.79 ± 0.703 (p = 0.2609)	-1.55 ± 0.772 (p = 0.0455)	-2.34 ± 1.056 (p = 0.0270)
Shift 1	0.79 ± 0.705 (p = 0.2596)	-1.39 ± 0.775 (p = 0.0740)	-2.18 ± 1.060 (p = 0.0397)	0.79 ± 0.705 (p = 0.2618)	-1.40 ± 0.775 (p = 0.0716)	-2.19 ± 1.060 (p = 0.0390)
Shift 2	0.80 ± 0.710 (p = 0.2620)	-1.24 ± 0.780 (p = 0.1128)	-2.03 ± 1.067 (p = 0.0566)	0.79 ± 0.710 (p = 0.2642)	-1.25 ± 0.780 (p = 0.1095)	-2.04 ± 1.066 (p = 0.0557)
Shift 3	0.80 ± 0.717 (p = 0.2658)	-1.09 ± 0.787 (p = 0.1665)	-1.89 ± 1.076 (p = 0.0797)	0.79 ± 0.717 (p = 0.2681)	-1.10 ± 0.786 (p = 0.1621)	-1.89 ± 1.076 (p = 0.0785)
Shift 4	0.80 ± 0.725 (p = 0.2709)	-0.94 ± 0.795 (p = 0.2373)	-1.74 ± 1.088 (p = 0.1103)	0.79 ± 0.725 (p = 0.2732)	-0.95 ± 0.795 (p = 0.2318)	-1.75 ± 1.088 (p = 0.1087)

Source: adapted from Table 12 CSR and Table 10 in CSR Addendum, verified by the reviewer.

3.2.4.3 Efficacy Results of the Secondary Endpoints

PGIC (Patient Global Impression of Change) and CGIC (Clinician Global Impression of Change) at Week 48

Table 10 Two Key Secondary Endpoints: PGIC and CGIC – Study 1402-PT2-FAS Population

	FAS	
	RTA408 (N=40)	Placebo(N=42)
PGIC		
LSM estimate	3.89	4.32
Difference	-0.43	---
p-value	0.13	---
CGIC		
LSM estimate	3.92	4.06
Difference	-0.13	---
p-value	0.53	---

Source: Table 13 CSR verified by the reviewer.
Comparison of PGIC and CGIC for patients treated with omaveloxolone 150 mg and placebo was estimated using analyses of covariance, with the following fixed factors: site, pes cavus status, and treatment. Missing data were imputed using multiple imputation.

If the primary endpoint was significant, then the two key secondary endpoints were tested hierarchically in the listed order, each at the 0.05 level. Mean PGIC and CGIC scores were not statistically different from placebo. The two key secondary endpoints of PGIC and CGIC at week 48 were not significant ($p > 0.05$), and therefore both and the subsequent secondary endpoints in the hierarchy were not statistically significant.

Other Secondary Endpoints

The other 5 secondary endpoints in the study: 9-HPT, 25-foot timed walk test, frequency of falls, peak work, and ADL scores at week 48, were to be analyzed using a fixed-sequence hierarchical approach after winning on the two key secondary endpoints to maintain the family-wise overall Type I error rate of 0.05. Out of the 5 secondary endpoints, only the ADL score at week 48 had a nominally significant p value of .04, while the other 4 had nominal p values larger than .05.

Table 11 Other Secondary Endpoints - Study 1402-PT2 - FAS Population

Endpoints at Week 48	RTA408 (N=40)	Placebo (N=42)	Difference in LSM	p-value
9-HPT (1/sec) (Analysis based on reciprocal of average time, nondominant hand)	-0.0014	-0.0001	-0.0013	0.18
25-foot timed walk test (1/sec) (Analysis based on reciprocal of average walk time)	-0.0169	-0.0226	0.0058	0.46
Frequency of Falls* (Median)	3.0	8.5	-0.32	0.28
Peak Work (watt/kg)	0.03	0.09	-0.06	0.22
Activities of Daily Living	-0.17	1.14	-1.30	0.04

*Comparison in the frequency of falls for treated patients versus placebo patients was estimated from the Poisson model with the natural logarithm of time on study (days) included as an offset term. Source: Table 16 CSR verified by the reviewer.

