STATISTICAL ANALYSIS PLAN

D1002001

A Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study of SM-13496 for the Treatment of Bipolar I Depression

AUTHOR: [redacted]

VERSION NUMBER AND DATE: VX.X, **DDMMMYYYY**30MAR2017SPONSOR NAME

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V1.0 (Dated 30MAR2017) for Protocol D1002001

	Name	Signature	Date
Author:	[redacted]		
Position:	Senior Biostatistician		
Company:	[redacted]		

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, authorizing that the content is acceptable for the reporting of this study.

	Name	Signature	Date
Approved By:	[redacted]		
Position:	Study Biostatistician, Biostatistics Group, Data Science		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		
Approved By:	[redacted]		
Position:	Head of Biostatistics, Group manager Biostatistics Group, Data Science		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		
Approved By:	[redacted]		
Position:	Medical Lead, Fellow CNS Research, Clinical Research,		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		

OUTPUT TEMPLATES SIGNATURE PAGE

Output Templates V1.0 (Dated 30MAR2017) for Protocol D1002001

	Name	Signature	Date
Author:	[redacted]		
Position:	Senior Biostatistician		
Company:	[redacted]		

Upon review of this document, the undersigned approves this version of the Output Templates, authorizing that the content is acceptable for the reporting of this study.

	Name	Signature	Date
Approved By:	[redacted]		
Position:	Study Biostatistician, Biostatistics Group, Data Science		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		
Approved By:	[redacted]		
Position:	Head of Biostatistics, Group manager Biostatistics Group, Data Science		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		
Approved By:	[redacted]		
Position:	Medical Lead, Fellow CNS Research, Clinical Research,		
Company:	Sumitomo Dainippon Pharma Co., Ltd.		

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
0.1	21OCT2015	[redacted]	Not Applicable – First Version
0.2	08FEB2016	[redacted]	Updates made based on sponsor comments. Updates still required for subgroups and sensitivity.
0.3	21MAR2016	[redacted]	Ongoing updates during BDR creation
0.4	NA	[redacted]	Version iterated with shells
0.5	19DEC2016	[redacted]	Reflect comments raised at BDR
1.0	20MAR2017	[redacted]	Addressed final comments and added information requiring study data information.
Final (1.0)	30MAR2017	[redacted]	Minor updates for consistency. Pooled site updated.

Table of Contents

1. INTRODUCTION

This document describes the rules and conventions to be used in the presentation and analysis of efficacy and safety data for Protocol D1002001. It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed.

This statistical analysis plan (SAP) is based on protocol version 1.04, dated 15th March 2016 and amendment 'Japanese Supplement A' version 1.00, dated 21st June 2013.

2. STUDY OBJECTIVES

2.1. PRIMARY OBJECTIVE

The primary objective is to compare the efficacy of SM-13496 (20-60 or 80-120 mg/day) monotherapy with that of placebo in patients with depressive symptoms associated with bipolar I disorder by assessing the change from baseline in the Montgomery-Åsberg Depression Rating Scale (MADRS) total score at Week 6.

2.2. SECONDARY OBJECTIVES

The secondary objectives of the study are as follows:

1) To evaluate the efficacy of SM-13496 (20-60 and 80-120 mg/day) monotherapy for 6 weeks in patients with depressive symptoms associated with bipolar I disorder by assessing the following:

- Change from baseline in the Clinical Global Impression Bipolar Version, Severity of Illness (CGI-BP-S) (depression) score at Week 6
- Change from baseline in the Sheehan Disability Scale (SDS) total score at Week 6
- The response rate and the remission rate

2) To evaluate the time course of efficacy of SM-13496 (20-60 and 80-120 mg/day) monotherapy for 6 weeks in patients with depressive symptoms associated with bipolar I disorder by assessing the change from baseline in the MADRS total score and in the CGI-BP-S (depression) score

3) To evaluate the following:

- The efficacy for anxiety symptoms associated with bipolar I disorder by assessing the change from baseline in the Hamilton Rating Scale for Anxiety (HAM-A) total score at Week 6
- The efficacy for manic symptoms associated with bipolar I disorder by assessing the change from baseline in the Young Mania Rating Scale (YMRS) total score at Week 6

