CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

761047Orig1s000

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

ADDENDUM

NDA/BLA #:	BLA 761-047
Drug Name:	Mepsevii (UX003, Recombinant Human Beta-Glucuronidase, rhGUS) Solution, Liquid
Indication(s):	Treatment of Mucopolysaccharidosis type VII (MPS VII, Sly syndrome)
Applicant:	Ultragenyx Pharmaceutical Inc.
Date(s):	Date Submitted: March 16, 2017 PDUFA Due Date: November 16, 2017
Review Priority:	Priority
Biometrics Division:	DBIII
Statistical Reviewers:	Feiran Jiao, Ph.D. Min Min, Ph.D.
Concurring Reviewers:	Yeh-Fong Chen, Ph.D.
Medical Division:	Division of Gastrointestinal and Inborn Error Products (DGIEP)
Clinical Team:	Medical Officer: Dina Zand, M.D. Medical Team Leader: Katie Donohue, M.D.
Project Manager:	Jenny Doan, MSN

In the CRF documentation (page 197) of the BLA application, for Subject (b) (6), at 8 week of UX003 (study week 24) limited walking without a walker was observed although she could walk fast with a walker. This subject used a mobility aide (specifically a walker) all the time after study week 24. We imputed zero for this subject's six minute walk distance (i.e., 6MWD) at 8 weeks of UX003 treatment in the ANCOVA (Table 6) for being conservative in statistical review, filed on 10/23/2017. During the labeling negotiation with the applicant, we were informed that this subject did not use walking aid during her actual 6MWT test (stated in CFR page 180) at 8 week of UX003. The applicant's ANCOVA analysis was shown in Table 1 and confirmed by the statistical reviewer. Based on Table 1 below and Table 6 in statistical review, now the mean difference between UX003 period and placebo period becomes larger (LS mean [SE]: -11 [24] vs. -21 [27]); accordingly, the overall improvement over time from 8 weeks to 16 weeks was decreased (34m to 24m).

Table 1: 6MWT (m): Least Square Mean Difference between UX003 Period and Placebo Period (No imputation for Subject (b) (6))

Treatment Duration	LS mean (SE)	P-value	95% Confidence Interval
8 weeks treatment	-11.07 (23.93)	0.66	(-69.6, 47.5)
16 weeks treatment	13.07 (31.68)	0.69	(-64.5, 90.6)
24 weeks treatment	17.74 (33.36)	0.61	(-63.9, 99.4)

Feiran Jiao, Ph.D.
Mathematical Statistician

cc: BLA 761-047

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

/s/

FEIRAN JIAO
11/01/2017

MIN MIN
11/01/2017

YEH FONG CHEN
11/02/2017

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 #:	BLA 761-047
Drug Name:	Mepsevii (UX003, Recombinant Human Beta-Glucuronidase, rhGUS) Solution, Liquid
Indication(s):	Treatment of Mucopolysaccharidosis type VII (MPS VII, Sly syndrome)
Applicant:	Ultragenyx Pharmaceutical Inc.
Date(s):	Date Submitted: March 16, 2017 PDUFA Due Date: November 16, 2017
Review Priority:	Priority
Biometrics Division:	DBIII
Statistical Reviewers:	Feiran Jiao, Ph.D. Min Min, Ph.D.
Concurring Reviewers:	Yeh-Fong Chen, Ph.D.
Medical Division:	Division of Gastrointestinal and Inborn Error Products (DGIEP)
Clinical Team:	Medical Officer: Dina Zand, M.D. Medical Team Leader: Katie Donohue, M.D.
Project Manager:	Jenny Doan, MSN

Keywords: BLA Review, Rare Disease, Blind-Start, Single-Crossover

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Ultragenyx Pharmaceutical has developed Mepsevii (UX003, rhGUS) as an intravenous (IV) enzyme replacement therapy (ERT) for the treatment of Mucopolysaccharidosis (MPS) type VII disease (Sly syndrome). To demonstrate the efficacy and safety of UX003 in treating patients with MPS VII, the applicant submitted one Phase 1/2 dose-finding study (CL201), one phase 2 study (CL203) in MPS VII patients under 5-year-old, and one randomized phase 3 placebo-controlled, blind start and single crossover (otherwise known as a delayed start design) study (CL301). In this review, our focus is the efficacy evaluation of UX003 from Study CL301.

Study CL301 contained 12 patients with MPS VII between the ages of 8 and 25 years (median age 14 years) who received UX003 4 mg/kg every two weeks for 24 to 48 weeks. Due to the heterogeneity of the patient population, efficacy of UX003 in treating MPS VII patients was based on the totality of the clinical evidence. No primary endpoint was declared beforehand. During an IND meeting for this clinical development program, (b) (4) Among the 6 key secondary endpoints included in a multi-domain clinical responder index (MDRI), the clinical team determined that only the Six Minute Walk Test (6MWT) after 24 weeks of treatment is suitable for efficacy evaluation. It is important to note that the applicant claimed UX003's efficacy solely through the comparison of patients' improvement from their baseline (i.e., placebo period prior to cross over) and they did not assess drug effect across cohorts which takes the true advantage of placebo data.

The statistical reviewers thoroughly evaluated the study data and the sponsor's supportive evidence for UX003's efficacy in this application. The applicant's analysis results for uGAG (the primary endpoint in the EU and rest of the world) were confirmed, as well as all results for the key secondary endpoints included in the MDRI (six-minute walk test (6MWT), forced vital capacity (FVC), shoulder flexion, visual acuity, Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) fine motor and gross motor). As previously noted, the applicant did not assess drug effect across cohorts which takes the true advantage of placebo data. In order to properly assess the UX003's effect, the statistical reviewer performed additional superiority analyses (i.e., ANCOVA, and permutation tests) to evaluate the superiority of UX003 period over placebo period using the endpoint of the change from baseline after 24 weeks of treatment in 6MWT.

According to the reviewer's thorough evaluation and the applicant's confirmation of the reviewer's results, no substantial evidence is demonstrated supporting the efficacy of UX003 in treating patients with MPS VII, i.e., Sly syndrome. We note, however, UX003's effect appears to increase gradually over time. While many other analyses could be performed, none will overcome the small sample size and lack of assessments in some patients for some of the outcomes.

2 INTRODUCTION

2.1 Overview

The applicant, Ultragenyx Pharmaceutical, has developed Mepsevii (Recombinant Human Beta-Glucuronidase[rhGUS], UX003) as an intravenous (IV) enzyme replacement therapy (ERT) for the treatment of Mucopolysaccharidoses (MPS) VII. MPS are a group of inherited lysosomal storage disorders caused by the deficiency of any of the enzymes involved in the stepwise degradation of glycosaminoglycans (GAGs). MPS VII (MPS 7, Sly syndrome), one of the least common forms of MPS, is estimated to occur in fewer than one in 250,000 and less than 100 living patients identified worldwide. MPS VII is caused by deficiency of one specific enzyme, beta-glucuronidase (GUS). Most MPS VII patients die before the second or third decade of life due to complicating medical problems. The presence and severity of symptoms in MPS VII patients are highly variable. Currently there is no approved treatment for MPS VII.

Based on the following summary of regulatory history, a novel blind-start design has never been agreed upon by the Agency. In addition, the Agency had major concerns regarding (b) (4) impact of missing data and absence of rigorously collected natural history data to ensure a reliable comparator.

- (b) (4)

- On August 13 2014, the applicant submitted a request for Type B meeting and the Agency recommended the following:
  "Previous ERTs have shown efficacy in MPS diseases in a timeframe of 6 months and we strongly recommend a pathway for registration that targets full approval using primary clinical endpoints for your registration trial, (b) (4) *."*
  "The Agency suggested that the individual clinical endpoints would be defined for each patient, which could include one or two endpoints, such as but not limited to 6MWT or FVC, an alternative endpoint could be individually selected that is appropriate for the patient. The Agency agrees that a pre-specified minimally important difference (MID) is important in assessing a clinically meaningful response. The sponsor should provide justification for the MID that is selected for each endpoint used in the trial (MID for the endpoint of visual acuity should be a change in 3 lines, with corrected vision)."

