

Title:

Statistical Analysis Plan for CDISCPILLOT01 – Initial Case Study of the CDISC SDTM/ADaM Pilot Project

Title of Case Study: Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS) in Patients with Mild to Moderate Alzheimer’s Disease.

Authors:

Cathy Barrows, Joel Hoffman, Susan Kenny, Sandy Lei, and Arline Nakanishi for the CDISC SDTM/ADaM Pilot Project Team

TABLE OF CONTENTS

	PAGE
1. BACKGROUND	5
1.1. CDISC SDTM/ADaM Pilot Project.....	5
1.2. Description of Clinical Study.....	5
2. PURPOSE OF THIS ANALYSIS PLAN	6
3. STUDY OBJECTIVE(S) AND ENDPOINT(S)	6
3.1. Study Objective(s).....	6
3.1.1. Primary.....	6
3.1.2. Secondary.....	6
3.2. Study Endpoint(s).....	6
3.2.1. Primary.....	6
3.2.2. Secondary.....	7
3.3. Statistical Hypotheses	7
4. STUDY DESIGN	7
5. SAMPLE SIZE CONSIDERATIONS	8
6. ANALYSIS POPULATIONS	8
7. GENERAL CONSIDERATIONS FOR DATA ANALYSES	8
7.1. Multi-center Studies.....	9
7.2. Examination of Subgroups	9
7.3. Multiple Comparisons and Multiplicity	9
8. DATA HANDLING CONVENTIONS.....	9
8.1. Early Termination and Missing Data	9
8.2. Assessment Windows	10
8.3. Laboratory Data.....	11
8.4. Values of Clinical Concern	11
8.4.1. Laboratory Values.....	11
9. STUDY POPULATION.....	12
9.1. Disposition of Subjects	12
9.2. Demographic and Baseline Characteristics	12
9.3. Treatment Compliance	13
10. EFFICACY ANALYSES	13
10.1. Primary Efficacy Endpoints	13
10.1.1. ADAS-COG (11)	13
10.1.2. CIBIC+	14
10.2. Secondary Efficacy Endpoints.....	14
10.2.1. NPI-X	14
11. SAFETY ANALYSES	14

11.1.	Extent of Exposure	14
11.2.	Adverse Events	14
11.3.	Deaths and Serious Adverse Events.....	15
11.4.	Adverse Events Leading to Discontinuation of Investigational Product and/or Withdrawal from the Study and Other Significant Adverse Events	15
11.5.	Clinical Laboratory Evaluations	16
11.6.	Other Safety Measures	16
12.	REFERENCES	17
13.	ATTACHMENTS	17
13.1.	Table of Contents for Data Display Specifications	17
13.1.1.	Tables	17
13.1.2.	Figures	18
13.1.3.	General Comments for Data Displays	18
13.2.	Templates for Data Displays	18
14.	APPENDIX 1 – ADDITIONAL STATISTICAL DETAILS.....	44
14.1.	Scoring algorithms for Efficacy Endpoints.....	44
14.1.1.	ADAS-Cog(11)	44
14.1.2.	CIBIC+	45
14.1.3.	NPI-X	45
14.2.	Handling missing item scores within efficacy data	46
15.	APPENDIX 2 – DEVIATIONS FROM PROTOCOL-SPECIFIED ANALYSES	47
15.1.	Deviations from Protocol-Specified Efficacy Analyses	47
15.2.	Deviations from Protocol-Specified Pharmacokinetic Analyses	47
15.3.	Deviations from Protocol-Specified Safety Analyses	47
15.4.	OMISSIONS FROM LEGACY DATA	48
15.5.	Coding of Adverse Events.....	48
16.	APPENDIX 3 - ANALYSES NOT PRE-SPECIFIED IN PROTOCOL.....	49
16.1.	CIBIC+.....	49

List of Abbreviations

AD	Alzheimer's Disease
ADaM	Analysis Dataset Model Team
ADAS-Cog	Alzheimer's Disease Assessment Scale - Cognitive Subscale
AE	adverse event
ANCOVA	analysis of covariance
ANOVA	analysis of variance
BMI	body mass index
CDISC	Clinical Data Interchange Standards Consortium
CIBIC+	Video-referenced Clinician's Interview-based Impression of Change
cm ²	centimeters squared – measure of area
CMH	Cochran-Mantel-Haentzel
DAD	Disability Assessment for Dementia
ECG	Electrocardiogram
ET	Early Termination visit
FDA	Food and Drug Administration
ITT	Intent-to-treat
PT	Preferred Term
LOCF	last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligrams
MMRM	Mixed-effects Model Repeated-Measure
MMSE	Mini-Mental State Examination
NINCDS-ADRDA	National Institute of Neurologic and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (developed criteria for the diagnosis of Alzheimer's disease)
NPI-X	Revised Neuropsychiatric Inventory
RT	Retrieval visit
SDTM	Study Data Tabulation Model
SOC	System Organ Class
TTS	Transdermal Therapeutic System
XAN	Xanomeline

1. BACKGROUND

1.1. CDISC SDTM/ADaM Pilot Project

The CDISC SDTM/ADaM Pilot Project team will produce a case study to demonstrate the effective transformation of legacy data into CDISC SDTM domains and ADaM datasets and their associated metadata. The resulting “pilot submission” will be delivered to FDA reviewers for their evaluation in a mock review, assessing whether data submitted to the FDA using the CDISC Standard will meet the needs and expectations of both medical and statistical FDA reviewers.