4) To evaluate the overall safety of SM-13496 (20-60 and 80-120 mg/day) by assessing the following:

- Adverse events, adverse drug reactions
- Laboratory measures

- Vital signs and electrocardiography (ECG) measurements

5) To evaluate the influence on the extrapyramidal symptoms by assessing the following:

- Extrapyramidal adverse events, extrapyramidal adverse drug reactions
- Proportion of patients using concomitant antiparkinson drugs
- Change from baseline in the Drug-Induced Extrapyramidal Symptoms Scale (DIEPSS) total score (except for the overall severity score) at the final assessment
- Change from baseline in the individual DIEPSS symptoms scores at the final assessment

6) To evaluate the influence on the following:

- QTc
- Body weight (change from baseline in body weight at the final assessment)
- Prolactin (change from baseline in serum prolactin concentration at the final assessment)
- Glucose metabolism (change from baseline in fasting blood glucose, HbA1c, and glycoalbumin at the final assessment)
- Lipid metabolism (change from baseline in total cholesterol, LDL cholesterol, HDL cholesterol, and triglyceride at the final assessment)

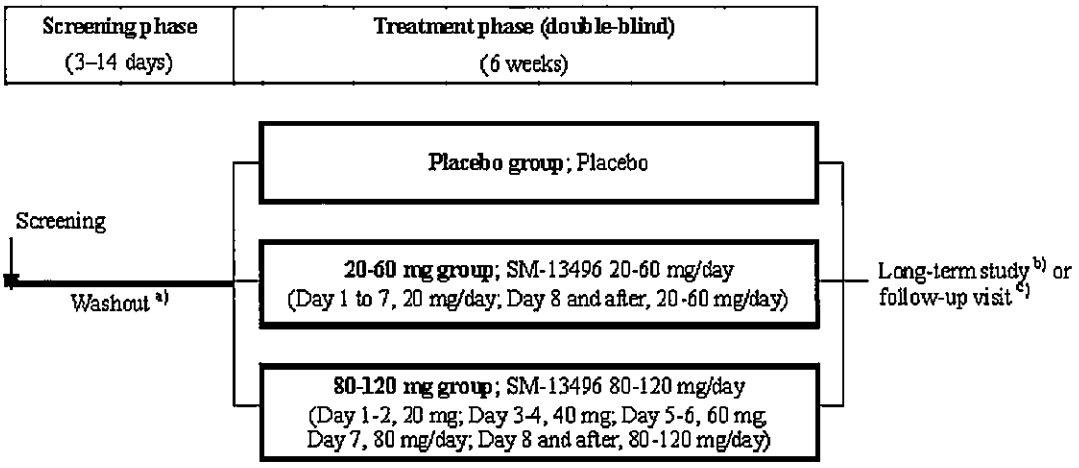
3. STUDY DESIGN

3.1. GENERAL DESCRIPTION

The study will be conducted in a multicenter, randomized, double-blind, parallel-group, placebo-controlled manner.

The study is a placebo-controlled phase 3 study to evaluate the efficacy and safety of SM-13496 monotherapy at doses of 20-60 or 80-120 mg/day, at the same doses for which efficacy was confirmed in the PREVAIL 1 and 2 studies. The primary variable will be the change in the MADRS total score, and the duration of treatment will be 6 weeks. Patients completing the present study can be enrolled in a long-term study (Study D1002002) to evaluate the long-term safety and efficacy of SM-13496.

Table A: Study Flow Chart



a: If the patients receive any psychotropic drugs (e.g., antipsychotics, antidepressants, mood stabilizers) at screening other than antiparkinson agents, anxiolytics, hypnotics (see protocol Section 5.3, Page 39), all the psychotropic drugs will be terminated at least 3 days before the initiation of study treatment after titrated down as needed.
b: The patients who complete the 6-week treatment phase will proceed to the long-term study after being deemed eligible and providing written informed consent to participate in the study.
c: Patients who are not enrolled in the long-term study will visit the study site for a follow-up on 7 days (±2 days) after the final administration of study drug.