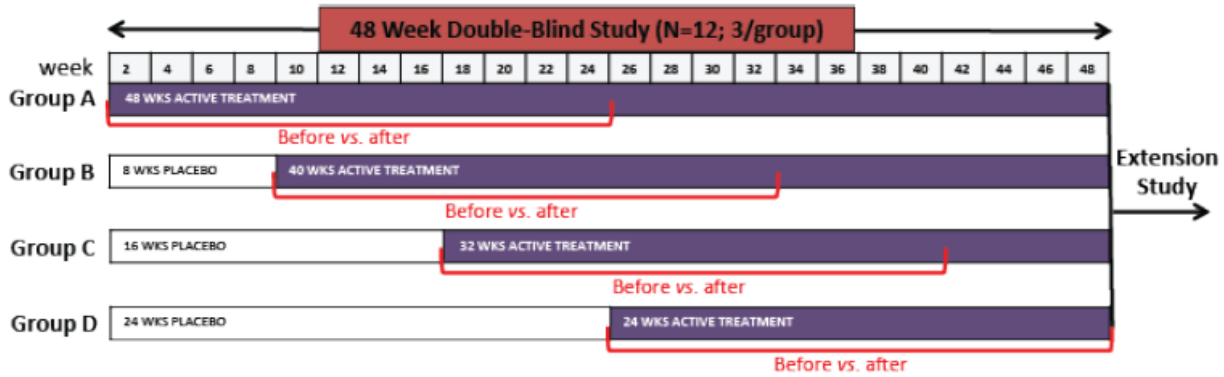
"We cannot agree with the study design at this time. The Blind Start trial design may be appropriate in certain circumstances, such as studies with objective endpoints like laboratory measurements. However, we need to have further discussion regarding study endpoints before we can agree on the study design."

- At the pre-BLA meeting (10/14/2016), the Agency stated the following:
  "We have reservations as to whether you have provided enough evidence of effectiveness to allow for a meaningful review. The absence of rigorously collected natural history data to ensure a reliable comparator for efficacy analyses, combined with very heterogeneous clinical manifestations expected and observed in this small cohort, raise serious challenges in assessing and demonstrating a treatment effect to support definitive efficacy conclusions for your product."
  "We are concerned about missing data, which will significantly affect efficacy analyses."

This application includes data for a phase 3, multicenter, randomized, placebo-controlled, blind-start, single-crossover study (Study CL301), an open-label phase 1/2 study (Study CL201) and an ongoing open-label phase 2 study (Study CL203). Study CL301 is the pivotal efficacy and safety study that assessed the efficacy and safety of UX003 in 12 MPS VII subject aged 8 to 25 years old who received UX003 4mg/kg every two weeks for 24 to 48 weeks. Since Study CL301 is not a traditional placebo-controlled study, in the rest of this review we will not keep this phase. Study CL201 is the dose-finding study conducted and assessed safety, dose, efficacy and pharmacokinetics (PK) of UX003 in 3 MPS VII subjects aged 5 to 30 years old. Study CL203 was conducted to gain additional safety and efficacy data from 4 younger MPS VII subjects (less than 5 years of age) who were not eligible to participate in the phase 3 study.

An overview of the relevant clinical studies is presented in Table 1; and Study CL301 study design is illustrated in Figure 1. This review focuses on Study 301 for evaluating efficacy of UX003 in MPS VII patients.

Figure 1: Study UX003-CL301 Study Design



Source: Applicant's Figure 2.1 of Report Body.pdf of Study UX003-CL301

Table 1: Relevant Clinical Studies

Study ID	Phase and Design	Study Population	Dose	N	Duration	Centers
CL301	Phase 3 study: Randomized blind-start, single-crossover design	MPS VII Aged 5-35 years	UX003 QOW 4mg/kg by IV	12 (3 per arm)	24-48 weeks	4 centers United States
CL201	Phase 1/2 study: Open-label, uncontrolled	MPS VII Aged 5-30 years	UX003 QOW IV Serial dose escalation: 2 mg/kg (14 wks) → 1 mg/kg (8 wks) → 4 mg/kg (8 wks) → 2 mg/kg (44 wks) → 4 mg/kg (up to 168 wks)	3	Up to 240 weeks	3 centers UK, (b) (6)
CL203	Phase 2 study: Open-label, uncontrolled	MPS VII < 5 years	UX003 QOW 4 mg/kg by IV	4	Up to 240 weeks	3 centers US, (b) (6)

2.2 Data Sources

This electronic submission was conformed in the eCTD format. The applicant's original BLA submission, including the data sets, is stored at the following link: \\cdsesub1\evsprod\BLA761047\0001\m5. During the review, the Agency sent two major information requests to the applicant for additional analyses. Please refer to Table 9 in Appendix for details.

The extension data sets for Study CL202, submitted on 8/10/2017 were stored at the following link:\\cdsesub1\evsprod\BLA761047\0039\m5.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The statistical reviewer was satisfied with the data quality of this submission. The applicant's analysis results for the phase 3 pivotal study CL301 can be confirmed. In terms of the analyses, the reviewer found a potential issue is the impact of missing data. There were some non-assessable clinical outcomes at baseline identified and listed below for each endpoint:

- 6 Minute Walk Test (6MWT): 3 subjects could not walk at baseline
- Forced vital capacity (FVC): 9 subjects were too young or cognitively impaired to perform this complex test at baseline
- Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) fine motor: 1 subject was too cognitively impaired to perform the fine motor tasks at baseline
- BOT-2 gross motor: one subject was too cognitively impaired and 4 subjects were too physically impaired to complete the gross motor tasks at baseline
- Visual acuity: 5 subjects could not understand the eye chart test at baseline

3.2 Evaluation of Efficacy

The following sections contain the study description, where mostly was extracted directly from the applicant's Clinical Study Report (CSR).

3.2.1 Study Objectives and Study Design

Study CL301 is titled as "A multicenter, randomized, Blind Start, single crossover Phase 3 study to assess the efficacy and safety of 4 mg/kg UX003 in MPS VII subjects aged 5 to 35 years". It was conducted from 02 Dec 2014 through 04 May 2016 in the United States. Kakkis and Signorovitch (2014) proposed a new study design, a randomized Blind Start trial, to evaluate clinical outcomes in rare, heterogeneous patient population, in which the traditional study design has critical limitations The Blind Start was named due to the fact that neither the subjects nor investigators know when the subject began UX003 therapy. The subjects and investigators also did not know when the subject completed 24 weeks of UX003 therapy.

Figure 1 provides a schematic of the overall study design of CL301. 12 MPS VII subjects enrolled in the study were randomized 1:1:1:1 to one of four groups (Group A, B, C or D; 3 subjects per group) with no stratification:

- Group A: UX003 treatment for 48 weeks
- Group B: placebo treatment for 8 weeks, then crossover to UX003 for 40 weeks
- Group C: placebo treatment for 16 weeks, then crossover to UX003 for 32 weeks
- Group D: placebo treatment for 24 weeks, then crossover to UX003 for 24 weeks

Subjects were dosed every other week (QOW) through Week 46. All groups received a minimum of 24 weeks of UX003 treatment with 4mg/kg QOW. Group assignments were coded and blinded to the applicant, investigators, observers and subjects.

The objectives of Study CL301 are listed in the Table 2.

Table 2: List of Study Objectives of Study CL301

In the US	In the EU & rest of the world	Study Objective
Primary		Efficacy of UX003 in MPS VII subjects based on the totality of the clinical data on a per subject basis. ***No primary endpoint was declared.***
Secondary	**Primary**	Efficacy of UX003 in MPS VII subjects as determined by the percent reduction of uGAG excretion after 24 weeks of treatment relative to the pre-treatment baseline.
Secondary	Secondary	Safety and tolerability following up to 48 weeks of UX003 exposure
Key Secondary	Secondary	Efficacy of UX003 in MPS VII subjects as measured by a multi-domain clinical responder index (MDRI) following 24 weeks of UX003 exposure
Key Secondary	Secondary	Efficacy following 24 weeks of treatment as determined by the proportion of subjects achieving a positive individualized clinical response (ICR) outcome

Key Secondary	Secondary	Clinical effects including pulmonary function, walking distance, shoulder flexion, fine motor function, and gross motor function following 24 weeks of UX003 exposure

The primary efficacy endpoint in the EU and rest of world is the specific uGAG analyte (dermatan sulfate [DS] derived fragments) measured by LC-MS/MS, percent reduction from baseline over 24 weeks of UX003 treatment.

Several key secondary efficacy endpoints were claimed in the study: MDRI (6MWT, FVC, shoulder flexion, visual acuity, and BOT-2 fine and gross motor); pulmonary function testing; 6MWT; shoulder flexion maximum range of motion (goniometry); visual acuity (corrected and uncorrected); BOT-2 (test of motor proficiency); fatigue (Pediatric Quality of Life Multidimensional Fatigue Scale [PedsQL-Multidimensional Fatigue Scale]); ICR (a single clinical measure with the highest impact on the individual subject), and scoring of impactful clinical problems.

Other endpoints included in the study were additional GAG measures (chondroitin sulfate [CS] and heparan sulfate [HS] uGAG by LC-MS/MS; and non-reducing end [NRE] CS/DS and HS GAG in serum and urine); hepatosplenomegaly; cardiac ventricular mass; 3-Minute Stair Climb Test (3MSCT); 2-minute Walk Test (2MWT); growth; subject-reported disability and pain: The MPS Health Assessment Questionnaire (MPS HAQ), and Childhood Health Assessment Questionnaire (CHAQ) or PROMIS Health Assessment Questionnaire (PROMIS HAQ); Physician and Subject/Parent/Caregiver Clinical Global Impression (CGI) scales.