The pilot submission will include SDTM datasets, analysis datasets, all relevant metadata, analysis results, and an abbreviated report (including only the necessary documentation).

The legacy data being used in CDISCPIL0T01 were provided by Eli Lilly and Company (Legacy Sponsor) for the purposes of this pilot project. The data were de-identified and documents were redacted prior to release to the pilot project team.

The submission will not reproduce all of the Legacy Sponsor’s analyses and reports. Instead only the more common elements of a submission will be addressed. These will include safety data, the primary outcome, and at least one secondary outcome. A representative set of analyses will be chosen. Deviations from the protocol-specified analyses are described in [Appendix 1](#). Additional variables and flags may be included in the analysis datasets, but may not be used in the analyses included in the report.

1.2. Description of Clinical Study

Study Title: Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS) in Patients with Mild to Moderate Alzheimer’s Disease.

The study was a prospective, randomized, multi-center (17), double-blind, placebo-controlled, parallel-group study. The objectives of the study were to evaluate the efficacy and safety of transdermal xanomeline, 50 cm² and 75 cm², and placebo in patients with mild to moderate Alzheimer’s disease.

Xanomeline or placebo was administered daily in the morning, with the application of two adhesive patches, one 50 cm² in area, the other 25 cm² in area. Doses were measured in terms of the xanomeline base, and were 54mg for the 50 cm² patch and 27mg for the 25 cm² patch. Placebo was identical in appearance to the primary study material. The total doses being compared are therefore 0 (both patches placebo), 54mg (large patch active drug, small patch placebo), and 81mg (both patches active drug). The treatment groups referred to throughout the pilot submission will be “xanomeline high dose,” “xanomeline low dose,” and “placebo”.

Patients were males or females of non-childbearing potential, 50 years of age or older, had probable Alzheimer's disease according to the NINCDS-ADRDA criteria, and an MMSE score of 10 to 23. The duration of treatment was 26 weeks, with 24 weeks of active treatment. A total of 295 patients were randomized into 1 of 3 treatment groups: xanomeline high dose, 97 patients; xanomeline low dose, 98 patients; and placebo, 100 patients; 166 were females and 129 were males.

2. PURPOSE OF THIS ANALYSIS PLAN

This analysis plan describes the analyses to be performed in the context of the first iteration of the CDISC SDTM/ADaM Pilot Submission, CDISCPIL0T01. It should be noted that this document is not meant to represent all of the measures assessed or analyses performed in the original study.

3. STUDY OBJECTIVE(S) AND ENDPOINT(S)

3.1. Study Objective(s)

3.1.1. Primary

The primary objectives of this study are

- To determine if there is a statistically significant relationship (overall Type 1 error rate, $\alpha=.05$) between the change in both the ADAS-Cog (11) and CIBIC+ scores, and drug dose (0, 50 cm² [54 mg], and 75 cm² [81 mg]).
- To document the safety profile of the xanomeline TTS.

3.1.2. Secondary

A secondary objective of this study is:

- To assess the dose-dependent improvement in behavior. Improved scores on the Revised Neuropsychiatric Inventory (NPI-X) will indicate improvement in these areas.

3.2. Study Endpoint(s)

3.2.1. Primary

- Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)] at Week 24
- Video-referenced Clinician's Interview-based Impression of Change (CIBIC+) at Week 24

3.2.2. Secondary

Secondary Efficacy Endpoints

- Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)] at Weeks 8 and 16
- Video-referenced Clinician's Interview-based Impression of Change (CIBIC+) at Weeks 8 and 16
- Mean Revised Neuropsychiatric Inventory (NPI-X) from Week 4 to Week 24

Safety Endpoints

- Adverse events
- Vital signs (weight, standing and supine blood pressure, heart rate)
- Laboratory evaluations

3.3. Statistical Hypotheses

The statistical hypotheses for the 2 primary endpoints are based on the primary analysis, which is a test for dose response. The primary analysis for ADAS-Cog (11) at Week 24 is based on an ANCOVA model, which includes the baseline score, site, and treatment as continuous variable. The statistical hypothesis is:

$H_0: b = 0$, where b is the coefficient for treatment

$H_1: b \neq 0$, where b is the coefficient for treatment

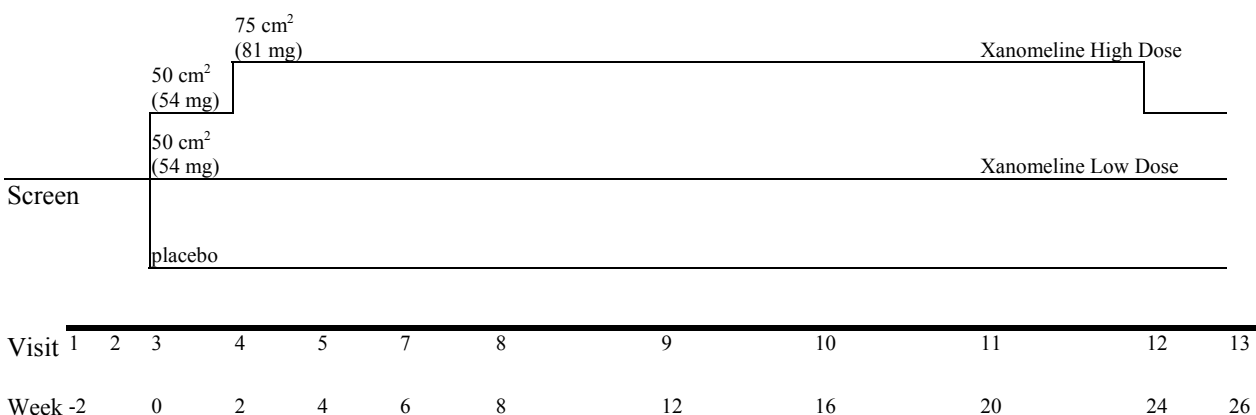
The primary analysis for CIBIC+ at Week 24 is based on an ANOVA model which includes site and treatment as continuous variable. The statistical hypothesis is:

$H_0: b = 0$, where b is the coefficient for treatment

$H_1: b \neq 0$, where b is the coefficient for treatment

4. STUDY DESIGN

Patients with probable mild to moderate AD will be studied in a randomized, double-blind, parallel (3 arm), placebo-controlled trial of 26 weeks duration. The study will be conducted on an outpatient basis. Approximately 300 patients will be enrolled.



5. SAMPLE SIZE CONSIDERATIONS

Approximately 100 patients will be randomized to each of the 3 treatment groups. Previous experience with the oral formulation of xanomeline suggests that this sample size has 90% power to detect a 3.0 mean treatment difference in ADAS-Cog ($p < .05$, two-sided), based on a standard deviation of 6.5. Furthermore, this sample size has 80% power to detect a 0.36 mean treatment difference in CIBIC+ ($p < .05$, two-sided), based on a standard deviation of 0.9.

6. ANALYSIS POPULATIONS

For this study, the following definitions are used:

Screen Failures	Patients entered into the study are those from whom informed consent for the study has been obtained. Patients entered into the study but not assigned to a treatment group are considered to be screen failures. Demographic data for screen failures will be included in the data tabulation datasets, but not in the analysis datasets or in the analyses.
Randomized	Patients who are enrolled in the study are those who have been assigned to a treatment group. Patients who are entered into the study but fail to meet criteria specified in the protocol for treatment assignment will not be enrolled in the study. Patients are randomly assigned to treatment groups at Week 0 (Visit 3).
ITT Population	All patients randomized
Safety population	All patients randomized and known to have taken at least one dose of randomized drug
Efficacy population	All patients who were randomized and took drug, and have at least 1 post-baseline measure for both ADAS-Cog and CIBIC+
Completers	All patients in the efficacy population who completed their Week 24 visit (Visit 12).

7. GENERAL CONSIDERATIONS FOR DATA ANALYSES

All statistical tests will be 2-sided with a significance level of 0.05. One-sided p-values will not be reported. Summary statistics for continuous variables will include the number of observations, mean, standard deviation, median, minimum, and maximum. Summary statistics for categorical variables will include the frequency and percentage.

7.1. Multi-center Studies

Sites that enroll fewer than 3 patients in any one treatment group will be grouped together, with a new pooled site identifier assigned for the purpose of analysis. If this combination still results in a treatment group having fewer than 3 patients in any one treatment group, then this group of patients will be combined with the next fewest enrolling site. In the event that there is a tie for fewest-enrolling site, one of these will be chosen at random by a random-number generator.

7.2. Examination of Subgroups

The effect of gender upon efficacy will be evaluated if sample sizes are sufficient to warrant such analyses.

Covariates for age, race, baseline disease severity as measured by MMSE, and patient education level will be included in analysis datasets as appropriate for exploratory analysis.

7.3. Multiple Comparisons and Multiplicity

There are 2 primary efficacy endpoints in this study, each of which will be tested at a significance level of 0.05. No adjustments for multiple primary endpoints will be made. Nominal 2-sided p-values will be reported for each primary efficacy endpoint.

The primary analyses for the 2 primary efficacy endpoints will be a test for dose response. Within each primary efficacy endpoint, 3 sets of pairwise comparisons for the 2 active treatment groups and placebo will only be performed if the test for dose response is significant.

8. DATA HANDLING CONVENTIONS

“End of treatment” will refer to the subject’s Week 24 visit or early termination visit.
“End of study” will refer to the subject’s Week 26 visit or early termination visit.

8.1. Early Termination and Missing Data

If possible, data for subjects who terminated the study early will be collected at the scheduled Week 24 visit.

