STATISTICAL ANALYSIS PLAN
Protocol ID:00103 / Autism

A Double Blind Randomized Placebo Controlled Study of CM-AT for the Treatment of Autism in Children With All Levels of Fecal Chymotrypsin

Document Date: 27 December 2017
Upload Date: 2 May 2021

STATISTICAL ANALYSIS PLAN
Amendment 1

A Double Blind Randomized Placebo Controlled Study of CM-AT for the Treatment of Autism in Children with All Levels of Fecal Chymotrypsin

PROTOCOL No. 00103 Autism

Sponsor: **Curemark, LLC**

Date: December 27, 2017

Approvals (sign and date):

Curemark, LLC

Joan Fallon, CEO

Joseph Massaro, PhD, Statistician

TABLE OF CONTENTS

1. Study Objectives

To examine children with autism at all levels of fecal chymotrypsin using the Aberrant Behavior Checklist.

Endpoint (outcome) measurements to determine efficacy of treatment with CM-AT versus placebo for changes in the Aberrant Behavior Checklist subscale for Irritability/Agitation (ABC-I) between Baseline (defined as score at Screening Part 1 for efficacy assessments; in the event a subject has more than one Screening Part 1 value, the value to be used will be the last value of the variable measured prior to start of randomized treatment) and Week 12 visit and changes in the Aberrant Behavior Checklist subscale for Lethargy/Social Withdrawal (ABC-L) between Baseline and Week 12 of treatment.

To continue to monitor the safety of CM-AT in this population of children.

2. Study Design

This is a Phase III, multi-center, randomized, double-blind, placebo-controlled study of approximately 300 subjects diagnosed with autism from approximately 35 clinical sites in the United States. Subjects will be between 3 and 8 years of age (inclusive) at the time of consent. All study participants must have an initial ABC-I score $\geq$ 11 and a diagnosis of Autistic Disorder as demonstrated by the Diagnostic and Statistical Manual for Mental Disorders IV Text Revision (DSM-IV-TR) and by the Autistic Diagnostic Interview – Revised (ADI-R). Subjects will also be screened using the Social Communication Questionnaire (SCQ) and must achieve a score $\geq$ 15 to be considered eligible.

After initial screening criteria are met, including the diagnosis of Autistic Disorder as demonstrated on the DSM-IV-TR, SCQ and ADI-R, the study participants will be administered the ABC. If their ABC-I score is greater than or equal to 11 ($\geq$ 11) they can be enrolled and given a two-week initial blinded placebo run-in. At the end of the run-in, subjects will be evaluated and randomized if all inclusion criteria are satisfied.

The compliance and inclusion criteria required for randomization are: 1) subjects must take the study drug on the prescribed regimen and 2) continue to meet inclusion criteria. At the Visit 1 Visit, they will be reevaluated for their level of Irritability/Agitation with the ABC-I scale. If a subject's score on the ABC-I scale has increased or decreased by 30% or more, or is <9 on the retest, they will be discontinued from the study and not randomized. Qualified subjects can be randomized into the study if they meet all inclusion/exclusion criteria.

After enrolling in the study at a clinical site, subjects ultimately randomized to treatment will receive study drug or placebo three times daily with food for a

period of 12 weeks. Parents will keep a missed dose log. Subjects will be seen at Week 2, 4, 6, 8, 10, and 12, post-randomization. Standardized measures of behavior, health, and quality of life will be utilized. In addition, measures of overall health including growth and weight will be obtained.

Subjects will be randomized in a 1:1 ratio to CM-AT or Placebo (both biomarker negative (>12.6 U/g) and positive (≤12.6 U/g) will be included). Randomization will be based upon study drug vs placebo.

Subjects will be randomized to ensure 82 subjects per study arm of the low-level fecal chymotrypsin (FCT) children are evaluable for the primary statistical analysis (treatment comparison on the primary endpoint of change in ABC irritability from Baseline to Week 12 on the primary analysis set of children with FCT ≤12.6 in randomized subjects). The FCT value used in determination of whether the patient has FCT ≤ 12.6 is the Screening Part 1 value; if the patient has more than one screening FCT, the last FCT measured prior to randomization will be used to determine whether the patient has FCT ≤ 12.6. It is anticipated that each site will enroll approximately 3 to 5 subjects per month and the study will be open for a period of twelve months. A power recalculation may be performed midway through Study Protocol 00103 to determine if sufficient numbers of biomarker negative children have been enrolled.

A subject may enroll in Study Protocol 00104 Open Label Extension after successfully completing Week 12 of Study Protocol 00103. The analyses of data collected in Study Protocol 00104 Open Label Extension will be discussed in a separate analysis plan.