Reviewer's Note:
For rare, heterogeneous patient populations traditional study designs have critical limitations including 1) there are too few patients who share similar characteristics available to perform traditional randomized control trials; 2) power is limited to assess the clinical outcomes due to the small sample size. FDA indicated during the type B, pre-IND meeting held on July 29, 2014 that (b) (4) .
The clinical reviewer also indicated that due to the degree of heterogeneity of presentation and progression of disease, not all endpoints within the MDRI are appropriate to assess efficacy of UX003 in MPS VII (i.e. FVC, visual acuity and shoulder flexion). Therefore, in assessing efficacy for Study CL301, interpretable endpoints are 6MWT, BOT-2 balance, and BOT-2 running speed and agility. Of note, the 6MWT or 12 MWT was previously used to demonstrate efficacy for other ERTs in treating patients with MPS I (laronidase), MPS II (idursulfase), MPS IVA (elosulfase alfa) and MPS VI (the only one by 12 MWT).

3.2.2 Statistical Methodologies

All efficacy analyses were based on the full analysis set, which was defined as all subjects who were randomized and received at least one dose of study drug or placebo.

For the primary endpoint in the EU and the rest of world the applicant used Generalized Estimating Equation (GEE) modeling to analyze the mean percent reduction in uGAG excretion (DS) over 24 weeks of treatment for significant reduction from the baseline (defined as the average of all assessments prior to beginning UX003 treatment). The following changes were pooled across the following four study Groups:

- Group A: change from Study Weeks 0 to 24
- Group B: change from Study Weeks 8 to 32
- Group C: change from Study Weeks 16 to 40
- Group D: change from Study Weeks 24 to 48

The null hypothesis of no change in mean uGAG (DS) percent reduction at 24 weeks of treatment was tested. Statistical significance was assessed at the two sided 5% level. Of note, the applicant did not assess drug effect across cohorts.

Among the key secondary endpoints, the MDRI was the first one to be tested at the two sided 0.05 level. The applicant used a t-test to detect the difference of the mean change in the MDRI between pretreatment baseline and 24 weeks of treatment. For each of the six domains in MDRI, the applicant pre-defined Minimal Important Differences (MID) described in Table 3. If the change is equal or greater than one MID, the score of +1 is provided. If changed negatively by one MID or greater, the score is -1. If there is no clinically significant change (>-1.0 to <1.0 MID), the score is 0. However, as stated in 8/13/2014 meeting minutes, the Agency expressed a concern of using the MDRI as the primary or key secondary clinical endpoint.

Table 3: Definition of the Minimal Important Difference

Domain	Minimal important difference (MID)
6MWT	23 meter AND 10% change from baseline
$FVC_{\%pred}$	5% absolute change or 10% relative change from baseline in $FVC_{\%pred}$
Shoulder flexion	20 degree change of passive shoulder range of motion
Visual acuity	3 lines (corrected, both eyes)
BOT-2 fine motor	Fine Motor Precision: change of 0.72 Manual Dexterity: change of 1.47
BOT-2 gross motor	Balance: 0.57 Running speed and agility:0.59

Source: Applicant's Table 3.9.1 in statistical-methods.pdf of Study UX003-CL301

The applicant also analyzed other repeated continuous secondary efficacy variables (i.e. pulmonary function testing, 6MWT, shoulder flexion maximum range of motion, visual acuity, BOT-2 motor, and fatigue) using GEE modeling.

GEE modeling was the primary analysis method for all repeated measures endpoints. In general, missing data were not replaced unless otherwise specified.

Reviewer's Note:
The applicant did not assess drug effect across cohorts which takes true advantage of placebo data. This reviewer conducted ANCOVA analyses to evaluate the change from baseline of UX003 period over that of placebo period in 6MWT at different treatment durations (8 weeks, 16 weeks, and 24 weeks); the covariates include a categorical group cohort (Group A, B, C or D), patients' baseline 6MWT, and age. This reviewer noted that the patients' age range in this study population was big and that the mean ages were very different across the four group cohorts, as well as their baseline 6MWT values. Age and baseline 6MWT are included as continuous covariates in the analysis model. Due to the small sample size of Study CL301, in order to ensure the robustness of ANCOVA analysis results, permutation tests were further conducted to evaluate the significance of the treatment effect. One major issue we identified was that 9 out of the total 12 patients (Group B, C, and D patients) contributed both placebo and UX003 data and thus the final analysis should take into account of the cohort effect to improve the variance estimation.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

The MPS VII population is extremely limited worldwide. A total of 12 subjects were randomly assigned to each of the four groups and completed the 48-week study. There were no deaths or AEs leading to treatment discontinuation and no withdrawals from study. There were two screen failures, and in both cases the reason was withdrawal of consent prior to first treatment.

A total of 99% and 96% of the planned UX003 and placebo volumes, respectively, were infused during the study. There were only 2 missed doses during the study: Subject (b) (6) (Group A) missed a UX003 infusion at Study Week 42 and Subject (b) (6) (Group C) missed a placebo infusion at Study Week 6.

Of the 12 subjects enrolled in the study, 4 were male and 8 were female; 9 were pediatric subjects (< 18 years old) and 3 were adult subjects (Listing 16.2.4.1). The age range of the subjects was 8 to 25 years of age (median age 14.0 years). The majority of the subjects were white (75%) and 50% were of Hispanic ethnicity. Four subjects were relocated to the US from (b) (6) (1 subject), (b) (6) (2 subjects), and (b) (6) (1 subject) in order to obtain a sufficient number of subjects to participate in the study.

3.2.4 Applicant's Results and Conclusions

The applicant first provided the following efficacy results of uGAG (DS); after 24 weeks of UX003 treatment, a rapid, marked, and sustained reduction in uGAG (DS) was achieved in all subjects with a least squares (LS) mean change (±SE) of -64.8% (±2.468%) ($p<0.0001$). The applicant claimed that all 12 subjects were indeed responders, defined by a ≥ 50% reduction in uGAG on at least one post-treatment visit.

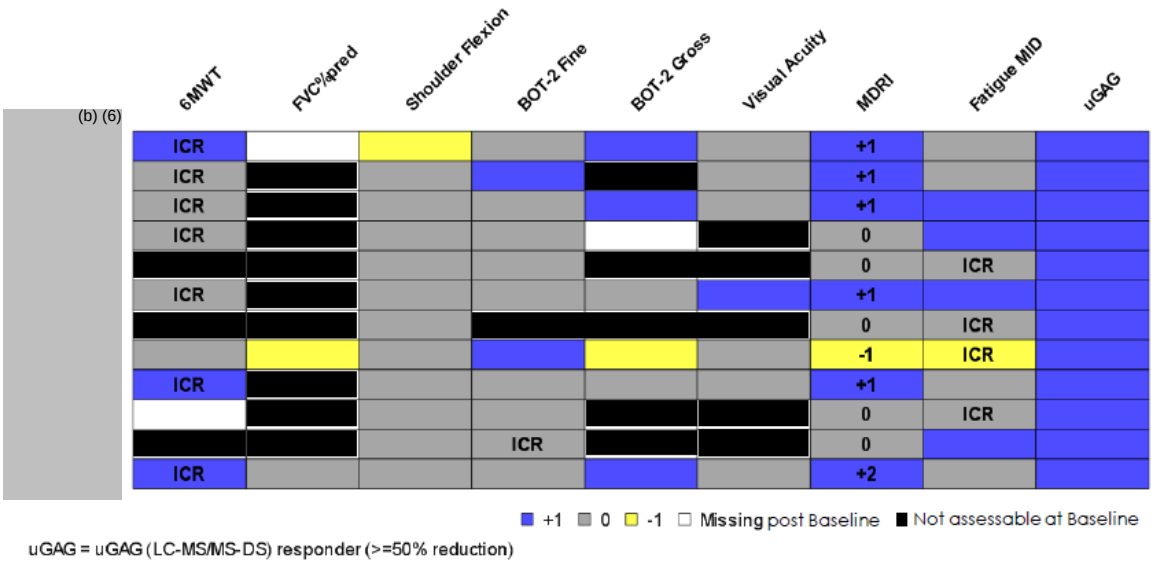
Table 4: Percentage Change from Baseline in uGAG at Week 24

Percent Change from Baseline in uGAG (LC-MS/MS-DS) at Week 24					
Min	Median	Max	LS mean (SE)	95% CI	P-value
-78.1	-61.66	-50.2	-64.8 (2.468)	(-69.66, -59.98)	<.0001

Source: Applicant's Table 14.2.1.1.1.2.1 in the Report Body.pdf of Study UX003-CL301

The applicant analyzed MDRI as the main secondary endpoint. In particular, MDRI had an overall positive mean change (SD) of +0.5 domains (±0.8) at 24 weeks of UX003 treatment (p=0.0527). Figure 2 depicts detailed response information of all 12 subjects for each interested key secondary endpoint. As noted, six of twelve subjects were responders based upon a positive MDRI score.