3.2. SCHEDULE OF EVENTS

Schedule of events can be found in Section 3.2 of the protocol.

3.3. CHANGES TO ANALYSIS FROM PROTOCOL

There were no changes made to the statistical analysis as planned in the final study protocol.

4. PLANNED ANALYSES

The following analyses will be performed for this study:

- Final Analysis

4.1. DATA MONITORING COMMITTEE (DMC)

There will be no DMC for this study.

4.2. INTERIM ANALYSIS

There is no interim analysis planned for this study.

4.3. FINAL ANALYSIS

All final, planned analyses identified in this SAP will be performed by [redacted] following Sponsor Authorization of this Statistical Analysis Plan, Database Lock, Sponsor Authorization of Populations and Unblinding of Treatment.

5. ANALYSIS POPULATIONS

Agreement and authorization of patients included/ excluded from each analysis population will be conducted prior to the unblinding of the study. The populations will be summarized for each treatment group. This summary will be repeated separated by country.

5.1. INTENTION-TO-TREAT POPULATION (ITT)

The ITT population will consist of all patients who fulfill the following conditions:

- Patients who are randomized and take at least one dose of the study drug in the treatment phase, regardless of any protocol deviations
- Patients who have baseline and at least one post-baseline evaluable MADRS total score (this means that all components of the MADRS questionnaire must be completed).

For analyses and displays based on ITT, patients will be classified according to randomized treatment.

5.2. PER PROTOCOL POPULATION [PP]

The PP population will consist of patients from the ITT population who fulfill the following conditions:

- Patients who fulfill the inclusion criteria (3) to (7), and (10) (see Protocol Section 4.1.1) and do not meet any of the exclusion criteria (1), (5) to (12), (18) to (21), and (28) (see Protocol Section 4.1.2)
- Patients who have been treated for 14 days or longer
- Patients who receive 75% to 125% of study treatment in the treatment phase

- Patients who do not receive any psychotropic medication other than antiparkinson drugs, anxiolytics, and hypnotics during the treatment phase

- Patients who have no important protocol deviations as determined by a blinded data review

The first 3 conditions will be verified using SAS programming. A medical review will be conducted by [redacted] or Sponsor Medical team to identify psychotropic medications. A review sheet will be provided for identifying medications and the completed sheet will be incorporated into SAS derived datasets.

For identifying other important protocol deviations, a blind data review meeting will be convened with representatives from [redacted] and the Sponsor to determine whether the deviations require the patient to be removed from the PP population.

For analyses and displays based on PP, patients will be classified according to randomized treatment.

5.3. SAFETY POPULATION [SAF]

The safety population will consist of all patients who are randomized and receive at least one dose of the study drug in the treatment phase. Safety population summaries and analyses will be presented using the actual treatment received.

6. GENERAL CONSIDERATIONS

6.1. REFERENCE START DATE AND STUDY DAY

Study Day will be calculated from the reference start date, and will be used to show start/ stop day of assessments and events.

Reference start date is defined as the day of the first dose of study medication, (Day 1 is the day of the first dose of study medication).

- If the date of the event is on or after the reference date then:

Study Day = (date of event – reference date) + 1.

- If the date of the event is prior to the reference date then:

Study Day = (date of event – reference date).

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings, and Study Day, and any corresponding durations will be presented based on the imputations specified in Appendix 2; Partial Date Conventions.

6.2. BASELINE

Unless otherwise specified, baseline is defined as the last non-missing measurement taken on or prior to reference start date (including unscheduled assessments). In the case where the last non-missing measurement and the reference start date coincide, that measurement will be considered pre-baseline, but Adverse Events (AEs) and medications commencing on the reference start date will be considered post-baseline.

Post-baseline data will be defined as non-missing data on or after Day 1 and through 7 days after the final administration of study drug in the treatment phase (excluding baseline).

6.3. DERIVED TIMEPOINTS

6.3.1 LOCF Timepoint

The last non-missing post-baseline measurement collected within the double-blind treatment phase will be carried forward and will be defined as the Week 6/last observation carried forward (LOCF) endpoint record. Measurements eligible for LOCF endpoint will be restricted to post-baseline measurements through either the Week 6 visit day (for patients entering the long term study) or 7 days after double-blind study drug discontinuation (for discontinued patients not entering the open-label phase). Unscheduled visits are considered useable for derivation of the LOCF endpoint in all domains.