3. Study Analysis Sets

One-hundred sixty-two (162) subjects will be randomized per treatment arm, or 324 subjects total. These subjects will be randomized from approximately 35 sites throughout the United States. Approximately 350 subjects will be enrolled into the placebo run-in period to ensure that 324 will be randomized. At the completion of the two-week placebo run-in, the subject will be assessed for compliance and ABC-I scores. If a subject's score on the ABC-I scale has increased or decreased by 30% or more, or is <9 on the retest, they will be discontinued from the study and not randomized.

The study population will consist of children, both males and females, between 3 and 8 years of age inclusively at the time of consent. All study participants must have an initial ABC-I score of ≥11 and a diagnosis of Autistic Disorder as demonstrated by the DSM-IV-TR and by the ADI-R. Subjects will also be screened using the SCQ and must achieve a score of ≥15 to be considered eligible.

The following populations form the basis of the analysis sets that will be included in efficacy and safety analyses (not listed in order of analysis below. For order of analysis see Section 6.0):

Intent-to-Treat (ITT): All randomized subjects who received study drug and had at least one post-Baseline ABC-I assessment.

Low-level FCT ITT: All randomized subjects with pathologically low levels of FCT (≤12.6 U/g) who received study drug and had at least one post-Baseline ABC-I assessment. Subjects are analyzed under the treatment to which they are randomized. This is the primary analysis population for efficacy.

Non-Low-Level FCT (FCT >12.6) ITT: All non-low-level (FCT >12.6 U/g) randomized subjects who received study drug and had at least one post-Baseline ABC-I assessment. Subjects are analyzed under the treatment to which they are randomized. This is an exploratory analysis population for efficacy.

All Randomized: All randomized subjects. This is an exploratory analysis population for efficacy. Subjects are analyzed under the treatment to which they are randomized.

Low-level FCT All Randomized Subjects: All randomized subjects with pathologically low levels of FCT (≤12.6 U/g). This is a population for use in a sensitivity analysis of efficacy. Subjects are analyzed under the treatment to which they are randomized.

Non-Low-Level FCT (FCT >12.6) All Randomized Subjects: All non-low-level (FCT >12.6 U/g) randomized subjects. Subjects are analyzed under the treatment to which they are randomized. This is an exploratory analysis population for efficacy.

Per-Protocol (PP): All ITT subjects who have completed a Week 12 visit and have a Week 12 ABC-I score. Subjects are analyzed under the treatment to which they are randomized. This is an exploratory analysis population for efficacy.

Low-Level FCT (FCT ≤12.6) Per Protocol (PP): All low-level (FCT ≤12.6 U/g) PP Subjects. Subjects are analyzed under the treatment to which they are randomized. This is an exploratory analysis population for efficacy.

Non-Low-Level FCT (FCT >12.6) PP: All non-low-level (FCT >12.6 U/g) PP Subjects. Subjects are analyzed under the treatment to which they are randomized. This is an exploratory analysis population for efficacy.

Run-in Analysis Set: All subjects receiving at least one dose of placebo during the placebo-run in phase. All subjects will be analyzed as one group.

Safety Analysis Set: All subjects receiving at least one dose of randomized treatment. Subjects are analyzed under the treatment received. This is a primary analysis set for safety.

Low-Level FCT (FCT ≤12.6) Safety Analysis Set: All Low-Level subjects (FCT ≤12.6) receiving at least one dose of randomized treatment. Subjects are analyzed under the treatment received. This is the secondary analysis set for safety.

4. Test product, Dose and Mode of Administration

900mg of CM-AT (pancreatic enzyme concentrate encapsulated in hydrogenated soybean oil) 3x daily sprinkled over food. To be taken with a meal or large snack.

5. Demographics, Other Baseline Characteristics (other than efficacy and safety variables), Study Drug Compliance

"Baseline" for variables other than efficacy and safety refers to the last value of the variable measured prior to start of randomized treatment. For efficacy and safety, "Baseline" is defined as the value measured at Screening Part 1, denoted as "Screening" below.

5.1 Demographics – Baseline demographics of age, gender (male; female), ethnicity (Hispanic or Latino; not Hispanic or Latino), race (American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or other Pacific Islander); height (cm), weight (kg), body mass index (BMI).