Figure 2: Endpoints Defined with MID or Response at UX003 Treatment Week 24



Source: Applicant's Figure 10.2.1.1 in the Report Body.pdf of Study UX003-CL301
ICR: Individualized Clinical Response for each subject

Reviewer's Note:

As noted earlier, to assess the drug's efficacy on the proposed endpoints, the applicant only compared individual patients' data of the crossover period to those of the placebo period (prior to cross-over). They did not perform any post-baseline comparison of UX003 period vs. placebo period. It is important to assess the treatment effect of UX003 period over placebo period. Of note, 9 subjects have received both placebo and UX003 treatments. As stated in 08/13/2014 Type B meeting minutes, (b) (4) . For US approval there is no agreed upon uGAG endpoint, one should keep in mind the applicant's uGAG results can be exploratory due to their post-hoc nature.

3.2.5 Statistical Reviewer's Findings and Comments

In this section, efficacy findings are mainly derived from the endpoint 6MWT. Only 9 subjects have data available; 3 subjects could not walk in Study CL301. Of note, Subject (b) (6), (b) (6) and (b) (6) used walking assistive devices when performing 6MWT.

- Subject (b) (6) and (b) (6) used walking assistive devices at all times (Screening visit, baseline visit, Treatment week 24 visit and Study week 48 visit)
- Subject (b) (6) used walking assistive devices after 8 weeks exposure of UX003 (after Study week 24)

Simply analyzing data using the available 6MWT value for all subjects may lead to biased results. This reviewer imputes 6MWT value as 0 for subjects who used walking assistive devices at certain visit.

1. **(Patient profile)** Since 3 subjects could not walk, the following figure shows the patients profile for only 9 subjects. No imputation for subjects who used walking assistive devices is done in the plot.

Figure 3: Patients Profile for 6MWT

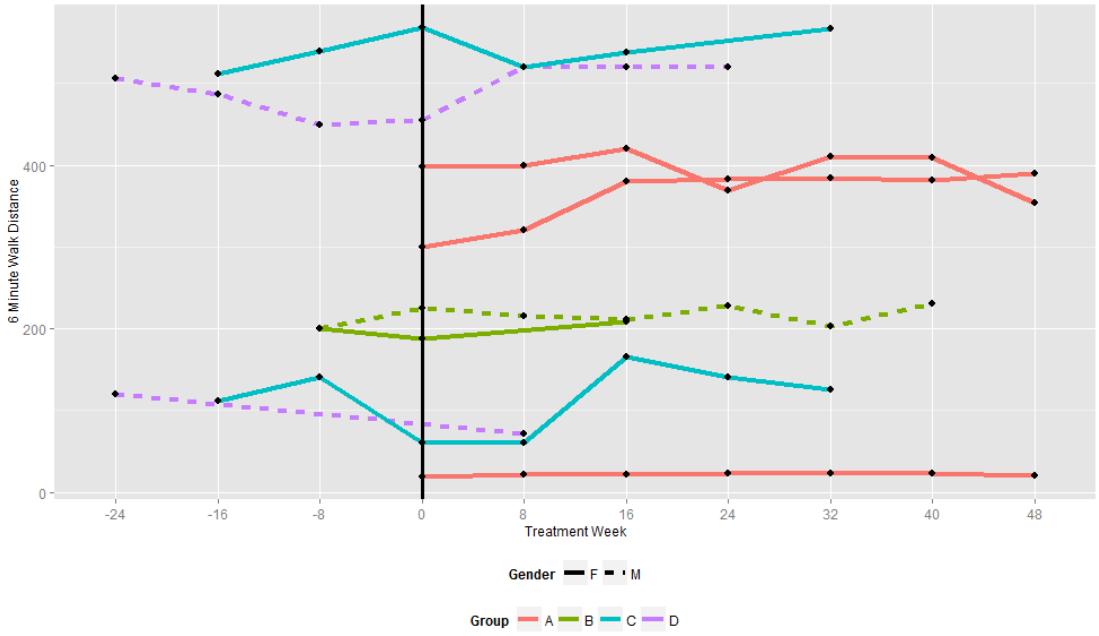
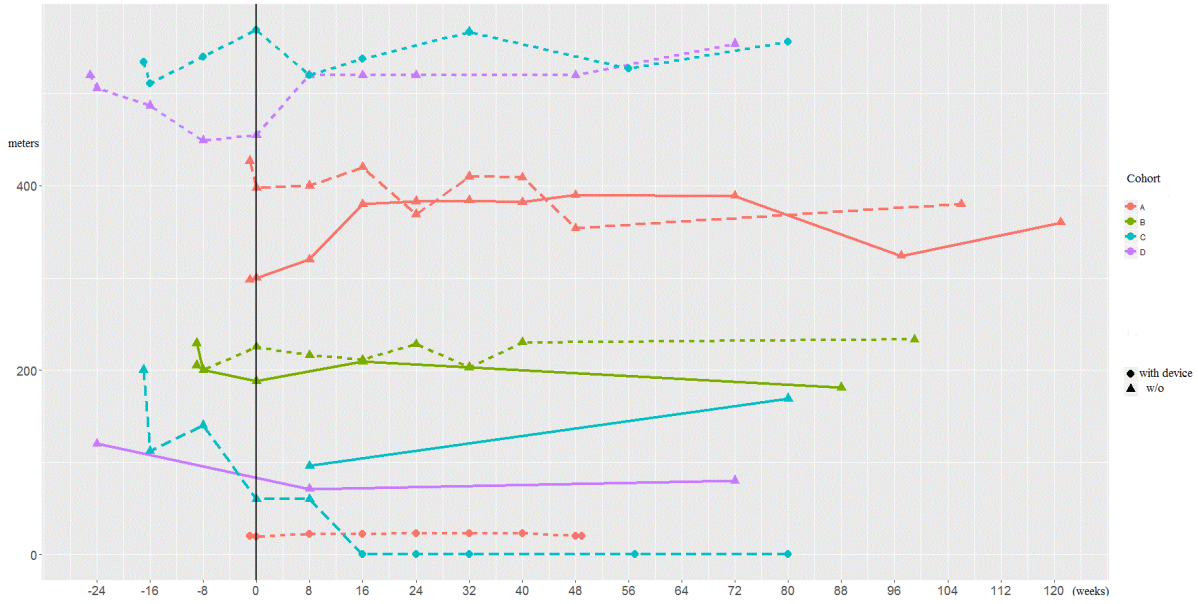


Figure 4 imputes 6MWT values as zeros for subjects who used walking assistive devices at certain visit time point.

Figure 4: Patients Profile for 6MWT (Impute 0 for subjects used walking aid for Subject (b) (6))



Placebo Period UX003 Period (weeks)

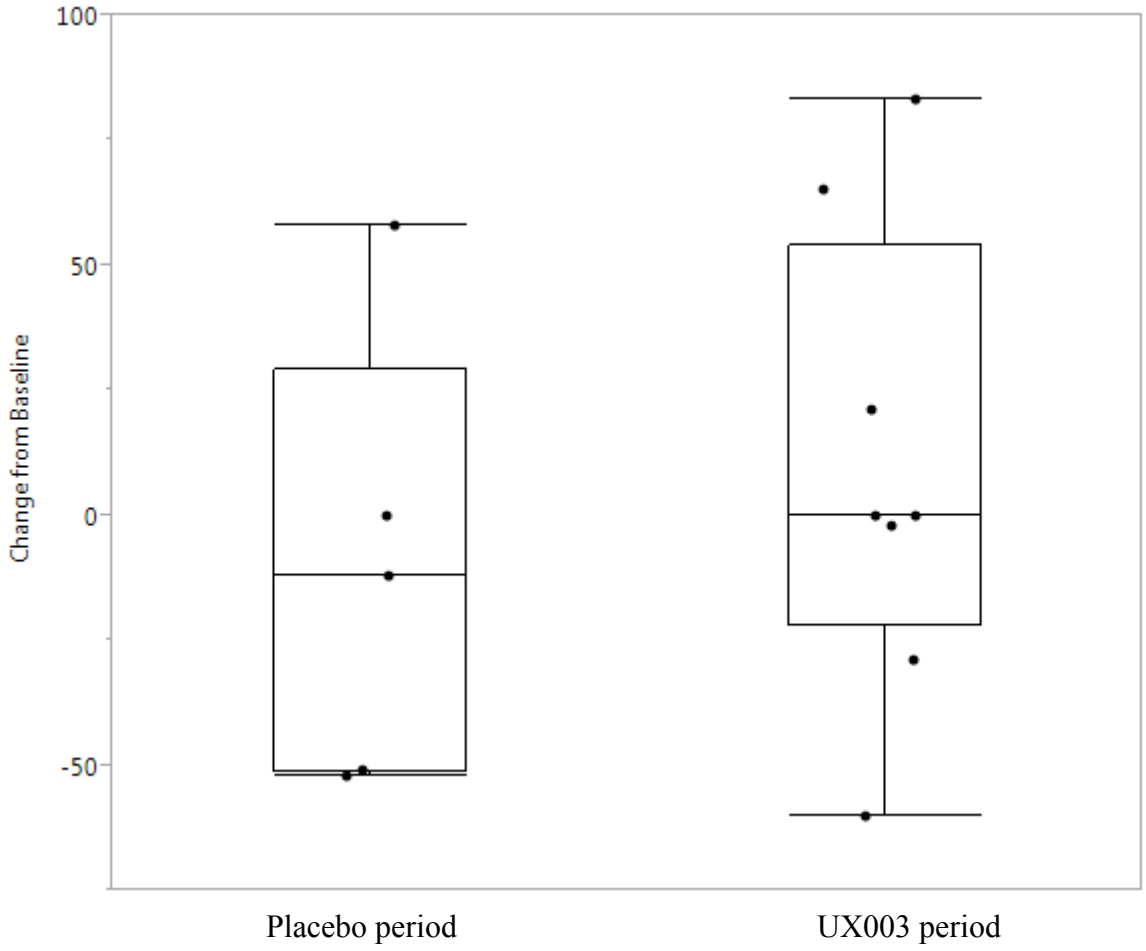
Variability of 6MWT baseline values ranges from 0 to nearly 600 meters. The other important confounding factor is the age of these subjects. Table 5 summarizes patients' age for different treatment cohorts. We noted that patients in Group A and B (i.e., those received longer time of treatment) were relatively younger.