6.4. RETESTS, UNSCHEDULED VISITS AND EARLY TERMINATION DATA

In general, for by-visit summaries, data recorded at the nominal visit will be presented. Post-baseline unscheduled measurements will not be included in by-visit summaries, but will contribute to the baseline, LOCF endpoint value, or best/ worst case value where required (e.g. shift table).

The data collected at discontinuation visit will be mapped to the next scheduled visit from the actual discontinuation date. Data collected at the discontinuation visit will be eligible for LOCF in cases where the study is not completed, provided the discontinuation is within 7 days of the last dose of study treatment.

Listings will include scheduled, unscheduled, retest and early discontinuation data.

6.5. WINDOWING CONVENTIONS

There will be no visit windowing for the analyses performed for this study. All data will be organized and analyzed according to the scheduled times as outlined in the protocol and by the visit denoted on the eCRF. When early termination occurs post-baseline within 7 days after treatment discontinuation, the assessment information will be assigned to the next planned visit. This will apply to all data points used in efficacy and safety analyses. All tables and figures presenting data by visit will present those timepoints where the applicable assessment was scheduled to be collected. Unscheduled and early termination data will be included for definition of LOCF endpoint or overall assessments. Data listings will present all data regardless of visit.

6.6. STATISTICAL TESTS

The default significance level will be (5%); confidence intervals will be 95% and all tests will be two-sided, unless otherwise specified in the description of the analyses.

6.7. COMMON CALCULATIONS

For quantitative measurements, change from baseline will be calculated as:

- Test Value at Visit X – Baseline Value

6.8. SOFTWARE VERSION

All analyses will be conducted using SAS (UTF-8 encoding) version 9.4 or higher.

7. STATISTICAL CONSIDERATIONS

7.1. ADJUSTMENTS FOR COVARIATES AND FACTORS TO BE INCLUDED IN ANALYSES

The following covariates and factors are used in the analyses. For details of their inclusion in the models, see the specific analysis section.

- Scheduled Visit
- Baseline Questionnaire Scores (dependent upon the questionnaire being analyzed).
- Center (centers to be confirmed). Centers may be subject to pooling, please see section 7.2
- Some separate analyses will be performed relating to subgroups, please see section 7.5.

7.2. MULTICENTER STUDIES

This study will be conducted by multiple investigators at multiple centers internationally.

All centers with eleven or fewer randomized patients will be pooled by size within country or geographic region. No pooled center will contain more than 25 randomized patients. If the pooling cannot be performed within country, other centers in the same geographic region will be considered. The list of pooled centers for the present study based on the pooling strategy previously described is provided in Table B.

Table B: Pooled Sites

Pooled Center ID	Center ID (Number of Subjects)	Country	Number of Subjects
901	[redacted]	JPN	21
902		JPN	21
903		JPN	13
904		JPN	23
905		JPN	22
906		JPN	22
907		JPN	24
908		JPN	17
909		JPN	16
910		LTU	22
		RUS	
911		MYS	19
912		PHL	24
		TWN	
913		RUS	22
914		RUS	24
915		RUS	17
916		RUS	24
917		RUS	18
918		RUS	13
919		RUS	15
920		SVK	16
921		UKR	19
922		UKR	17
923		UKR	17
924		UKR	15
925		UKR	15
926		UKR	14
927		UKR	13
928		UKR	21

A term for treatment by pooled center interaction will not be included in the primary analysis model, however, a sensitivity analysis, based on the primary model but including a term for treatment by pooled center interaction, will be performed.

7.3. MISSING DATA

Missing efficacy data will be handled as described in section 15.1.2 of this analysis plan.

Missing safety data will be subject to LOCF analysis as specified in relevant safety domain sections.

7.4. MULTIPLE COMPARISONS/ MULTIPLICITY

An analysis must be conducted taking into account the multiple dose regimens for primary efficacy analysis, as there are 2 active treatment groups being compared to a single placebo treatment group.