5.2 Physical examination abnormalities by body system (HEENT, Respiratory, Cardiovascular, Abdomen, Musculoskeletal, Central Nervous System, Dermatologic, General Appearance, Other)

5.3 Responses to each question on Study Specific Medical History questionnaire

5.4 Response to each question on Neurodevelopmental history questionnaire

5.5 Responses to each question on Family History questionnaire

5.6 Responses to each question on Parent Visit questionnaire

5.7 Screening ABC-I score. The scoring of this subscale is given in Appendix A.

5.8 Screening DSM-IV-TR total score and sub-scores (Social Interaction; Communication; Behavior) as described in Appendix C.

5.9 Screening SCQ total score and sub-scores (Social Interaction; Communication; Restricted, Repetitive, and Stereotyped Patterns of Behavior).

The total score and the subscales of Reciprocal Social Interaction, Communication, and Restricted, Repetitive, and Stereotyped Patterns of Behavior domain are given in Appendix B.

5.10 Screening ADI-R (Total A, Verbal Total B, Nonverbal Total B, Total C and Total D). These scores are determined by the rater and recorded directly on the questionnaire.

5.11 Study Drug compliance will be assessed based upon the sachet usage. The calculation for percentage of compliance is 100 times the number of sachets dispensed at the previous visit minus the number of sachets returned, divided by the days between visits times 3 (not including days after the last visit at which sachets were dispensed if the subject did not return after that last visit). Study drug compliance during the 2-week placebo run-in period and during the randomized double-blind treatment period will be analyzed.

6. Efficacy Variables

The following lists the efficacy variables that will be collected on each subject and analyzed. Section 7 of this document provides the details of the analysis methodology.

6.1 **The primary efficacy variable** is the change from Baseline (score at Screening Part 1) to Week 12 in the Aberrant Behavior Checklist (ABC) subscale for Irritability/Agitation (ABC-I) in Low-level FCT ITT subjects. The subscale item components are given in Appendix A. Only the raw scores will be analyzed.

6.2 The **secondary efficacy variable** is the change from Baseline (score at Screening Part 1) to Week 12 for the Aberrant Behavior Checklist subscale for Lethargy/Social Withdrawal (ABC-L) in Low-level FCT ITT subjects. The scoring of this subscale is given in Appendix A. Only the raw scores will be analyzed.

6.3 Exploratory Efficacy Variables include:

- Behavioral changes from Baseline (score at Screening Part 1 unless otherwise specified) to Week 12 (the scoring of these subscales is given in Appendix A):

Aberrant Behavior Checklist			
	FCT ≤12.6 U/g	FCT >12.6 U/g	FCT – All Levels
ABC Irritability		X	X
ABC Lethargy		X	X
ABC Hyperactivity	X	X	X
ABC Stereotypy	X	X	X
ABC Inappropriate Speech	X	X	X
ABC Total Score	X	X	X
Additional Exploratory Endpoints			
Clinical Global Impression – Improvement	X	X	X
Clinical Global Impression – Severity (Baseline is the value measured at the randomization visit)	X	X	X
Social Responsiveness Scale ((Baseline is the value measured at the randomization visit)	X	X	X
Pediatric Quality of Inventory			
Physical Functioning	X	X	X
Emotional Functioning	X	X	X
Social Functioning	X	X	X
Cognitive Functioning	X	X	X
Communication	X	X	X
Worry	X	X	X
Daily Activities	X	X	X
Total Score	X	X	X

- Gastrointestinal measures will be performed at each regularly scheduled visit, with abnormality status assessed at each visit and with change defined as the difference between Baseline (measures from Screening Part 1) to Week 12.
- Weight will be assessed at each regularly scheduled visit,
- Weight gain is defined as Post-Baseline weight ≥ Baseline weight (Baseline weight is the weight at Screening Part 1. If a patient has more than one Screening Part 1 weight, the last weight measured prior to randomization will be used).
- Weight loss is defined as Post-Baseline weight ≤ Baseline weight.

Stool Test Panel & Weight Change			
	FCT ≤12.6 U/g	FCT >12.6 U/g	FCT – All Levels
C.Diff. Tox A&B	X	X	X
A1-Antitrypsin	X	X	X
Fecal Chymotrypsin	X	X	X
Cryptosporidum AG	X	X	X
H Pylori Ag,	X	X	X
pH Stool	X	X	X
Weight	X	X	X

7. Safety Parameters

7.1 Adverse Events – Incidence of treatment emergent adverse events are collected for each subject. A run-in emergent adverse event is defined as an event starting or worsening after the start of placebo run-in phase and before the start of double-blind treatment. A treatment emergent adverse event (TEAE) is defined as an event starting or worsening on or after the start of double-blind treatment. Each adverse event will be coded to a MedDRA system organ class (SOC) and preferred term (PT). Severity, seriousness and relationship to study drug for each event will also be collected.