Table 5: Subjects' Age in Study CL301

Group	Age (SUBJID)
A	14.7 ((b) (6)) 13.4 ((b) (6)) 11.5 ((b) (6))
B	8.5 ((b) (6)) 16.5((b) (6)) 12.8 ((b) (6))
C	22.6 ((b) (6)) 22.5 ((b) (6)) 17.4 ((b) (6))
D	10.1 ((b) (6)) 25.3 ((b) (6)) 10.5 ((b) (6))

2. **(Boxplot: Compare placebo vs UX003 at treatment week 24)** Due to the design of Study CL301, only patients in Group D have 24 weeks placebo treatment. Consider the small sample size, this reviewer includes all data from placebo treatment in the following boxplot; Group B patients contribute 8 weeks placebo treatment and Group C patients contribute 16 weeks placebo treatment. Subject (b) (6) completed 6MWT at screening, baseline and treatment 16 visits. Subject (b) (6) completed 6MWT at screening, baseline and treatment 32 visits. These two subjects are included in the treatment period using their week 16 and week 32 6MWT values respectively.

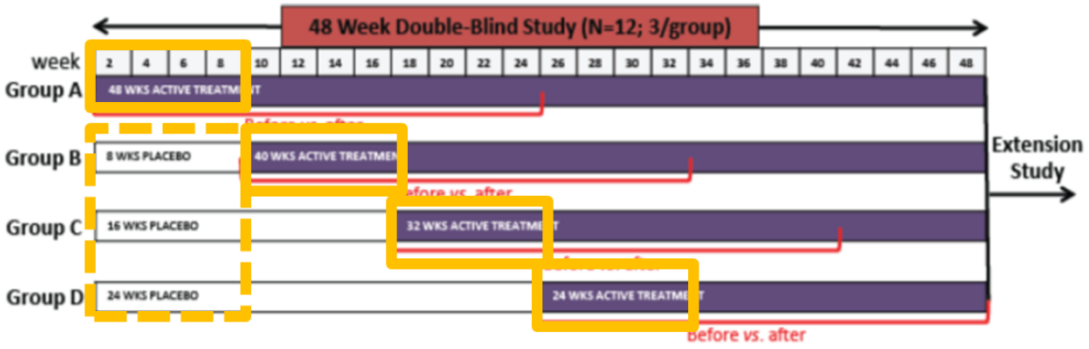
Figure 5: Change from Baseline (Placebo Period vs. UX003 Period) at Treatment Week 24 for 6MWT



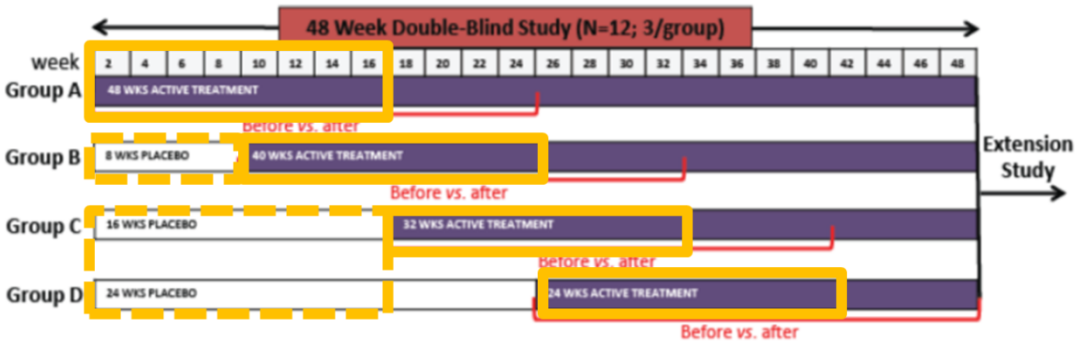
Median of UX003 period is 0 at treatment week 24, while that of placebo data is -12. The ranges of variability of both periods are similar. No significant positive change from baseline in 6MWT is detected after 24 weeks of UX003 treatment. Considering the small sample size and that some of the patients were included in both placebo and UX003 periods, it is hard to tell the difference in change from baseline of UX003 period over that of placebo period is significant.

3. **(ANCOVA analyses at different treatment durations)** This reviewer conducted ANCOVA analyses for evaluating the change from baseline between UX003 and placebo periods at different treatment durations (8 weeks, 16 weeks, and 24 weeks) in 6MWT. Group cohort (i.e., Group A, B, C and D) is included as a categorical covariate. This reviewer noted that the patients' age range in this study population was big and that the mean ages were very different across four group cohorts, as well as their baseline 6MWT values; therefore age and baseline 6MWT are included as continuous covariates in the analysis model. Parallel group analyses considered to evaluate the treatment effect. See the following diagrams at different treatment duration (yellow solid vs. yellow dashed).

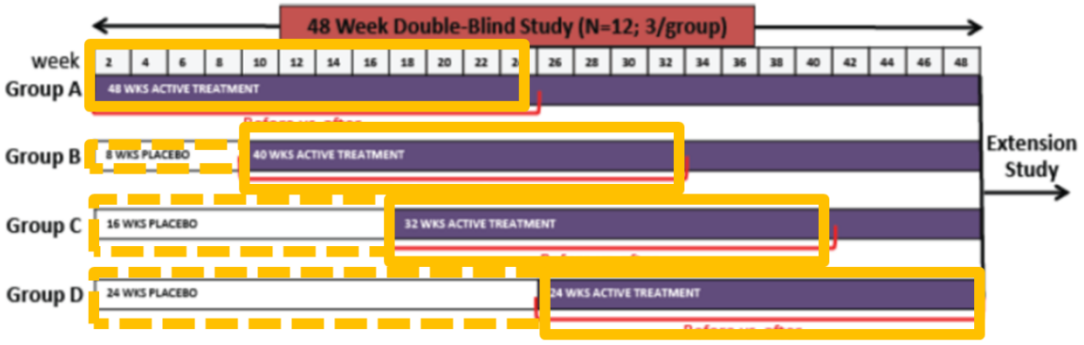
- *8 weeks treatment*



- *16 weeks treatment*



- *24 weeks treatment*



At 24 weeks treatment, clinical reviewer suggested imputing 6MWT value for Subject (b) (6) and (b) (6) with their 16 weeks treatment and 32 weeks treatment values respectively.

Table 6: Least Square Mean Difference between UX003 Period and Placebo Period in 6MWT (m)

Treatment Duration	LS mean (SE)	P-value
8 weeks treatment	-21.85 (26.70)	0.45
16 weeks treatment	12.36 (32.43)	0.72
24 weeks treatment	16.96 (34.58)	0.64

Table 6 summaries the least square mean difference between UX003 period and placebo period in 6MWT at different treatment durations. After 24 weeks of treatment, the least

square mean difference between UX003 and placebo periods is approximately 17 meters with a relative large standard error 34.6.
We also conducted the sensitivity analysis by including the 6MWT values for those subjects who used the same walking assistive devices throughout the trial, before and after initiating treatment ((b) (6) and (b) (6)) and only imputed 6MWT values as zero for subjects who began using a device after initiating treatment ((b) (6)). Table 7 summaries our results, which are very close to those in Table 6.