Type I error for rejection of the null hypothesis (see section 15.1.3) will be controlled at α=0.05 (two-sided) using the Hochberg procedure (Hochberg, 1988).

The hypotheses being tested are:

- $H_{0(1)}$: $\mu_{20\text{-}60}=\mu_{placebo}$ versus the alternate H_{11}: $\mu_{20\text{-}60}\neq\mu_{placebo}$.

- $H_{0(2)}$: $\mu_{80\text{-}120}=\mu_{placebo}$ versus the alternate H_{12}: $\mu_{80\text{-}120}\neq\mu_{placebo}$.

Let p_{20} and p_{80} denote the p-values for the two primary comparisons on the primary efficacy parameter; then define p_{min} = min (p_{20}, p_{80}), and p_{max}=max (p_{20}, p_{80}), representing the two dosing regimens being evaluated. Within the Hochberg multiple comparison procedure one of the following three cases will apply:

- If $p_{max} \leq 0.050$ then both primary tests are classified as statistically significant;
- If $p_{max} > 0.050$ and $p_{min} \leq 0.025$ then the test corresponding to p_{min} is classified as statistically significant, and the test corresponding to p_{max} is classified as not statistically significant; or
- If $p_{max} > 0.050$ and $p_{min} > 0.025$ then both primary tests are classified as not statistically significant.

In summary, if both tests have p≤0.05, then both null hypotheses are rejected; if the less significant test (p_{max}) has p≥0.05, then that null hypothesis ($H_{0(2)}$) is not rejected, and the other null hypothesis ($H_{0(1)}$) is tested at the 0.025 level.

This analysis will be conducted using the MULTTEST procedure in SAS.

7.5. EXAMINATION OF SUBGROUPS

Subgroup analyses will be conducted on change from baseline results for MADRS total score and CGI-BP-S depression score from Week 1 to Week 6, including Week 6 (LOCF) endpoint. It should be noted that the study was not designed to detect treatment differences with high statistical power within subgroups.

The efficacy analyses to be repeated on subgroups are the MMRM and ANCOVA methods for both MADRS and CGI-BP-S. This analysis will only be performed on the ITT population. Additionally, SDS, HAM-A and YMRS will be analyzed separately by country.

A subgroup analysis will also be performed upon a selection of safety parameters. These parameters are DIEPSS

(ANCOVA only) as well as weight, BMI, QTc, and selected laboratory parameters (ANCOVA only). A small number of specified Adverse Event tables will also be split by subgroup. Safety parameter subgroup analyses will be performed on the Safety population.

The full analysis methods for subgroups are detailed in section 15.2.4.

The following subgroups will be assessed and described within the exploratory analysis sections:

- Sex:
  - Male
  - Female
- Age (years):
  - <55
  - ≥55

  And
  - <65
  - ≥65
- Bipolar I Diagnosis Subtype (Efficacy Analyses only)
  - Rapid Cycling
  - Non-rapid Cycling
- Country
  - Japan
  - Ukraine
  - Philippines
  - Taiwan
  - Russia
  - Malaysia
  - Lithuania
  - Slovakia

8. OUTPUT PRESENTATIONS

Appendix 1 shows conventions for presentation of data in outputs.

The templates provided with this SAP describe the presentations for this study and therefore the format and content of the summary tables, figures and listings to be provided by [redacted]

Summaries for categorical data will include frequency counts and percentages. For continuous data points, summaries will include a count, mean, standard deviation, median, minimum and maximum.

9. DISPOSITION AND WITHDRAWALS

All patients who provide informed consent and are screened will be accounted for in this study.

Subject disposition and withdrawals will be presented for all enrolled patients. Protocol violations (as defined in section 5.2), including inclusion and exclusion criteria, will be presented for the ITT population. Number of patients screened, randomized, completed, discontinued and continuing into the extension study will be summarized, along with reasons for discontinuation as required.

This summary will be repeated separated by country.

10. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic data and other baseline characteristics will be presented for the ITT population, safety population and PP population. This summary will be repeated separated by country for the ITT population.

No statistical testing will be carried out for demographic or other baseline characteristics.