7.2 A complete physical exam is conducted at the start of the placebo run-in phase. Brief physical exams are conducted at the randomization visit (prior to randomization) and during the double-blind phase. If clinically significant findings are identified during the physical exam, an adverse event form will be completed.

7.3 Stool pH is collected at Baseline (Screening Part 1) and at several Post-Baseline visits including Week 12. The value at each visit and the change from Baseline at each Post-Baseline visit will be analyzed.

7.4 For each subject at each visit at which Stool pH is collected, the incidence of abnormal outcomes for stool pH (<5.92 or >8.00), fecal globin (present) and fecal occult blood (present) are collected.

7.5 A1-Antitrypsin in the feces from the Fecal Panel is collected at Baseline (Screening Part 1) and Post-Baseline. It will be the variable for analysis along with the change from Baseline at each Post-Baseline visit. Additionally, at each time point for which α1-antitrypsin is measured, each subject's incidence of A1-Antitrypsin <55 and ≥55 will be recorded.

7.6 Vital signs - The analysis variables are the change from Baseline (Screening Part 1) at each Post-Baseline at each visit for visit for weight (kg), systolic blood pressure (mmHg), diastolic blood pressure (mmHg) and pulse (beats per minute).

8. Statistical Methodology

8.1 Sample Size

Sample size is based on:

- The primary efficacy variable -
  - Change from Baseline (score at Screening Part 1) to Week 12 for the Aberrant Behavior Checklist (ABC) subscale for Irritability

- The primary analysis set of Low-level FCT ITT subjects.

- 1:1 randomization scheme

Previously reported Sponsor data indicating that 60% of autistic children have pathologically low levels of FCT formed the basis for the following calculations. Assuming a -3.7 ± 7.0 ABC Irritability change from Baseline to Week 12 of for CM-AT and -0.6 ± 5.0 for Placebo, a sample size of 82 evaluable low-level subjects per treatment group (active and placebo) yields 90% power to declare a statistically significant difference between treatments at a two-sided 0.05 level of significance. To account for an anticipated 15% early withdrawal, 97 low-level subjects will be randomized per treatment group (active and placebo). A total of 162 subjects per group will be randomized to achieve 97 randomized low-level subjects per group (active and placebo).

8.2 General Statistical Issues

The primary efficacy analysis set is the Low-Level FCT ITT set defined in Section 3. Any efficacy analyses will be presented on other populations from Section 3 will be noted below. The primary safety analysis set is the Low-Level FCT Safety Analysis set. All safety analyses in the double-blind phase will be performed on this analysis set and, secondarily, on the entire Safety Analysis set defined in Section 3.

For quantitative (continuous) variables summaries will include the sample size, mean, median, standard deviation, minimum and maximum. For qualitative (i.e., categorical) variables, summaries will include the number and percent of subjects for each outcome.

Subjects may have a scheduled or unscheduled visit outside the protocol-specified time windows for any protocol-specified visit. Windows of scheduled visits will be expanded to have adjacent start/stop periods so that all visits (scheduled or unscheduled) fall within a scheduled visit window. For example, efficacy data are scheduled to be collected for each subject at screening, Baseline (Screening Part I), Randomization visit, and at Weeks 2, 4, 6, 8, 10 and 12 after randomization. The Week 2 visit will be expanded to allow any visit from 1 – 20 days post-randomization visit to be considered as the Week 2 visit. The Week 4 visit will be expanded to allow any visit from 21 – 34 days post-randomization visit to be considered as the Week 4 visit, etc.

Subjects may have more than one follow-up appointment that falls within a window for a given visit. In such a case, the follow-up data collected closest to the target date the visit should have occurred (given the number of weeks since randomization) will be used in the analysis. If two visits are equidistant from the target date of the scheduled visit, the later visit will be used in the analysis.

Pooling of study centers (i.e., combining study centers with small numbers of patients) may be required prior to performing efficacy analyses that include Study Center as a variable in the analysis. The following is the pooling algorithm that will be used (the algorithm described below is for the Low-Level FCT (≤12.6 U/g) analysis set; the algorithm will be repeated separately for every analysis set on which efficacy data will be analyzed):

If one or more study centers with Low-Level FCT (≤12.6 U/g) subjects have fewer than 10 subjects for the given analysis population, pooling of study centers will be required. Pooling of small study centers will be implemented as follows;

- Order the study centers, starting with those with the lowest numbers of Low-Level FCT (•12.6 U/g) subjects.
- The centers will then be combined in a consecutive upward order unless the total number of subjects exceeds the median number of subjects of all sites with 10 or more subjects.
- If this process does not capture all the study centers with fewer than 10 subjects, a second pooled study center will be created using the same process.
- This process will be continued until all study centers with fewer than 10 subjects are accounted for.