Laboratory values collected after the discontinuation of study drug will be used. The assessment date will be compared with the last dose of study drug to determine if the assessment was made within the defined washout period and a flag will be set to indicate this status.

Missing postbaseline efficacy values will be imputed using last observation carried forward (LOCF). Missing baseline values and missing safety data will not be imputed. For the efficacy data, the last observation carried forward will be based on the targeted assessments (i.e. those assigned to be the analyzable assessment based on the assessment windows).

8.2. Assessment Windows

In general, assessments will be assigned to visits as collected on the CRFs, and will disregard the actual date of the assessment. For example, if an assessment is recorded on the Visit 10 CRF page, the assessment will be assigned to Week 16 (Visit 10).

The ADAS-Cog (11), CIBIC+, and NPI-X assessments will also be assigned to visits based on the actual visit dates, as will laboratory assessments. Actual visit days will be determined relative to the date of randomization, using the algorithm {day = visit date – randomization date}. If multiple assessments fall into the same visit window (windows defined in following table), then the assessment closest to the target day will be selected. [Note that retrieval visits (visit number 201) are included for the purpose of selecting assessments for the week 24 visit window.] If two assessments are equidistant from the target day, then the assessment prior to the target day will be selected. In situations where imputation of missing values is also involved, imputation will use the targeted assessments within the windows.

Variable	Scheduled Visit	Time Interval (label on output)	Time Interval (Day)	Target Time Point (Day)
ADAS-Cog CIBIC+	3	Baseline	≤ 1	1
	8	Week 8	2-84	56
	10	Week 16	85-140	112
	12	Week 24	>140	168
NPI-X	3	Baseline	≤ 1	1
	4	Week 2	2-21	14
	5	Week 4	22-35	28
	7	Week 6	36-49	42
	8	Week 8	50-63	56
	8.1	Week 10 (Tel)	64-77	70
	9	Week 12	78-91	84
	9.1	Week 14 (Tel)	92-105	98
	10	Week 16	106-119	112
	10.1	Week 18 (Tel)	120-133	126
	11	Week 20	134-147	140
	11.1	Week 22 (Tel)	148-161	154
	12	Week 24	162-175	168
	13	Week 26	>175	182

8.3. Laboratory Data

Multiple laboratory assessments within visit/week windows may be collected. In part, this will be a function of safety monitoring procedures as described in the protocol. Additional and unscheduled labs may also be collected for other reasons. Only planned laboratory values collected at scheduled visits are used for analysis. Additional lab values will be included in datasets, but not incorporated into analyses. A flag will be set to indicate the last on-treatment (prior to or at week 24) observation for each lab parameter.

A change from baseline laboratory value will be calculated as the difference between the baseline lab value and the endpoint value (i.e., the value at the specified visit) and the end of treatment observation.

In addition, each laboratory value, including the baseline value, will be categorized with reference to the lab normal range as

- “L” - less than or equal to the lower limit of normal
- “N” – Greater than the Lower Limit of Normal and less than the Upper Limit of Normal
- “H” – Greater than or equal to the Upper Limit of Normal.

Laboratory values will be assigned a flag of abnormal (high or low) if the value is outside the threshold range (defined as significantly beyond the normal range, i.e., $>1.5 \times \text{ULN}$ or $<.5 \times \text{LLN}$) or if the value is significantly different from the value observed at the preceding scheduled visit (i.e., absolute value of the change from previous value is larger than the 50% of the normal range, $\text{ULN} - \text{LLN}$). In addition, the ratio of the value to its LLN (i.e., value/LLN) and to its ULN (i.e., value/ULN) will be calculated for inclusion in the datasets, but not analyzed. These ratios will allow a quick searching for subjects with values greater than $1.5 \times \text{ULN}$, for example, without the use of additional flags.

A separate analysis dataset will be provided containing the lab parameters needed for the assessment of Hy's Law. An abnormal flag will be assigned to indicate whether a subject meets the criteria for a modified Hy's Law assessment of liver function, defined as:

1. Transaminase (SGPT/ALT or SGOT/AST) elevations (i.e., $>1.5 \times \text{ULN}$) and
2. Bilirubin elevated to greater than $1.5 \times \text{ULN}$.

8.4. Values of Clinical Concern

8.4.1. Laboratory Values

Laboratory values will be assigned a flag of abnormal if the value is significantly beyond the normal range or if the value is significantly different from the proceeding value.

In addition, elevated liver function values will be used to trigger retesting and possible discontinuation of study drug as described in the protocol. These values of clinical concern are:

- Patients with ALT/SGPT levels >120 IU

- Patients with ALT/SGPT values >400 IU, or alternatively, an elevated ALT/SGPT accompanied by GGT and/or ALP values >500 IU

9. STUDY POPULATION

9.1. Disposition of Subjects

The number of subjects randomized, number of subjects in the ITT population, number of subjects in the safety population and number of subjects in the efficacy population will be summarized by treatment group. The number and percentage of subjects who complete the study as well as subjects who withdraw prematurely from the study will be displayed. The reasons for early termination will be summarized.