The following demographic and other baseline characteristics will be summarized and listed for this study. Data points denoted by * are reported in a listing only:

1) Demographics

- Age
- Sex
- Race
- Ethnicity
- Height
- Country
- Weight
- BMI

2) Disease data

- Date of initial onset of bipolar I disorder
- DSM-IV-TR diagnostic code (regarding severity/psychotic/remission specifiers) of the current major depressive episode or other current episode
- Date of onset of the current major depressive episode
- Presence or absence of characteristics of rapid cycling disease course
- Number of mood episodes for the consequent 12 months, by episode type

- Date of onset of the most recent manic episode, hypomanic episode, and mixed episode*
- Date of most recent hospital discharge if hospitalized because of manic, mixed manic or depressive episodes*
- Number of hospitalizations due to bipolar I disorder
- Psychiatric disease other than bipolar I disorder; diagnosis, DSM-IV-TR diagnostic code, date of onset*
- Baseline scores for MADRS, CGI-BP-S and SDS.

3) Bipolarity Index (BPI)*

The BPI is a diagnostic instrument that considers 5 "dimensions" of bipolarity:

1. Hypomania or mania
2. Age of onset of first mood symptoms
3. Illness course and other features generally only apparent over time
4. Response to medications (antidepressants and mood stabilizers)
5. Family history of mood and substance use problems

11. SURGICAL AND MEDICAL HISTORY

Medical History information will be presented for the ITT population and the SAF population.

- Medical History will be coded using **MedDRA Version 19.1**
  - Medical History conditions are defined as those conditions beginning prior to screening which stop prior to/are ongoing at Screening.
  - Summarized by System Organ Class (SOC) and Preferred Term (PT).

12. CONCOMITANT MEDICATIONS

Concomitant Medications will be presented for the Safety population and coded using **WHO Drug Dictionary Enhanced /WHO Herbal Dictionary v01DEC2013.**

ATC coding will be performed for the study, including screen failed patients. Medications will be summarized using ATC level 3 coding and Preferred Term.

See Appendix 2 for handling of partial dates for medications, in the case where it is not possible to define a medication as prior, concomitant, or post treatment, the medication will be classified as both prior and concomitant.

- 'Prior' medications are medications which started prior to the first dose of study medication. If they are ongoing at start of dosing then the medication will be classified as both prior and concomitant.
- Concomitant' medications are medications which:

- started prior to, on or after the first dose of study medication and started no later than one day before the end of the double blind phase
- AND ended on or after the date of first dose of study medication or were ongoing at the end of the study.

- 'Post' medications are medications which started on the same day, or after the end of study treatment for patients who are not continuing into the extension study. For patients continuing into the extension study, the medication should be captured in the extension study database.

A summary of patients with concomitant use of antiparkinson medication will be summarized by treatment group. An additional summary will be produced categorizing medication categories related to Anxiolytics and Hypnotics.

13. STUDY MEDICATION EXPOSURE

Exposure to study medication in days will be presented for the ITT population and SAF population.

The date of first and last study medication administration will be taken from the eCRF drug administration and accountability form. Return of the blister card at any given visit is not a prerequisite for inclusion in exposure summary statistics if corresponding dose start and end dates are reported.

Interruptions, compliance, and dose changes are not taken into account for duration of exposure.

The average daily assigned dose will be summarized for the Safety and ITT populations by treatment group stratified for all patients and completers. The modal daily assigned dose will likewise be summarized based on overall study drug exposure; if there is more than one modal daily assigned dose for a subject, the last modal assigned dose administered will be used in the overall summary. Return of the blister card at any given visit is not a prerequisite for inclusion in average daily assigned dose summary statistics or modal daily assigned dose summary statistics if corresponding dose start and end dates are reported. For the purpose of computing average daily dose and modal daily dose (overall), dose level will be imputed during Week 1 for patients randomized to the 80-120 mg treatment arm because dose titration was required for this group during this week (see Protocol Section 3.3), and will be based on the assumption that dosing instructions associated with titration have been followed correctly (i.e. 20 mg tablets taken on Days 1-2, 40 mg tablets taken on Days 3-4, 60 mg tablets taken on Days 5-6, and 80 mg tablets taken on Day 7).