Table 7: Least Square Mean Difference between UX003 Period and Placebo Period in 6MWT (m) based on Sensitivity Analysis

Treatment Duration	LS mean (SE)	P-value
8 weeks treatment	-28.11 (26.26)	0.33
16 weeks treatment	8.75 (30.26)	0.78
24 weeks treatment	17.54 (34.53)	0.63

We would like to emphasize that due to the small sample size of Study CL 301 (a total of 12 subjects; 3 subjects per group), it is questionable that the ANCOVA analysis results are valid. Therefore, results should be interpreted with caution.

4. **(Permutation tests at different treatment durations)** This reviewer conducted (stratified) permutation tests to evaluate the change from baseline of the treatment effect in 6MWT at different treatment durations (8 weeks, 16 weeks, and 24 weeks). Three different cohorts are considered as blocking factors in the permutation tests:
    - Age < 12, 12 ≤ Age < 18, or Age ≥ 18
    - Randomization Baseline < 300 meters vs Randomization Baseline ≥ 300 meters
    - Group A, B, C or D

Table 8 shows that no strong evidences are detected to indicate the superiority of UX003 period over placebo period.

Table 8: Two-sided P-values for Treatment Effect in Permutation Tests under Different Cohorts

Treatment Duration	No Stratification	Stratified by Group	Stratified by Baseline 6MWT	Stratified by Age
8 weeks treatment	0.49	0.37	0.41	0.50
16 weeks treatment	0.35	0.73	0.39	0.50
24 weeks treatment	0.41	0.62	0.47	0.54

Of note, the validity of permutation test results depends on the data exchangeability, which cannot be justified due to small sample size.

3.3 Evaluation of Safety

The safety evaluation is not conducted in this review. Refer to the clinical review and division director summary for an evaluation of safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

No subgroup analyses were performed due to the small sample size.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Due to the rarity of MPS VII, the applicant only submitted one efficacy study CL301 with 12 subjects enrolled to support the use of UX003 for the treatment of MPS VII patients.

During an IND meeting for this clinical development program, the FDA notified the applicant (b) (4) . Among the 6 key secondary endpoints included in multi-domain clinical responder index (MDRI), the clinical team determined that only the Six Minute Walk Test (6MWT) after 24 weeks of treatment is suitable for efficacy evaluation.

Multiple issue and challenges were identified in reviewing study CL301. One major challenge is 9 out of the total 12 patients (Group B, C, and D patients) contributed both placebo and UX003 data and thus the final analysis should take into account of the cohort effect. Otherwise, the independence assumption of each data point in ANCOVA will be violated and lead to invalid results. Another issue that may impact the efficacy analysis is 3 subjects utilized walking assistive devices when performing 6MWT. Simply analyzing data using the available 6MWT value for all subjects may lead to biased results.

The statistical reviewer evaluated the study data and the sponsor's supportive evidence for UX003's efficacy in this application. The applicant's analysis results for uGAG (the primary endpoint in the EU and rest of the world) were confirmed, as well as all results for the key secondary endpoints included in MDRI (six-minute walk test (6MWT), forced vital capacity (FVC), shoulder flexion, visual acuity, Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) fine motor and gross motor). It is important to note that the applicant claimed UX003's efficacy solely through the comparison of patients' improvement from their baseline (i.e., placebo period prior to cross over) which did not take the advantage of the placebo data in the blind-start design even though the Agency did not agree upon the use of this design for this drug development In particular, the applicant did not assess drug effect across cohorts. As a result, the statistical reviewer performed additional superiority analyses (i.e., ANCOVA, and permutation tests) to evaluate the superiority of UX003 period over placebo period using the end point of the change from baseline after 24 weeks of treatment in 6MWT.

5.2 Conclusions and Recommendations

The applicant concluded, based on their analysis results of uGAG and MDRI data, that CL301 was a positive study providing compelling evidence for the efficacy of UX003 in treating MPS

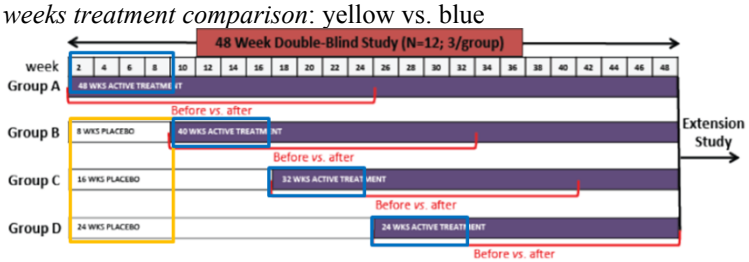
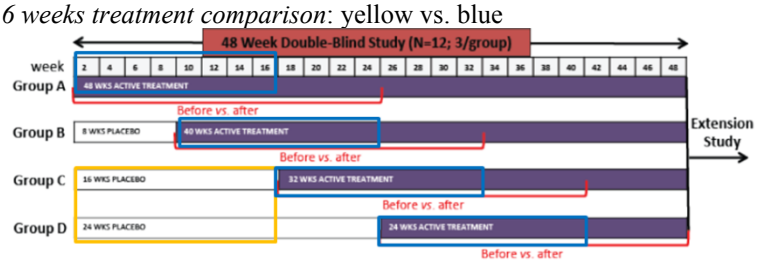
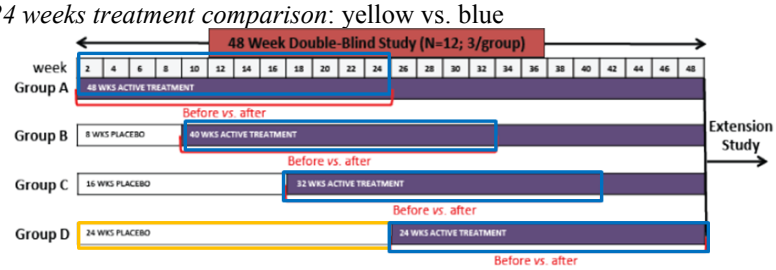
VII patients. This conclusion needs to be reevaluated due the lack of comparison with placebo period data based on blind-start design, the concerns of using MDRI and uGAG, and the exploratory analyses conducted after the fact.

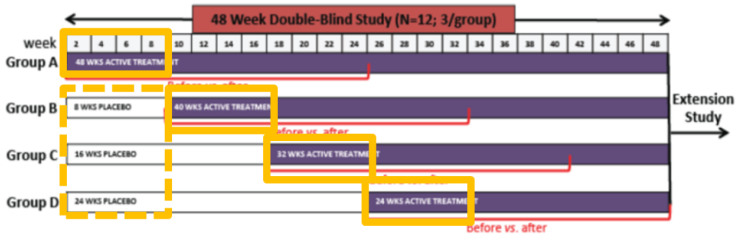
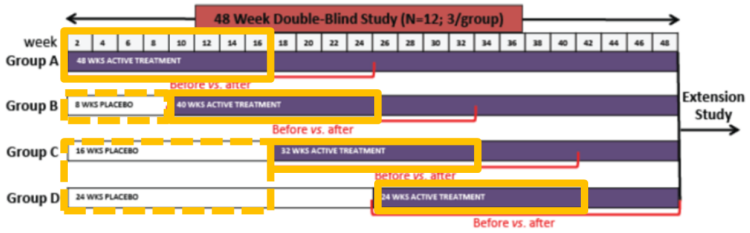
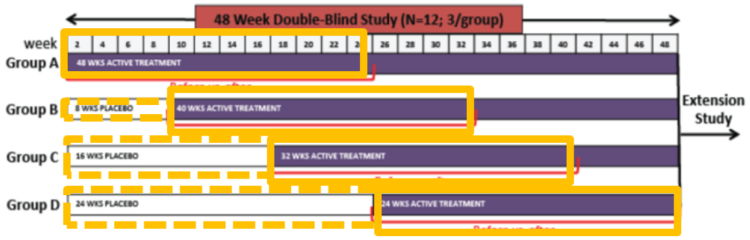
According to the reviewer's thorough evaluation and the applicant's confirmation of the reviewer's results through 48 weeks, no substantial evidence is demonstrated supporting the efficacy of UX003 in treating patients with MPS VII, i.e., Sly syndrome. We note, however, UX003's effect appears to increase gradually over time. While many other analyses could be performed, no quantitative analysis will overcome the small sample size and lack of assessments in some patients for some of the outcomes.