Again, this pooling algorithm will be repeated separately within each analysis population for which efficacy analyses will be run.

Unless otherwise indicated, p-values are two-sided and statistical significance is declared if the two-sided p-value is ≤0.05. There is only one primary efficacy endpoint and one efficacy secondary endpoint. The primary endpoint will be tested at the two-sided 0.05 level of significance, and if significant, the secondary endpoint will be tested at the two-sided 0.05 level of significance. All other endpoints are considered exploratory and there will be no adjustment for multiple comparisons across these exploratory endpoints.

8.3 Subject Disposition

The number and percentage of subjects in each analysis set will be presented. With the exception of the Run-in Analysis Set and the All Randomized Analysis set, percentages will be based on the total number of randomized subjects who received study drug and had at least one post-Baseline ABC-I assessment (the ITT analysis set). For the All Randomized Analysis Set and the Run-in Analysis Set, no percentages will be presented; just the number of subjects will be presented. Randomized subjects without a baseline FCT measurement will not be counted in any Low-Level FCT (≤12.6 U/g) or Non-Low-Level FCT analysis set (>12.6 U/g) but will be included in all other analysis sets in which they reside.

Disposition of all subjects over the course of the double-blind portion of the trial will be presented for the Low-Level FCT ITT analysis set, the ITT analysis set, the Low-Level FCT All Randomized analysis set and the All Randomized analysis set. This presentation will include the number of subjects who completed the study through 12 weeks and the number who discontinued early as indicated on the End of Study Case Report Form page (along with the reason for study discontinuation as indicated on the same page).

The number and percent of subjects, by treatment group and overall, who discontinue during the double-blind phase will be summarized by reason for discontinuation. Percentages are based on the total number of subjects randomized in the given analysis set. The reasons for discontinuation as indicated on the End of Study Case Report Form page are:

- Adverse Event
- Noncompliance
- Protocol violation
- Withdrawal of consent

- Lost to follow-up
- Investigator decision
- Trial terminated by sponsor
- Other

-

8.4 Demographics, Other Baseline Characteristics (other than efficacy and safety variables)

Descriptive statistics of demographic characteristics will include:

- Age, gender (male; female)
- Ethnicity (Hispanic or Latino; not Hispanic or Latino), race (American Indian or Alaska Native; Asian; Black or African American, Native Hawaiian or other Pacific Islander)
- Height (cm)
- Weight (kg)
- Body mass index (BMI)

These characteristics will be summarized using the appropriate descriptive statistics by treatment group and overall for the Low-Level FCT ITT, the ITT, the Low-Level All Randomized the All Randomized analysis sets.

Other Baseline characteristics listed in Section 5 will be summarized using the appropriate descriptive statistics by treatment group and overall for the Low-Level FCT ITT analysis set and for the ITT analysis set.

8.5 Study Drug Usage

The following analyses will be presented for the Low-Level FCT ITT analysis set, the ITT analysis set, the Low-Level FCT All Randomized analysis set and the All Randomized analysis set. For each treatment group compliance will be calculated at each visit as well as for the overall study.

The calculation for percentage of compliance is 100 times the number of sachets dispensed at the previous visit minus the number of sachets returned, divided by the days between visits times 3 (not including days after the last visit at which sachets were dispensed if the subject did not return after that last visit).

If a subject was dispensed sachets at a given visit and then did not return to the study, the dispensation information at this last visit will not be used in the calculation of number of sachets used.

Summary statistics for these variables will include the number of subjects, mean, standard deviation, median, minimum and maximum.

8.6 Efficacy Analyses

The Analyses described below will be performed on every ITT, All Randomized and PP analysis set. **The Low-Level ITT analysis will be considered the primary analysis set.**

8.6.1 Primary Efficacy Analysis

The primary efficacy endpoint is the change from Baseline (score at Screening Part 1) to Week 12 visit in the ABC Irritability/Agitation (ABC-I) in the Low Level FCT ITT population as defined in Section 3. Descriptive statistics of ABC-I and change from Baseline ABC-I will be presented at Week 12 for each treatment group. Mean change in Baseline ABC-I at Week 12 will be compared across treatment groups using a mixed model repeated measures (MMRM) model with terms for treatment, study center, visit, and the treatment by visit interaction. The model will also include Baseline ABC-I score as a covariate and subject will be a random effect. An unstructured covariance matrix will be used for the within-subject correlation. Kenward-Rogers' approximation will be used to estimate degrees of freedom. The SLICE option will be employed to examine differences across treatments at specific time points, with the focus being on the Week 12 change from Baseline. Two-sided 95% confidence intervals of the difference between treatment means of the primary endpoint will be constructed using MMRM model estimates for the mean. Tests of hypotheses will be two-sided and will be based on the contrasts between CM-AT and placebo within the MMRM model.