3.3 Evaluation of Safety

Please refer to the clinical review for details on safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age and Geographic Region

The demographic characteristics (i.e., age, sex, and race) and region of the FAS population for Study 1402-Pt2 were summarized by treatment group.

Table 12 Age, Gender, Race, Region by Treatment Group- Study 408-C-1402-PT2-FAS population

	Age (Median)	Age (%<18)	Gender (%F)	Race (%White)	Region (%US)
RTA408 (N=40)	23	17.50	60.00	100.00	65.00
Placebo (N=42)	21	30.95	33.33	95.24	66.67
Overall (N=82)	21	24.39	46.34	97.56	65.85

(Reviewer's result.)

The observed treatment effect appears larger in males than females. Because of limited data, this difference may be due to chance. The number of patients of races other than White was very small: 0 in the treated group and 2 in the placebo group. Therefore, the differences in how the drug worked in other races could not be determined. There is virtually no difference in treatment effects by age subgroups as defined.

Table 13 Findings in Subgroup Populations – Study 408-C-1402-PT2-FAS population

mFARS	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)
	Age < 18				Age >= 18			
150 mg	7	44.56 (10.73)	0.05 (-5.55, 5.64)	-1.59 (-8.54, 5.36)	33	40.18 (10.32)	-1.41 (-2.81, 0.00)	-1.75 (-3.82, 0.33)
placebo	13	40.42 (10.31)	1.63 (-2.23, 5.50)		29	38.03 (11.43)	0.34 (-1.11, 1.79)	
	Gender = Female				Gender = Male			
150 mg	24	41.45 (10.21)	-1.01 (-3.04, 1.01)	-1.86 (-5.50, 1.78)	16	40.18 (10.95)	-2.99 (-5.40, -0.58)	-4.00 (-7.06, -0.93)
placebo	14	33.01 (7.97)	0.84 (-1.81, 3.50)		28	41.65 (11.32)	1.01 (-0.73, 2.75)	
	Race = White				Race = Other			
150 mg	40	40.94 (10.39)	-1.57 (-2.92, -0.23)	-2.93 (-4.67, -0.89)	0			
placebo	40	38.89 (10.93)	1.20 (-0.08, -2.49)		2	36.5 (17.68)		
	Region = US				Region = All Other			
150 mg	26	43.50 (10.34)	-2.35 (-3.97, -0.74)	-3.70 (-5.90, -1.50)	14	36.19 (9.00)	-0.24 (-2.79, 2.31)	-0.03 (-3.74, 3.68)
placebo	28	38.88 (12.33)	1.34 (-0.11, 2.80)		14	38.57 (8.23)	-0.21 (-2.72, 2.30)	
	mFARS Data Entry Error Correction Addendum							
mFARS	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)
	Age < 18				Age >= 18			
150 mg	7	44.58 (10.70)	0.02 (-5.59, 5.63)	-1.61 (-8.58, 5.36)	33	40.18 (10.33)	-1.41 (-2.82, 0.00)	-1.75 (-3.82, 0.33)
placebo	13	40.42 (10.31)	1.63 (-2.23, 5.50)		29	38.04 (11.43)	0.33 (-1.11, 1.78)	
	Gender = Female				Gender = Male			
150 mg	24	41.46 (10.20)	-1.02 (-3.04, 0.99)	-1.86 (-5.50, 1.77)	16	40.19 (10.96)	-3.00 (-5.41, -0.59)	-4.01 (-7.06, -0.95)
placebo	14	33.01 (7.97)	0.84 (-1.81, 3.48)		28	41.66 (11.32)	1.00 (-0.74, 2.74)	
	Race = White				Race = Other			
150 mg	40	40.95 (10.39)	-1.58 (-2.93, -0.24)	-2.78 (-4.68, -0.89)	0			
placebo	40	38.89 (10.93)	1.20 (-0.08, -2.48)		2	36.5 (17.68)		
	Region = US				Region = All Other			
150 mg	26	43.52 (10.34)	-2.37 (-3.99, -0.76)	-3.71 (-5.91, -1.52)	14	36.19 (9.00)	-0.24 (-2.79, 2.31)	-0.03 (-3.74, 3.68)
placebo	28	38.88 (12.32)	1.34 (-0.11, 2.79)		14	38.57 (8.23)	-0.21 (-2.72, 2.30)	

(Source: reviewer's analysis.)