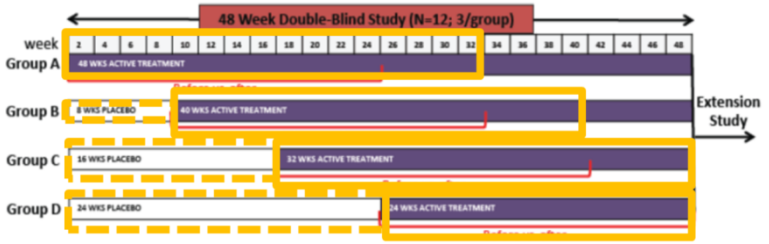
6 APPENDIX

6.1 SUMMARY OF INFORMATION REQUESTS (IR)

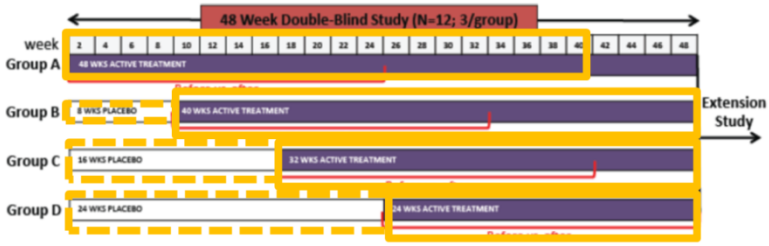
Table 9: Summary of Main Statistics Related Information Requests (IR)

IR Sent date	IR Response date	Major Issue(s) and IR response (Datasets) location
06/07/2017	06/16/2017	• Provide boxplots for the six endpoints (6MWT, FVC%Pred, BOT-2 Fine, BOT-2 Gross, Shoulder flexion and Visual Acuity), considering parallel group design at different treatment durations (8 weeks, 16 weeks, and 24 weeks). *8 weeks treatment comparison*: yellow vs. blue  *16 weeks treatment comparison*: yellow vs. blue  *24 weeks treatment comparison*: yellow vs. blue  • Provide median test for the change from baseline of the 6MWT and uGAG between the placebo and UX003 periods at different treatment durations (8 weeks, 16 weeks and 24 weeks; see yellow vs. blue above for reference). • Apply a bootstrapping method to obtain the mean difference as well as median difference in change between UX003 and placebo periods for 6MWT, uGAG and BOT2-Fine. Provide boxplots for Week 8, Week 16 and Week 24 comparisons. • Provide a side-by-side spaghetti plot of 6MWT with one line for each patient over time. Use different style lines for the four groups (A, B, C and D). Use different colors for males and females. • Provide a side-by-side spaghetti plot of uGAG with one line for each patient over time. Use different style lines for the four groups (A, B, C and D). Use different colors for males and females. • Provide heat maps for each of the six endpoints. For each week (8, 16, 24, 32, 40 and 48), calculate the following values. Median of each group

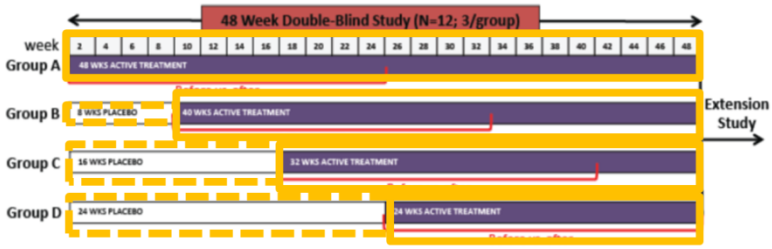
		should be used. A: A-B B: A-C, B-C C: A-D, B-D, C-D \\cdsesub1\evsprod\bla761047\0022\m5\53-clin-stud-rep\535-rep-effic-safety-stud\mps-7\5351-stud-rep-contr\ux003-cl301\listings-figures-info-req-06jun2017.pdf
07/26/2017	08/07/2017	• For both of the 6MWT and uGAG data, conduct the following analyses for evaluating the change from baseline between UX003 and placebo periods at different treatment durations (8weeks, 16 weeks, 24 weeks, 32 weeks, 40 weeks and 48 weeks). Include the group cohort (i.e., Group A, B, C and D) as a categorical covariate, baseline 6MWT (or baseline uGAG) and age as continuous covariates in the analysis model. For subjects who utilized any walking assistive devices (e.g., Subject (b) (6), (b) (6), and (b) (6)) when performing 6MWT, impute data using two approaches: 1) impute 6MWT value (uGAG value) as zero; 2) impute 6MWT (uGAG) change from baseline value as zero. For handling missing data, consider two approaches: 1) observed cases; 2) last observation carried forward (LOCF). *8 weeks treatment*  *16 weeks treatment*  *24 weeks treatment*  *32 weeks treatment*



40 weeks treatment



48 weeks treatment



1. ANCOVA Analysis
2. Provide stratified permutation tests to evaluate the change from baseline of the treatment effect in 6MWT at different treatment durations (8 weeks, 16 weeks, 24 weeks, 32 weeks, 40 weeks and 48 weeks). In particular, permute outcomes under different cohorts, and then conduct permutation tests (see diagrams in (a) as reference). Permute the outcomes under the following three cohorts:
   - Age < 18 vs Age >= 18
   - Baseline < 300 meters vs Baseline > 300 meters
   - Group A, B, C or D

   For example, for subjects who were less than 18 years old, permute the outcomes. Also permute outcomes for subjects who were older than 18. Conduct permutation test to see if the outcomes and the treatment periods (UX003 or placebo) are independent. Consider the ways of handing missing data and imputing data for subjects who used walking aids, a total of 4*18=72 stratified permutation tests need to be conducted.

- Provide Generalized Estimating Equation (GEE) analysis to evaluate the change from baseline of the UX003 period over that of the placebo period in uGAG excretion, including baseline value and age as continuous covariates and group (Group A, B, C or D) as categorical covariate. The covariance structure within subjects can be assumed to be exchangeable. Consider change from study week 0 to 48 (every 2 week increment), including missing data in the analysis.
- Provide Generalized Estimating Equation (GEE) analysis to evaluate the

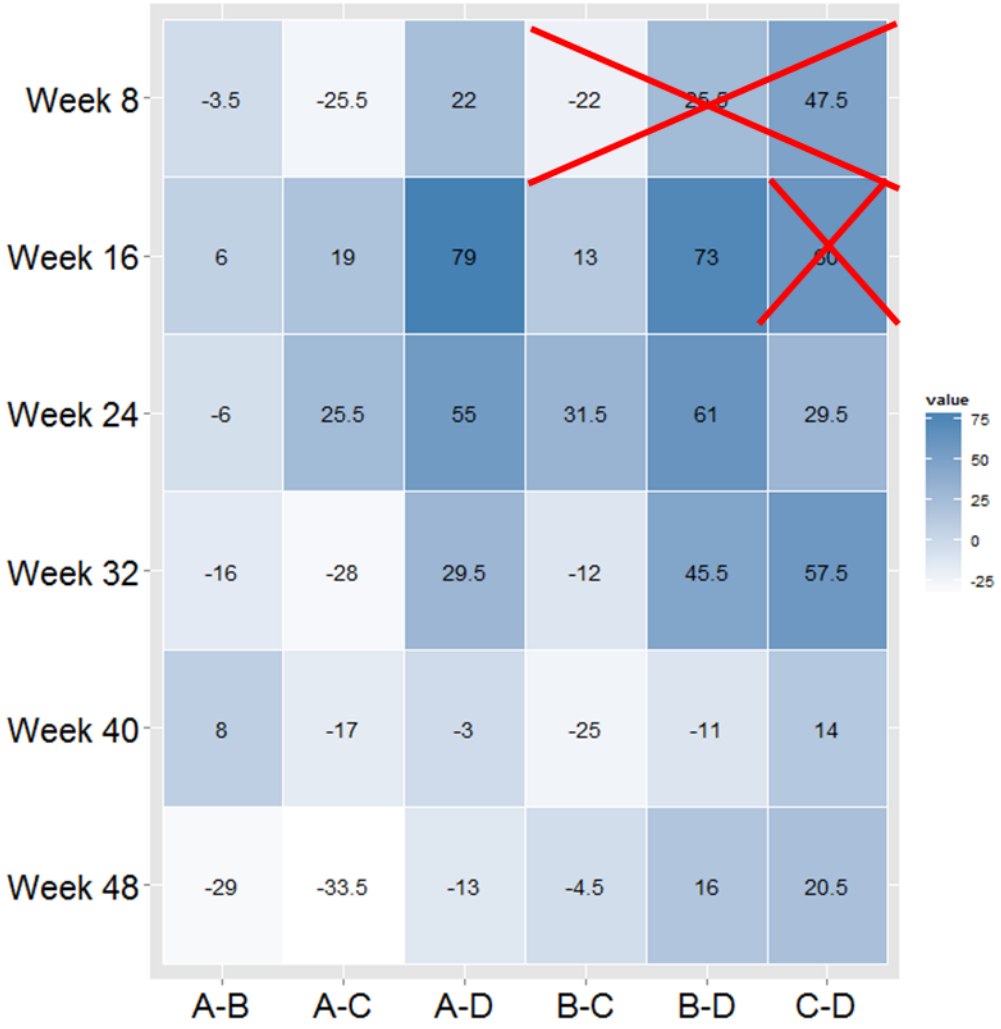
		change from baseline of the UX003 period over that of the placebo period in 6MWT, including baseline value and age as continuous covariates, and group (Group A, B, C or D) as categorical covariate. The covariance structure within subjects can be assumed to be exchangeable. Impute change from baseline values for subjects who used any type of walking aid devices as zeros. Consider change from study week 0 to 48 (every 8 week increment), including missing data in the analysis. \\cdsesub1\evsprod\bla761047\0036\m5\53-clin-stud-rep\535-rep-effic-safety-stud\mps-7\5351-stud-rep-contr\ux003-cl301\tables-info-req-26jul2017.pdf

6.2 Reviewer's Additional Analysis Results

1. **(Heat maps for change from baseline in 6MWT)** Red crossed boxes should have no improvement.