An assessment of treatment-by-study center interaction on the change from Baseline (score at Screening Part 1) to Week 12 ABC-I will be carried out using analysis of variance (ANOVA). Included in the ANOVA model will be main effects for treatment, investigative center and Baseline ABC-I and the treatment-by-center interaction effect. A 0.10 level of significance will be used to assess the interaction. A non-significant interaction or an interaction significant but only quantitative in nature will support the pooling of results across study centers for the final analysis. For the PP, ITT, and All Randomized analysis sets, a treatment-by-low-level status interaction will be assessed in a similar manner as the treatment-by-study center interaction.

If statistical significance, in favor of CM-AT, is shown on the primary efficacy endpoint (FCT ≤12.6 U/g) in the Low-Level ITT analysis set, the study will be considered a success statistically.

For completeness, though not of primary interest, descriptive statistics of ABC-I and of the change from ABC-I will be presented at each visit, and treatment group comparison p-values on mean change from Baseline ABC-I to each visit will be calculated from the MMRM model discussed above.

Missing ABC-I Data: The primary analyses will be based on the MMRM model with available data discussed above. As a sensitivity analysis, the above analyses will be repeated where, first, multiple imputation using the monotone linear regression approach will be used to impute missing primary endpoint data at each visit via SAS PROC MI; included as covariates in the linear model will be age, sex, race, ABC Irritability measured at previous visits (including the randomization visit), Baseline FCT, and investigative center.

Ten complete datasets will be created (the linear regression approach to multiple imputation assumes monotone missing data pattern; if such a pattern does not exist, then prior to each of the 10 linear regression imputations, non-monotone missing ABC-I data will be imputed via the MCMC approach on ABC-I across visits to create a monotone missing data pattern). Once complete datasets are generated, the primary MMRM model will be applied to each of these datasets with the results will be combined into a single estimate of the treatment difference on the change from Baseline to each visit using SAS PROC MIANALYZE. The focus is on the treatment difference on the change from Baseline to the Week 12 ABC-I as the primary analysis.

As an additional tertiary sensitivity analysis, the Last-Observation Carried Forward (LOCF) method will be applied prior to carrying out the MMRM analysis. For subjects who prematurely withdrew, the last efficacy measurement obtained prior to premature withdrawal (even if the last measurement is the Baseline measurement or the measurement obtained at the randomization visit) will be carried forward to each visit following the withdrawal prior to carrying out the MMRM.

8.6.2 The secondary efficacy endpoint of ABC-L will be analyzed and compared between treatments in a similar manner as the primary endpoint. Again, the Low-Level FCT ITT analysis set is primary, but all ITT, All Randomized and PP analysis sets will be analyzed.

8.6.3 Exploratory Endpoints:

For continuous exploratory endpoints, treatments will be compared on change from Baseline in a similar manner as the primary endpoint at Week 12 and each Post-Baseline visit except for multiple imputation and assessment of treatment-by-study center interaction effects. For CGI-S, an endpoint for which there is no Baseline value, treatments will be compared in a similar manner as the primary endpoint on the CGI-S value itself (as opposed to a change from Baseline) without a Baseline value of the endpoint as a covariate. There will be no adjustment for multiple comparisons.

For dichotomous exploratory endpoints (stool test, weight gain/loss), no formal statistical hypothesis testing will be performed to compare the treatment groups on these results.

The number and percentage of subjects experiencing the dichotomous endpoint will be presented for each treatment group. For completeness, this analysis will be repeated for each visit.

8.7 Safety Analyses

The safety analysis will be performed on the Run-in Safety Analysis Set (run-in phase safety data only) where specified, on the Low-Level FCT Safety Analysis set (for double-blind safety data) and on the entire Safety Analysis set (for double-blind safety data). All subjects receiving at least one dose of randomized treatment is the primary safety analysis set. There will be no imputation of missing data and there will be no adjustment for multiple comparisons.