The number of subjects receiving each dose within a given study week will be presented. A summary of dose change from previous scheduled visit will be presented by assigned dose level for patients in the Safety and ITT populations. Return of the blister card at any given visit is not a prerequisite for inclusion in dose change summary statistics if corresponding dose start and end dates are reported.

13.1. DERIVATIONS

Duration of exposure (days) = date of last study medication administration – date of first study medication administration + 1.

14. STUDY MEDICATION COMPLIANCE

Compliance will be presented for the Safety and ITT populations by treatment group, and overall, both continuously (mean percentage) and categorically (compliant vs. non-compliant, non-compliant <75% and non-

complaint >125%, and patients with any missing compliance).

14.1. DERIVATIONS

At each visit, prior to dispensing study medication, previously dispensed study medication will be retrieved and assessed by tablet count. Compliance in the double-blind phase will be calculated overall as follows:

- Percent compliance = (number of tablets taken / number of tablets should have taken) x 100.
- Number of tablets taken = number of tablets dispensed – number of tablets returned. Lost tablets are considered to have been returned for the purpose of calculating compliance.
- Number of tablets should have taken = (number of tablets supposed to take in a day) x (number of exposure days).

If the number of tablets returned is missing, compliance will be missing. Non-compliance is defined as less than 75% or more than 125% non-missing overall compliance with the study medication. Patients with missing compliance are not classified as non-compliant, and are not considered to be protocol deviators.

15. EFFICACY OUTCOMES

15.1. PRIMARY EFFICACY

15.1.1. PRIMARY EFFICACY VARIABLE & DERIVATION

The primary efficacy variable is the change from baseline in the MADRS total score at Week 6 for testing superiority of SM-13496 (20-60 or 80-120 mg/day) to placebo. The rater-administered MADRS data will be used for all statistical analyses. The full data for the questionnaire is collected on the eCRF.

The total score is calculated as the sum of all the component scores of the MADRS scale. The MADRS total score ranges from 0 to 60. Higher scores are associated with greater severity of Bipolar I disorder.

The change from baseline at week 6 is calculated as MADRS total score at week 6 minus the MADRS total score at baseline. This value will be a negative value if the patient condition has improved.

15.1.2. MISSING DATA METHODS FOR PRIMARY EFFICACY VARIABLE(S)

In the event that any component element of a total score is missing, then the total score will be set to missing.

Missing data at a visit will be accounted for using a combination of approaches. The MMRM method being used for the primary analysis includes an assumption that data is missing at random.

The primary analysis will also have a sensitivity analysis conducted using a LOCF variable which will impute a missing week 6 endpoint value.

Additionally, 2 sensitivity analyses will be conducted investigating alternate assumptions about missing data. For full information about these please see section 15.1.4.

15.1.3. PRIMARY ANALYSIS OF PRIMARY EFFICACY VARIABLE(S)

15.1.3.1. Montgomery-Asberg Depression Rating Scale (MADRS)

The primary efficacy variable is the change from baseline in the MADRS total score at Week 6 for testing superiority of SM-13496 (20-60 or 80-120 mg/day) to placebo, versus the null hypothesis of no difference versus placebo. The hypothesis for each SM-13496 treatment group will be considered independently.

Let $\mu_{20\text{-}60}$, $\mu_{80\text{-}120}$, and $\mu_{placebo}$ represent the mean changes from baseline in the MADRS total score at Week 6 in the 20-60mg, 80-120mg, and placebo groups, respectively. The following 2 hypotheses will be tested to compare the mean changes at Week 6 of each SM-13496 group with that of the placebo group:

- H_{01}: $\mu_{20\text{-}60}=\mu_{placebo}$ versus the alternate H_{11}: $\mu_{20\text{-}60}\neq\mu_{placebo}$.
- H_{02}: $\mu_{80\text{-}120}=\mu_{placebo}$ versus the alternate H_{12}: $\mu_{80\text{-}120}\neq\mu_{placebo}$.

The rater-administered MADRS data will be used for all statistical analyses.