See Figure 6 below. Baseline values are set to be the values at study week 0 for all the groups. Numerical values in each box are the Group median difference of change from baseline values. For example, at study week 8, median change from baseline of Group A minus that of Group B is -3.5. We expected to have 32 out of 36 boxes showing improvement, however only 56% (18/32) of boxes showed improvement.

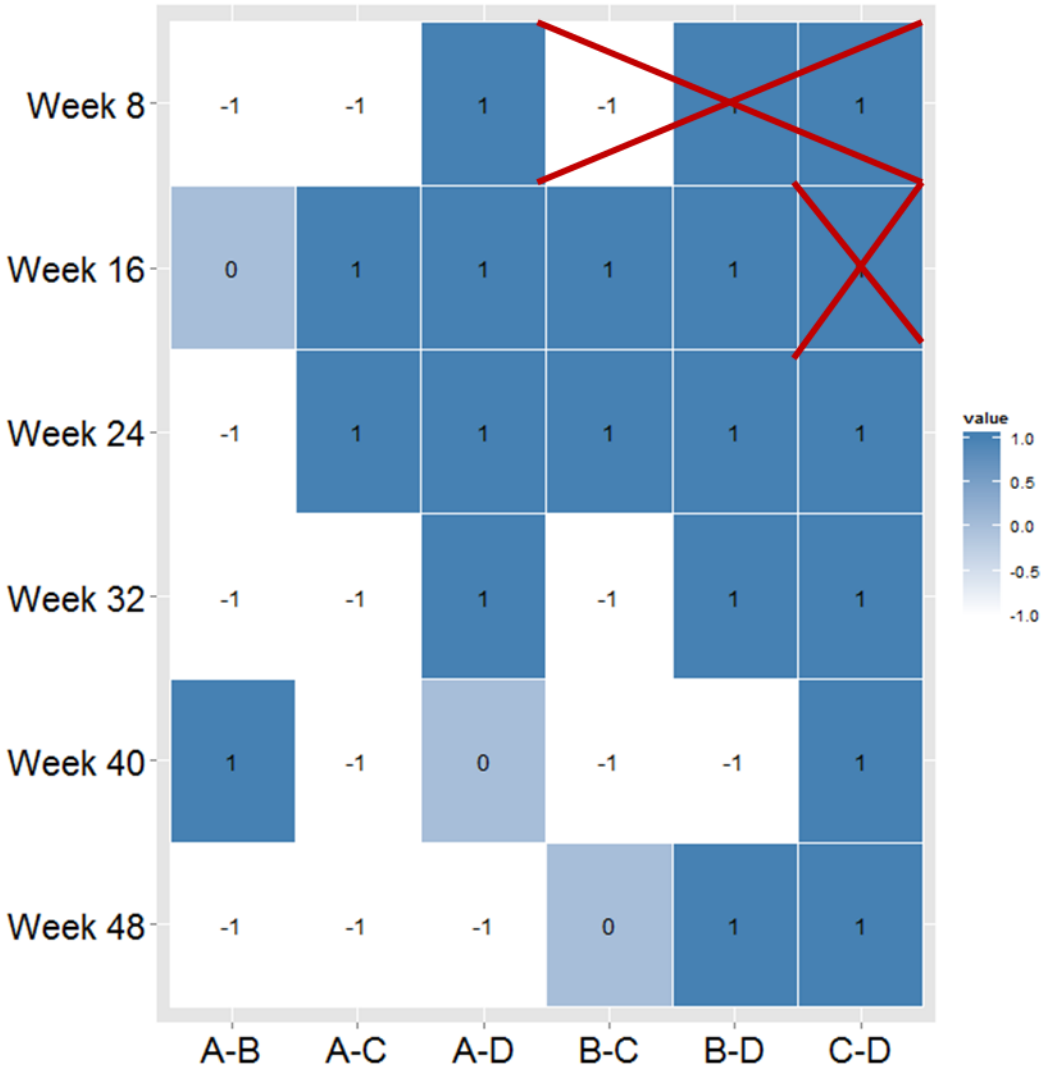
Figure 6: Heat Map (Group Median Difference)



For Figure 7, we define the following improvement criteria: improvement is +1 if the ratio of median change from baseline in 6MWT of Group i minus that of Group j to the absolute value of median change from baseline of Group i is greater than 30%; -1 if the ratio of median change from baseline in 6MWT of Group i minus that of Group j to the absolute value of median change from baseline of Group i is less than -30%; otherwise 0. As shown in the Figure, we observed 53% (17/32) improvement and 38% (12/32) worsening.

For both Figures 6 and 7, it is interesting to note that among the four cells, which were crossed and were not expected to show any important, three of them had values indicating positive findings (either positive or 1, respectively). It suggests that data of this study do not seem stable, which may be mainly due to the small sample size.

Figure 7: Heat Map (Improvement Criteria)



2. **(Correlations between endpoints)** Multivariate analysis was conducted to assess the pairwise correlations between any two of the endpoints (uGAG, 6MWT, BOT-2 Balance, BOT-2 fine motor precision, BOT-2 manual dexterity and BOT-2 running speed & agility) at treatment week 24. Figure 8 shows that uGAG, balance and running speed & agility are relatively highly correlated. Due to the small sample size of the study, the robustness of correlation results is questionable. The permuted correlation tests were also conducted for these endpoints. We noted that p-value for pairwise correlation between uGAG and balance is 0.078, while p-value between balance and motor prevision is 0.18. In addition, the correlation between uGAG and 6MWT appears to be very low (p is only around 0.25).

Figure 8: Heat map for correlation between endpoints

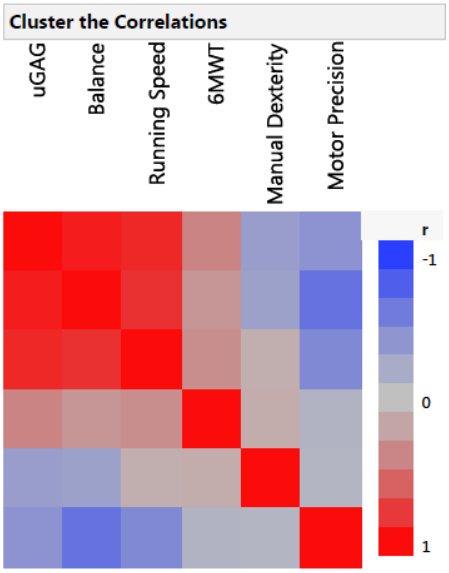
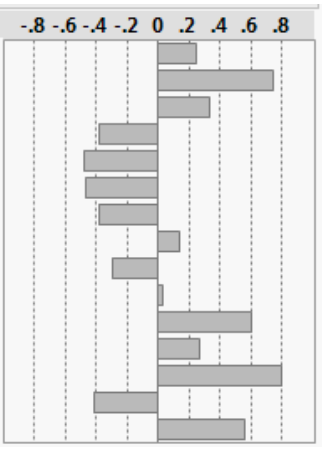


Figure 9: Pairwise Correlations between endpoints

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
6MWT	uGAG	0.2526	8	-0.5500	0.8126	0.5461
Balance	uGAG	0.7443	6	-0.1699	0.9700	0.0897
Balance	6MWT	0.3390	6	-0.6519	0.9023	0.5110
Motor Precision	uGAG	-0.3836	11	-0.7995	0.2809	0.2442
Motor Precision	6MWT	-0.4776	8	-0.8846	0.3423	0.2314
Motor Precision	Balance	-0.4671	6	-0.9272	0.5548	0.3503
Manual Dexterity	uGAG	-0.3733	11	-0.7951	0.2920	0.2582
Manual Dexterity	6MWT	0.1402	8	-0.6264	0.7689	0.7406
Manual Dexterity	Balance	-0.2909	6	-0.8919	0.6816	0.5759
Manual Dexterity	Motor Precision	0.0346	11	-0.5773	0.6216	0.9196
Running Speed	uGAG	0.6004	5	-0.5993	0.9692	0.2843
Running Speed	6MWT	0.2723	5	-0.8028	0.9309	0.6577
Running Speed	Balance	0.8018	5	-0.2750	0.9863	0.1027
Running Speed	Motor Precision	-0.4082	5	-0.9488	0.7409	0.4950
Running Speed	Manual Dexterity	0.5590	5	-0.6378	0.9652	0.3273



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

/s/

FEIRAN JIAO
09/29/2017

MIN MIN
09/29/2017

YEH FONG CHEN
09/29/2017