We note that the applicant's subgroup analysis added the subgroup variable to the MMRM model as covariate(s) in 4 terms: standalone, 2-way interactions with visit, with trt, and 3-way interaction with visit and trt which did not generate useful information for assessing potential heterogeneity of treatment effects in subgroups while the sample size is so small. Such details and changes to the model specifications for subgroup analyses would need to be described in the study SAP but were not.

For subgroup analysis, we will rerun the exact same MMRM model used in the primary endpoint analysis of the subsets defined by the subgroup variables (age, gender …), and report the estimate as descriptive statistics without p-values, as can be seen in the above table.

4.2 Other Special/Subgroup Populations

Two other subgroups of interest are GAA1 Repeat Length (≥675: Yes, No) and history of cardiomyopathy (Y, N). The observed treatment effect appears larger in subjects with GAA1 repeat length >=675 than those < 675. The treatment effect appears to be slightly larger in subjects with no history of cardiomyopathy. Because of limited data, these differences may be due to chance.

Table 14 GAA1 and Cardiomyopathy by Treatment Group- Study 408-C-1402-PT2-FAS population

	GAA1 Repeat Length (%>=675)	History of Cardiomyopathy (%Y)
RTA408 (N=40)	67.74	47.50
Placebo (N=42)	50.00	28.57
Overall (N=82)	58.21	37.80

(Reviewer's result.)

Table 15 Findings in Other Subgroup Populations – Study 408-C-1402-PT2-FAS population

mFARS	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)
	GAA1 repeat length >= 675				GAA1 repeat length < 675			
150 mg	21	44.26 (10.34)	-2.32 (-4.37, -0.27)	-4.34 (-7.43, -1.24)	10	39.80 (10.35)	-1.18 (-4.53, 2.17)	-1.78 (-6.08, 2.51)
placebo	18	42.46 (12.41)	2.02 (-0.13, 4.17)		18	34.97 (9.70)	0.60 (-1.80, 3.00)	
	History of cardiomyopathy =Y				History of cardiomyopathy = N			
150 mg	19	44.26 (10.37)	0.11 (-2.45, 2.68)	-2.04 (-6.40, 2.33)	21	37.94 (9.68)	-2.85 (-4.56, -1.14)	-3.07 (-5.34, -0.79)
placebo	12	38.97 (10.65)	2.15 (-0.97, 5.27)		30	38.70 (11.35)	0.21 (-1.17, 1.59)	
	mFARS Data Entry Error Correction Addendum							
mFARS	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)	N	Baseline Mean (SD)	Change at Week 48 (95% CI)	Diff at week 48 (95% CI)
	GAA1 repeat length >= 675				GAA1 repeat length < 675			
150 mg	21	44.27 (10.35)	-2.32 (-4.36, -0.28)	-4.34 (-7.42, -1.25)	10	39.82 (10.35)	-1.22 (-4.56, 2.12)	-1.82 (-6.10, 2.46)
placebo	18	42.46 (12.41)	2.02 (-0.13, 4.16)		18	34.98 (9.70)	0.60 (-1.79, 2.99)	
	History of cardiomyopathy =Y				History of cardiomyopathy = N			
150 mg	19	44.28 (10.37)	0.09 (-2.47, 2.65)	-2.06 (-6.42, 2.31)	21	37.94 (9.68)	-2.86 (-4.56, -1.15)	-3.06 (-5.33, -0.80)
placebo	12	38.97 (10.65)	2.15 (-0.97, 5.26)		30	38.70 (11.35)	0.21 (-1.17, 1.58)	

(Reviewer's result.)