Fisher's exact test will be used to analyze 3 reasons for study discontinuation (protocol completed, lack of efficacy, and adverse event).

9.2. Demographic and Baseline Characteristics

The following will be summarized by treatment group and across all treatment groups.

- Age
- Age category (<65, 65-80, >80)
- Sex
- Race
- Mini-Mental State
- Duration of disease [computed as months between date of Week -2 (Visit 1) and date of onset of the first definite symptoms of Alzheimer's Disease]
- Years of education
- Weight, height, BMI at Baseline (Visit 3 for weight and BMI, Visit 1 for height)
- BMI category (BMI<25, BMI 25-<30, BMI≥30)

The treatment groups will be compared by analysis of variance (ANOVA) for continuous variables and by Pearson's chi-square test for categorical variables. Note that because patients are randomized to 1 of the 3 treatment groups, any statistically significant treatment group differences are by definition a Type I error; however, the resulting p-values will be used as another descriptive statistic to help focus possible additional analyses (for example, analysis of covariance, subset analyses) on those factors that are most imbalanced (that is, that have the smallest p-values).

Baseline comparisons across treatment groups for the Mini-Mental State Examination (MMSE) will be made using analysis of variance with treatment and site as main effects.

9.3. Treatment Compliance

Treatment compliance will not be provided.

10. EFFICACY ANALYSES

Refer to Section [15.1 \(Deviations from Protocol-Specified Efficacy Analyses\)](#) for a description of how these analyses differ from the protocol-specified efficacy analyses.

A key difference from the protocol is that efficacy assessments are considered valid even if they occur while off study drug.

10.1. Primary Efficacy Endpoints

10.1.1. ADAS-COG (11)

The primary analysis of the ADAS-Cog (11) at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24. A secondary analysis will be performed for the Week 24 endpoint using the completers subset using observed data. For each of these analyses, an ANCOVA model will be used with the baseline score, site, and treatment included as independent variables. Treatment will be included as a continuous variable, and results for a test of dose response will be produced. Interaction terms will not be investigated. If the test for dose response is statistically significant, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.

Summary statistics will be generated for each visit including baseline using the efficacy population with LOCF imputation. The visits for ADAS-Cog (11) are baseline (Week 0), Week 8, Week 16, and Week 24.

A supportive analysis for the ADAS-Cog (11) will use a likelihood-based repeated measures (MMRM) analysis. In this analysis for the change from baseline in the ADAS-Cog (11) at Week 24, the independent variables included in the model are the fixed, categorical effects of treatment, site, time (week), and treatment by time interaction along with the continuous effects of baseline ADAS-Cog (11) score and baseline ADAS-Cog (11) score by time interaction. Barring a computational singularity, an unstructured covariance matrix will be used to model the within-subject errors in the MMRM analysis. The unstructured covariance matrix is chosen to allow the analysis to be unconstrained by the structure of the covariance. If there is any computational singularity, a Toeplitz covariance matrix will be used. The Toeplitz covariance structure provides reasonable flexibility in estimating correlation between visits.

Additional details regarding scoring and methods for handling missing data for ADAS-Cog (11) are in [Appendix 1](#) of this analysis plan.

10.1.2. CIBIC+

The primary analysis of CIBIC+ at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24. For this endpoint, an ANOVA model will be used with site and treatment included as independent variables. Interaction terms will not be investigated. Treatment will be included as a continuous variable, and results for a test of dose response will be produced. If the test for dose response is statistically significant, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.

Summary statistics will be generated for each visit using the efficacy population with LOCF imputation. The visits for CIBIC+ are Week 8, Week 16, and Week 24.

Additional details regarding scoring for CIBIC+ are in [Appendix 1](#) of this analysis plan.

10.2. Secondary Efficacy Endpoints

10.2.1. NPI-X

The primary analysis of mean NPI-X total score from Week 4 to Week 24 will use the efficacy population. This endpoint will be calculated as the mean of all available total scores between Weeks 4 and 24, inclusive. For this endpoint, an ANCOVA model will be used with the baseline score, site, and treatment included as independent variables. Interaction terms will not be investigated. Treatment will be included as a continuous variable, and results for a test of dose response will be produced. If the test for dose response is statistically significant, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.

The visits for NPI-X are baseline (Week 0), Week 2, Week 4, Week 6, Week 8, Week 10 (telephone), Week 12, Week 14 (telephone), Week 16, Week 18 (telephone), Week 20, Week 22 (telephone), Week 24, Week 26.

Additional details regarding scoring and methods for handling missing data for NPI-X are in [Appendix 1](#) of this analysis plan.

11. SAFETY ANALYSES

11.1. Extent of Exposure

Average daily dose and cumulative dose at end of study (Week 26 or early termination) will be computed for each subject. Summary statistics will be computed for each of the above quantities for each treatment group.