. **Adverse Events** – The number and percentage of subjects with TEAEs (Treatment Emergent Adverse Events) will be presented overall and by MedDRA System Organ Class (SOC) and Preferred Term (PT). Patients experiencing more than one occurrence of an event within a given SOC and PT will be counted only once within that SOC and PT, respectively. The analysis will be repeated for serious TEAEs, for TEAEs leading to premature withdrawal and for TEAEs deemed at least possibly related to study drug. No formal statistical hypothesis testing will be performed to compare the treatment groups on incidence of TEAEs.

The number and percentage of subjects with TEAEs will be presented by severity of TEAE (mild/moderate/severe) within each SOC and PT. A subject experiencing more than one occurrence of a TEAE within an SOC or

PT will be categorized under the maximum severity in which the subject experienced the TEAE.

The above will be repeated on the run-in TEAEs. The denominator will be the number of patients receiving placebo in the run-in phase, and there will only be one group analyzed (all patients receiving placebo in the run-in phase).

No formal statistical hypothesis testing will be performed to compare the treatment groups on the proportion of subjects with TEAEs.

Physical exams – The number and percentage of subjects with a clinically significant finding on the physical exams will be presented at each Post-Baseline double-blind visit. No formal statistical hypothesis testing will be performed to compare the treatment groups on the proportion of subjects with clinically significant findings.

Fecal Panel - For Fecal Chymotrypsin, pH and A1-Antitrypsin, descriptive statistics (number of subjects, mean, median, standard deviation, minimum and maximum) of Baseline (Screening Part I), each Post-Baseline visit and the change from Baseline at each Post-Baseline visit will be presented. Treatment groups will be compared on change from baseline at each visit using analysis of covariance adjusting for study center and Baseline. For Fecal Chymotrypsin, the number and percentage of subjects with values >12.6 U/g and ≤12.6 U/g at each visit will be presented. For pH, the incidence rate (number and percentage of subjects) with abnormal outcomes will be presented at each visit. For A1-Antitrypsin, the number and percentage of subjects with values <55 and ≥55 will be presented at each visit. Further, for fecal globin and fecal occult blood the incidence rate (number and percentage of subjects) with abnormal outcomes will be presented at each visit, where abnormal is defined as the variable being "present". No formal statistical hypothesis testing will be performed to compare the treatment groups on these results.

Vital signs – For each of the following measurements: weight (kg), systolic blood pressure (mmHg), diastolic blood pressure (mmHg) and pulse (beats per minute), descriptive statistics (number of subjects, mean, median, standard deviation, minimum and maximum) of Baseline, each Post-Baseline visit and the change from Baseline at each Post-Baseline visit will be presented. No formal statistical hypothesis testing will be performed to compare the treatment groups on these results.

Weight Loss by End of Study - The number and percentage of subjects losing weight by the end of the study will be tabulated for each visit for each treatment group. No formal statistical hypothesis testing will be performed

to compare the treatment groups on the proportion of subjects with weight loss.

9. Interim Efficacy Analysis

An Independent Statistical Group (ISG) will conduct one (1) formal interim efficacy analysis on the primary endpoint when at most 50% of the planned number of total Low-Level FCT All Randomized subjects have completed Week 12 evaluations (the exact timing of the interim analysis will depend on enrollment rate; if enrollment is quicker than anticipated, the interim analysis may take place when <50% of the subjects have completed Week 12 evaluations). The Sponsor, participating clinical investigators and any personnel involved in trial conduct will remain blinded to study treatment. The sole purpose of the interim analysis is to ensure proper power of the study.

The ISG will calculate the conditional power for potentially achieving a statistically significant treatment difference at the final analysis, conditioned on the interim results. The conditional power (CP) formula to be used is:

$$CP = \Phi\left(\frac{c_2\sqrt{n_2} - z_1\sqrt{n_1} - \frac{(n_2 - n_1)\Delta}{\sqrt{2\sigma^2}}}{\sqrt{n_2 - n_1}} \right)$$

where Φ is the cumulative distribution function of the standard normal distribution; c_2 is the critical value in the final analysis =1.96 (this critical value corresponds to a one-sided final Type I error rate of 0.025); n_2 is the planned final evaluable sample size per treatment group in the Low-Level FCT All Randomized group (=82); z_1 is the test statistic based on the observed interim data (=z-test statistic for the treatment effect from the MMRM model); n_1 is the sample size per group on which the interim observed rates are based (since the sample sizes may not exactly be equal between treatment groups at this stage despite the 1:1 randomization, the average of the two group sample sizes will be used as an approximation); σ^2 is the assumption of the within-group variance (this will be set to the interim pooled standard deviation between the groups), Δ is the assumed difference in treatment means under which the conditional power is calculated. This will be set to the estimate based on the interim power calculation.