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

No major statistical issues were identified.

There was only one single pivotal study to review. This study provided statistical evidence to support that RTA408 at 150 mg is effective in treating patients with FA based on the primary endpoint mFARS.

To show additional support of the single pivotal trial, the applicant has included 1 ref paper of a natural history study and submitted 1 post-hoc analysis of global test of multiple endpoints, as described below.

Natural history study data Patel [2016]

There is only one single natural history study with its result summarized in Patel [2016].

Patel [2016] study collected data over a 5-year period. The study cohort had a noticeable amount of missing data. The number of subjects with demographic data decline from 820 at Year 0 to 597, 479, 405, 352 and 290 at Year 1, 2, 3, 4 and 5. The missing rate of mFARS score which requires a clinical visit, is likely higher, although the details are not given in the ref paper.

The average of disease progression when limiting to the same (or similar) patient population in the trial (age 16 to 40, and other I/E criteria) and same duration 48 weeks (approx. 1 year) was between 1 to 2 points in Patel [2016]. We noted that the “pes cavus” status was not collected in the natural history study while it was a part of the definition for the FAS population in the pivotal trial (408-C-1402-PT2).

The result of the pivotal trial showed a 2.4 points advantage of Omav over placebo, which is larger than the 2 points per year, while the placebo showed 0.85 points worsening over 48 weeks period, which is approximately close to the average of 1 to 2 points disease progression per year. We also noted that the assessment window for the clinical visit for mFARS score in the natural history study was likely much wider than the pivotal trial.

The tipping point shift in the pivotal trial was identified at 2 points, which is within the range of 1 to 2 points per year disease progression if left untreated.

In sum, the primary endpoint result of the pivotal trial is statistically significant with $p = 0.0141$. Given the well-known possible issues with use of natural history data, we do not see that Patel [2016] with heavy missing data could provide additional strong support.

Post-hoc Global Test of Combining Endpoint(s)

The results of several global tests were submitted in SN 32 in September 2022, after the mFARS data entry error related submissions SN 22, 23, and 29, later than the planned end of July 2022 as noted in the mid-cycle meeting slides.

In general, global testing combines a set of endpoints into one statistical testing.

By combining the primary endpoint and secondary endpoints and analyzing the sum of ranks, the applicant's post-hoc global testing analyses show a smaller p value of 0.0124 for the primary endpoint, while the pre-specified MMRM model analysis p value was .0141 (after data error correction it should be .0138). The differences in these p-values are very small if any.

Among the secondary endpoints, the 5th other secondary endpoint (i.e., FA-ADL) is only one showing a nominal p-value = 0.0420. The applicant's further post-hoc global test combining the primary and this secondary endpoint shows a p value of 0.0041.

We have concerns with such global testing. First, all these analyses are post-hoc and do not follow the pre-specified multiplicity adjustment procedure. It is difficult to interpret such post hoc results because of the post hoc nature after looking at the trends to decide what and how to combine for global testing. Second, combining endpoints trending in the same direction tends to yield a smaller p-value. The null hypothesis tested under such global testing is not endpoint-specific hypothesis.

This kind of p-value for multiple null hypotheses considered jointly is not a measure of strength of evidence of individual endpoint-specific null hypothesis.

It does not control the overall type I error probability associated with testing the pre-specified set of endpoints (FDA Guidance for Industry, Multiple Endpoints in Clinical Trials https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry).

Such a global testing result adds little to help interpretation for strength of statistical evidence.

5.2 Conclusions and Recommendation

The efficacy results obtained from the statistical analyses of the one pivotal study (408-C-1402-PT2) support the conclusion that RTA408 at 150 mg is effective in treating patients with episodic FA, based on change from baseline in the primary endpoint mFARS at Week 48.