11.2. Adverse Events

For this submission, the adverse events will be recoded according to MedDRA. Treatment emergent adverse events will be cross-tabulated by System Organ Class (SOC)

and preferred term (PT). Please refer to [Appendix 15.5](#) for additional information about the MedDRA coding. The incidence of treatment emergent events grouped under preferred terms for each active treatment will be compared to placebo using Fisher's exact test. Treatment emergent adverse events are defined relative to the date of first dose [Week 0 (Visit 3) unless indicated otherwise] as

- events with a start date that is equal to or greater than the date of first dose
- events that start prior to the date of first dose and worsen after that date
- events that start and resolve prior to the date of first dose, but then recur after that the date of first dose.

If the recording of an adverse event start date is not complete, imputation of the start date will be done in a conservative manner. Adverse events will be considered treatment emergent if the year and/or the month is the same as the treatment start year and month. In the case of a completely missing adverse event start date, the start date will be imputed as the day of first dose. No imputation of adverse events dates where the partial date clearly indicates a start prior to the beginning of treatment will be done.

Due to the formulation of the clinical path, it is anticipated that there may be an increase in adverse events that are associated with the application of the skin path. For this reason, additional analysis of dermatological adverse events will be conducted. A category of special events will be created to identify the events that are considered dermatological events. These events will be determined by the medical review of blinded coded adverse event terms and all preferred terms that are considered to be dermatologic in nature, such as rash, pruritus, or dermatitis, will be flagged as adverse events of special interest. A complete list of preferred terms that are considered to be dermatologic events will be provided in the final analysis. The time to the first dermatological event will be compared across the treatment groups using Kaplan-Meier methods. Graphical displays of the survival curves will be presented.

11.3. Deaths and Serious Adverse Events

Serious adverse events (SAEs) will be summarized by SOC and PT. Incidence of SAEs will be compared between active drug groups and placebo, again using Fisher's exact test.

11.4. Adverse Events Leading to Discontinuation of Investigational Product and/or Withdrawal from the Study and Other Significant Adverse Events

For this submission, no formal summarization of adverse events leading to discontinuation or withdrawal from the study will be conducted. The analysis data will provide variables to identify these adverse events.

11.5. Clinical Laboratory Evaluations

Hematology, and clinical chemistry will be summarized for Baseline and Weeks 2, 4, 6, 8, 12, 16, 20, 24, and 26 (Visits 1, 4, 5, 7, 8, 9, 10, 11, 12, and 13, respectively). Urinalysis and other lab data will not be summarized, but will be included in the tabulation datasets. The baseline values will be those collected at Week -2 (Visit 1).

Four assessments of abnormality will be identified for each laboratory analyte, as described in [Section 8.4.1](#):

- Values outside the normal range
- Values significantly beyond the normal range (i.e., outside the threshold range)
- Values differing significantly from values at the previous scheduled visit,
- Abnormal values as defined by Hy's Law

The number of subjects with no abnormal measure during treatment and those with at least one abnormal measure during treatment will be summarized for each lab analyte. Two tables will be provided – one defining abnormal as beyond normal range (i.e, below LLN or above ULN) and the other defining abnormal as a clinically significant change from the previous visit. Fisher's exact test will be used to analyze the incidence of abnormal (high or low) measures during the post-randomization phase.

A display summarizing shifts from baseline by week in terms of abnormality based on threshold range will be provided. The data will be summarized using sets of 3x3 matrices comparing baseline and on drug categorization for each treatment group for each week for each laboratory analyte.

Shift tables summarizing whether or not a subject's status changed from baseline during the treatment period will be provided for changes based on threshold ranges and changes based on Hy's Law. In these tables a subject will be categorized as normal or abnormal (i.e., outside the threshold range) at baseline. During the treatment phase, the most extreme value will be used to categorize a patient as normal or abnormal during the treatment phase. The shift table will show the number of patients whose on treatment categorization was the same or shifted from the baseline categorization. The treatment period is defined as any planned visit after Week 0 (Visit 3), up to and including Week 24 (Visit 12). A Cochran-Mantel-Haenszel (CMH) test, stratifying by status at baseline, will be performed.

11.6. Other Safety Measures

Vital sign data (blood pressure supine, blood pressure standing 1 minute, blood pressure standing 3 minutes, heart rate supine, heart rate standing 1 minute, and heart rate standing 3 minutes) at baseline and Week 24 and end of treatment (last visit on or before Week 24 visit) will be summarized by treatment group. Change from baseline will also be summarized.

Weight data at baseline and Week 24 (with and without including early terminations) will be summarized by treatment group. Change from baseline will also be summarized.

The concomitant medication data will be coded using a publicly available sample of WHO Drug. Drugs not matching those in the sample will be considered “uncoded” for the purposes of this submission. The number and percent of subjects receiving each concomitant medication will be summarized. Concomitant medications will be reported by Body System and ingredient. Medications will be sorted in descending order of total incidence across treatment groups for the Body System and in descending order of total incidence for the ingredient within each Body System. If the total incidence for any two or more ingredients is equal, the events will be presented in alphabetical order.