If the conditional power is between 50% and 90%, the ISG may recommend increasing the final Low-Level FCT All Randomized sample size n_2 in the above CP formula to maintain 90% conditional power, following the algorithm in Chen, DeMets, and Lan (2004)[1]. Increasing the sample size to maintain CP of 90% at a one-sided 0.025 level of significance only when the CP is greater than 50% does not inflate the Type I error, as shown in Chen, DeMets and Lan if the sample size

increment is less than 151.3% of the original sample size. The value of 151.3% corresponds to the value of 100*R=100*1.513 in Table IV of Chen, Demets and Lan, where R satisfies:

$$\frac{\sqrt{1+R}\left(\sqrt{1+R}-1\right)}{\sqrt{1+R-t}} = \left(\frac{z_\beta}{z_\alpha}\right)$$

where t is the information fraction at which the interim analysis is carried out (=0.5) based on the original planned final sample size, z_β is the z-critical value corresponding to 90% power (=1.28) and z_α is the z-critical value corresponding to a one-sided 0.025 level of significance (or two-sided 0.05 level of significance) = 1.96.

Again, the above assumes t=0.5; note that if t<0.5, the value of R will be revised accordingly. If the Low-Level FCT All Randomized sample size is increased, the entire randomized sample size will be increased accordingly to ensure the required number of Low-Level FCT All Randomized subjects is met; enrollment will not be specifically focused on Low-Level FCT All Randomized subjects.

10. Software

All analyses, data listings and tables will be performed/produced using SAS® version 9.4 or higher.

[1]

[1] Chen, JYH, DeMets, DL, Gordon Lan, KK, *Increasing the sample size when the unblended interim result is promising, Statist.* Med. 2004; 23:1023-1038 (DOI: 10.1002/sim.1688)

Appendix A

Scoring of the Aberrant Behavior Checklist (ABC)

1. The subscale scores are determined by summing these results.

2. Hyperactivity/Non-compliance:

Item No.
1
7
13
15
18
21
24
28
31
38
39
44
48
51
54
56

3. Irritability/Agitation,

Item No.
2
4
8
10
14
19
25
29
34
36
41
47
50
52
57

Appendix A (cont.)

4. Lethargy/Social Withdrawal

Item No.
3
5
12
16
20
23
26
30
32
37
40
42
43
53
55
58

5. Stereotypic Behavior

Item No.
6
11
17
27
35
45
49

6. Inappropriate Speech

Item No.
9
22
33
46

Appendix B

Scoring of the Social Communication Questionnaire (SCQ)

1. If Question 1 is "yes" then score items 2 through 40. If Question 1 is "no" then score items 8 through 40.
2. For questions 2, 9, and 19 a response of "no" is scored as 1 and "yes" is scored as 0. For all remaining questions a response of "yes" is scored as 1 and "no" is scored as 0.
3. For the Reciprocal Social Interaction domain questions 9, 10, 19, 26-33, 36, 37, 39, and 40 are summed.
4. For Communication domain questions 2-6, 20-25, 34, and 35 are summed.
5. For Restricted, Repetitive, and Stereotyped Patterns of Behavior domain questions 7, 8, and 11-16 are summed.
6. The Total score is the sum of questions 2-40.
7. There is no recommendation by the innovators of the questionnaire regarding the handling of missing data. For this study, it is required that at least 50% of the questions must be answered for each subscale before the subscale can be determined. The same rule will be used for the Total score. For each subscale, missing data for individual questions will be estimated by the average of the remaining questions. Similarly, for the Total score missing data for individual questions will be estimated by the average of the remaining question.

Appendix C

Scoring of the DSM-IV Diagnostic Criteria for Autism (DSM-IV-TR)

1. For Parts A, B, and C a response of "Noted" is scored as a "1" and "Not Noted" is scored as "0".
2. The score for Part A is the sum of the four questions.
3. The score for Part B is the sum of the four questions.
4. The score for Part C is the sum of the four questions.
5. The Total score is the sum of the 12 questions in Parts A, B, and C.
6. The subject is classified as having Autism if the Total score is at least 6, the score for Part A is at least two and at least one for Part B and Part C, at least one of the three questions in Part II is checked as "abnormal", and both questions in Part III are checked "Yes".
7. There is no recommendation by the innovators of the questionnaire regarding the handling of missing data. If any of the questions in Part A, B, and C are missing the corresponding subscales will be considered missing. Additionally, if any of the 12 questions in Parts A, B, and C are missing the Total score will be considered missing. Finally, if any of these 12 questions are missing the Autism assessment will be considered missing unless the criteria for Autism is satisfied by the questions that have been answered.