5.3 Labeling Recommendation

Table 16 Efficacy Result for Study 408-C-1402-PT2 – FAS population

			mFARS Data Entry Error Correction Addendum	
Change from Baseline mFARS score at Week 48	RTA408 (N=42)	Placebo (N=40)	RTA408 (N=42)	Placebo (N=40)
MMRM Model Adjusted Analyses*				
LSM estimate	-1.55	0.85	-1.56	0.85
Difference in LSM (95% CI)	-2.40 (-4.31, -0.50)	---	-2.41 (-4.32, -0.51)	---
p-value	0.0141	---	0.0138	---

(Source: Table 10 CSR and Table 8 CSR Addendum, verified by reviewer.)

Adjusted analysis utilized a form of MMRM (Mixed Model Repeated Measures) which includes treatment (drug or placebo), visit (Week 4, 12, 18, 24, 36, and 48), treatment by visit interaction and baseline mFARS by visit interaction as fixed effects, baseline value of mFARS and site as covariates, subject as random effect, and assumes an unstructured covariance structure.

6 REFERENCES

[1] FDA Guidance for Industry, Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. http://www fda.gov/cder/guidance/1397fnl.pdf

[2] FDA Guidance for Industry, Multiple Endpoints in Clinical Trials https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry

APPENDIX – Data Entry Error

Data entry error regarding primary endpoint (mFRAS) was discovered late in the review, although it did not alter the efficacy conclusion and the impact was minimal. The data entry error occurred in a small number of subjects in both arms, slightly over stating the patient's condition measured by section E Upright Stability of the mFARS assessment by averaging total score over 3 instead of 2 because the third time was marked as "Ø" and mistakenly entered as a score "0" when the "Ø" meant the third measurement was not performed.

More details about the error:

> The specific subparts of the mFARS Section E Upright Stability exam: Section E Upright Stability subparts 2.a, 2.b, 3.a, and 3.b, required the investigator to make 3 assessments of the time the patient could stand if the patient could not stand for at least one minute on the first try. One clinical site did not consistently perform the third timed trial when the patient could not stand for at least one minute on the first trial but could stand for at least one minute on the second trial. The investigator accurately noted when the third trial was not attempted, but the site sometimes incorrectly recorded the third trial value as "0" rather than missing. This specific pattern of the error is referred to as the "#-0-0 pattern." This caused a small error in the computation of the mFARS score because the average value for the item was computed using a 3 in the denominator rather than a 2.

(Source: Page 3 applicant's "data-entry-error-report-408-c-1402.pdf" in SN 23)

Here's a list of subject IDs with visits at which the data entry error occurred:

Treatment Group	Analysis Population	Subject # (n=11)	Visit # (n=13)	Section E – Upright Stability Data Item # (n=14)	Source Doc Value Trial 1-Trial 2-Trial 3	eCRF Entry Trial 1-Trial 2-Trial 3
Omaveloxolone	ARP	(b) (6)	Week 24	2a	1-0-NA	1-0-0
			Screening	3a	2-0-NA	2-0-0
			Week 12	3a	1-0-NA	1-0-0
			Week 36	2a	3-0-NA [2]	3-0-0
	ARP, FAS		Day 1	3a	2-0-NA	2-0-0
			Week 18	2a	3-0-NA	3-0-0
			Week 4	2a	3-0-NA	3-0-0
			Day 1	2a	3-0-NA	3-0-0
Placebo	ARP		Week 24	3b	1-0-NA	1-0-0
	ARP, FAS		Week 36	3b	2-0-NA	2-0-0
			Day 1	3a	2-0-NA [1]	2-0-0
			Week 4	2b	2-0-NA	2-0-0
				3a	2-0-NA	2-0-0
			Week 18	2a	3-0-NA	3-0-0

Abbreviations: ARP=All Randomized Population; FAS=Full Analysis Set

[1] Contradiction between the score and the related time on the source document - third score has a "0" value but time is completed as "ø"

[2] Contradiction between the score and the related time on the source document - third score has a "0" value but time is completed as "Not Done"

(Source: Page 18 applicant's "data-entry-error-report-408-c-1402.pdf" in SN 23)

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/

JINNAN LIU
02/02/2023 03:40:32 PM

JOHN P LAWRENCE
02/02/2023 03:58:36 PM

HSIEN MING J HUNG
02/02/2023 04:06:14 PM