12. REFERENCES

13. ATTACHMENTS

13.1. Table of Contents for Data Display Specifications

13.1.1. Tables

1. Summary of Populations ([Template 1](#))
2. Summary of End of Study Data ([Template 2](#))
3. Summary of Demographic and Baseline Characteristics ([Template 3](#))
4. Summary of Number of Subjects by Site ([Template 4](#))
5. Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 – LOCF ([Template 5](#))
6. Primary Endpoint Analysis: CIBIC+ - Summary at Week 24 – LOCF ([Template 6](#))
7. ADAS Cog (11) - Change from Baseline to Week 8 – LOCF ([Template 5](#))
8. CIBIC+ - Summary at Week 8 – LOCF ([Template 6](#))
9. ADAS Cog (11) - Change from Baseline to Week 16 – LOCF ([Template 5](#))
10. CIBIC+ - Summary at Week 16 – LOCF ([Template 6](#))
11. ADAS Cog (11) - Change from Baseline to Week 24 – Completers at Week 24 - Observed Cases-Windowed ([Template 7](#))
12. ADAS Cog (11) - Change from Baseline to Week 24 in Male Subjects – LOCF ([Template 8](#))
13. ADAS Cog (11) - Change from Baseline to Week 24 in Female Subjects – LOCF ([Template 8](#))
14. ADAS Cog (11) - Mean and Mean Change from Baseline over Time ([Template 9](#))
15. ADAS Cog (11) – Repeated Measures Analysis of Change from Baseline to Week 24 ([Template 10](#))

14. APPENDIX 1 – ADDITIONAL STATISTICAL DETAILS

14.1. Scoring algorithms for Efficacy Endpoints

Described below are details for scoring the ADAS-Cog (11), CIBIC+, and NPI-X, and details on how to handle missing data.

14.1.1. ADAS-Cog(11)

ADAS-Cog Scoring Methods

Item No.	Description	Score Range
ITEM01	Word Recall Task	0-10
ITEM02	Naming Objects and Fingers	0-5
ITEM03 ^a	Delayed Word Recall	0-10
ITEM04	Commands	0-5
ITEM05	Constructional praxis	0-5
ITEM06	Ideational praxis	0-5
ITEM07	Orientation	0-8
ITEM08	Word recognition	0-12
ITEM09 ^a	Attention/Visual Search Task	0-5 Based on number of correct answers: >30 = 0 24-30 = 1 18-23 = 2 12-17 = 3 6-11 = 4 0-5 = 5
ITEM10 ^a	Maze Solution	0-5 Based on time (in seconds) and number of errors: If <2 errors then: 0-30 seconds = 0 31-60 = 1 61-90 = 2 91-120 = 3 121-239 = 4 ≥ 240 = 5 If 2 or more errors then the score is a 5
ITEM11	Spoken Language Ability	0-5
ITEM12	Comprehension of Spoken Language	0-5

ITEM13	Word Finding Difficulty in Spontaneous Speech	0-5
ITEM14	Recall of Test Instructions	0-5
TOT01	ADAS-Cog (11)	0-70
TOT02	ADAS-Cog (14)	0-90

^a Additional items for the ADAS-Cog (14) but not part of the ADAS-Cog (11).

14.1.2. CIBIC+

This assessment has a 7-point scale and is scored as:

- 1 = Marked improvement,
- 2 = Moderate improvement,
- 3 = Minimal improvement,
- 4 = No change,
- 5 = Minimal worsening,
- 6 = Moderate worsening,
- 7 = Marked worsening

14.1.3. NPI-X

The primary assessment of this instrument will be for the total score, not including the sleep, appetite, and euphoria domains. This total score is computed by taking the product of the frequency and severity scores and summing them up across the domains.

Severity:

Range 1-3

1 = mild, 2 = moderate, 3 = marked

Frequency:

Range 1-4

1 = occasionally, 2 = often, 3 = frequently, 4 = very frequently

Can be treated as continuous variables

Frequency × Severity for each NPI domain

Range 0-12

NPI-X Total (9) will be calculated as the sum of all individual domain scores (can be treated as continuous variable). If the domain is absent, then the score for the domain is 0. If the domain is not applicable then the score for the domain is set to missing. The range of NPI-X Total (9) is 0-108.

NPI-X Total (9) domains are:

- Delusions
- Hallucinations
- Agitation/Aggression
- Depression/Dysphoria
- Anxiety
- Apathy/Indifference
- Disinhibition
- Irritability/Lability
- Aberrant Motor Behavior

14.2. Handling missing item scores within efficacy data

The following applies to all totals and subtotals of ADAS-Cog(11) and NPI-X, and does not apply to CIBIC+.

Any computed total score will be treated as missing if more than 30% of the items are missing or scored “not applicable”. For example, when computing ADAS-Cog(11), if 4 or more items are missing, then the total score will not be computed. When one or more items are missing (but not more than 30%), the total score will be adjusted in order to maintain the full range of the scale. For example, ADAS-Cog(11) is a 0-70 scale. If the first item, Word Recall (ranges from 0 to 10), is missing, then the remaining 10 items of the ADAS-Cog(11) will be summed and multiplied by $(70 / (70-10))$, or 7/6.

15. APPENDIX 2 – DEVIATIONS FROM PROTOCOL-SPECIFIED ANALYSES

Some analyses specified in the original protocol will not be performed for the purposes of this pilot project. Where applicable, deviations are noted in the appropriate sections of this analysis plan. Otherwise, deviations from the protocol-specified analyses are described below.

15.1. Deviations from Protocol-Specified Efficacy Analyses

The following efficacy endpoints will not be used: ADAS-Cog (14) and DAD. ANOVA and ANCOVA models for the efficacy endpoints will not assess site*treatment interaction. Furthermore, the normality assumption for the efficacy endpoints will not be investigated and consideration for rank transformations will not be done.

The protocol proposes a number of secondary analyses for the efficacy endpoints. The following secondary analyses will not be performed:

- Observed cases at each timepoint for ADAS-Cog (11) and CIBIC+.
- Average of all postrandomization NPI-X scores including Weeks 2 and 26.
- Dichotomizing subjects for each behavior in the NPI-X instrument into those who experienced the behavior for the first time postrandomization and those who had the quotient between frequency and severity increase relative to baseline versus those who did not.

The protocol states that efficacy assessments are invalid if no study drug has been taken within 3 days prior to the assessment. This will not be considered in the pilot project. In addition, efficacy assessments occurring after the last dose of drug will be considered for windowing and for the efficacy analysis if they are collected at visit number 201.

Interim analyses will not be performed.

Covariate analyses examining the effect of Apo E on the efficacy measures will not be performed.

15.2. Deviations from Protocol-Specified Pharmacokinetic Analyses

Pharmacokinetic analyses will not be performed.

15.3. Deviations from Protocol-Specified Safety Analyses

ECG analyses will not be performed.

The caregiver's response about the patch will not be summarized.

No ANCOVA analyses will be performed for laboratory data. Instead, frequency tables based on the on-treatment period will include p-values.

15.4. OMISSIONS FROM LEGACY DATA

The data reflected in the submitted datasets will not include all of the subjects in the legacy data. This is because we do not have all of the data for the remaining subjects, so chose to omit them as we have an adequate number of subjects left for the purposes of the pilot.

The lab data included many analytes. Because of the large size of the datasets, it was decided to reduce these datasets by dropping less common analytes from the datasets. This was done by comparing the lab tests performed to a list of common lab tests found on the CDISC web site. Lab tests not in this list of common tests were dropped. In addition, lab tests with character results only were dropped, to simplify the analyses provided. These include “RBC Morphology,” “elliptocytosis,” “basophilic stippling,” and “target cells.”

15.5. Coding of Adverse Events

Due to licensing restrictions for MedDRA, all of the event terms, including verbatim text, LLT, PT, HLT, HLGT, and SOC were initially masked. Discussions with MSSO resulted in an agreement that all can be unmasked with the exception of HLT and HLGT, as long as text is what is provided and not the actual MedDRA numeric code. In the processing of providing the unmasking, it was simpler to also leave the verbatim text masked. Consequently, the SOC, PT, and LLT are real (i.e., not masked).

16. APPENDIX 3 - ANALYSES NOT PRE-SPECIFIED IN PROTOCOL

16.1. CIBIC+

At the request of the FDA reviewers, treatments will also be compared for overall differences in the CIBIC+ by using the CMH test, controlling for site. The template for this result is [Ad hoc Template 1](#).

Protocol: CDISCPILLOT01

Page 1 of n

Population: Efficacy

Ad hoc Template 1
CIBIC+ - Categorical Analysis - LOCF

	Placebo (N=xxx)	Xanomeline Low Dose (N=xxx)	Xanomeline High Dose (N=xxx)	p-value [1]
Week 8				
N	xx	xx	xx	x.xxx
Marked improvement	xx (xx%)	xx (xx%)	xx (xx%)	
Moderate improvement	xx (xx%)	xx (xx%)	xx (xx%)	
Minimal improvement	xx (xx%)	xx (xx%)	xx (xx%)	
No Change	xx (xx%)	xx (xx%)	xx (xx%)	
Minimal worsening	xx (xx%)	xx (xx%)	xx (xx%)	
Moderate worsening	xx (xx%)	xx (xx%)	xx (xx%)	
Marked worsening	xx (xx%)	xx (xx%)	xx (xx%)	
Week 16				
N	xx	xx	xx	x.xxx
Marked improvement	xx (xx%)	xx (xx%)	xx (xx%)	
Moderate improvement	xx (xx%)	xx (xx%)	xx (xx%)	
Minimal improvement	xx (xx%)	xx (xx%)	xx (xx%)	
No Change	xx (xx%)	xx (xx%)	xx (xx%)	
Minimal worsening	xx (xx%)	xx (xx%)	xx (xx%)	
Moderate worsening	xx (xx%)	xx (xx%)	xx (xx%)	
Marked worsening	xx (xx%)	xx (xx%)	xx (xx%)	

repeat above for week 24

[1] Overall comparison of treatments using CMH test (Pearson Chi-Square), controlling for site.