The MMRM method will be used in the ITT population. The MMRM model will include treatment as a categorical factor, visit (Week 1, 2, 3, 4, 5, and 6; as a categorical variable), pooled center, the MADRS total score at baseline as a covariate, and the treatment-by-visit interaction. An unstructured covariance matrix will be used for the within-patient correlation and the Kenward-Rogers approximation will be used to calculate the denominator degree of freedom. In the case of a convergence problem with the unstructured covariance matrix, the following structures with robust sandwich estimator for the standard error of the fixed effect estimates will be assessed in a sequential fashion: heterogeneous Toeplitz, heterogeneous first-order autoregressive, and Toeplitz. Among these 3 covariance structures, the first structure yielding convergence will be used for the MMRM analysis. P-values for comparisons of primary efficacy variable at Week 6 between SM-13496 groups and the placebo group will be adjusted for multiplicity using a Hochberg procedure (section 7.4).

Effect size at Week 6 will also be presented and will be derived as the absolute value of the LS mean difference between the treatment groups divided by the model estimate of the standard deviation. To calculate the standard deviation of change from baseline at Week 6 based on MMRM model, the square root of the variance estimate for that visit, obtained from the SAS-estimated R matrix will be taken (O'Kelly & Ratitch, 2014). The variance estimates for the visits are obtained differently dependent upon the covariance structure being applied for the model. If the MMRM converges using the planned unstructured covariance structure then the diagonal of the R matrix will be used for variance estimate at each visit. In the case of non-convergence and the use of a heterogeneous Toeplitz based MMRM or first order auto-regressive MMRM then the diagonal cell corresponding to Week 6 from the R matrix can still be used to estimate the variance. However, in the case of non-convergence and the use of a regular Toeplitz based MMRM, the diagonal from the first column of the R matrix can be used as the variance estimate.

The mean change from baseline in the MADRS total score will be plotted by visit of discontinuation to assess whether there are dropout patterns that appear to violate the missing at random (MAR) assumption, which the MMRM is based upon.

15.1.4. SENSITIVITY ANALYSIS OF PRIMARY EFFICACY VARIABLE

15.1.4.1. ANCOVA

A sensitivity analysis for the primary efficacy variable will be performed using an analysis of covariance (ANCOVA) model on the ITT population. The response variable for the model will be the change from baseline in the MADRS total score at Week 6 (LOCF). The ANCOVA model will include treatment as a categorical factor, pooled study site and the MADRS total score at baseline as a covariate..

Treatment differences for baseline MADRS data will be based on an ANOVA model adjusting for pooled center. Within-treatment post-baseline differences will be assessed with a one-sample t-test using the LS mean and estimated standard error of the mean change from baseline difference. Least square means for each treatment group, the mean difference between treatment groups, the standard error of the mean difference between treatment groups, the associated two-sided 95% CI for the difference between the treatment groups, and p-values for between-treatment tests of differences will be presented. The same pairwise comparisons will be presented as for the MMRM model. Within-group effect size will be calculated as the absolute value of the mean change from baseline divided by the standard deviation of the change from baseline and will be displayed for all post-baseline visits. Between-group effect size will be calculated as the absolute value of the LS mean difference from placebo divided by the model estimate of the pooled standard error (the standard error of the LS mean difference divided by the square root of the sum of inverse treatment group sizes). This analysis will be conducted for the ITT and PP populations. All treatment groups will be included in the ANCOVA models.

15.1.4.2. Random Effects Pattern Mixture Model

A random effects pattern mixture model (Hedeker & Gibbons, 1997) will be applied as one of two sensitivity analyses to explore the robustness of the MMRM results for the primary and key secondary efficacy variables for the ITT population.

Using a random effects PMM, one can examine (a) the degree to which the groups defined by the dropout patterns differ in terms of the outcome variable (i.e., main effect of the dropout pattern variable) and (b) the degree to which the dropout pattern moderates the influence of other model terms (i.e., interactions with the dropout pattern). In deciding on an appropriate grouping of the dropout patterns, an important consideration is the sparseness of the data for each pattern. If a pattern has very few observations, it may not make sense to treat it as a separate group in the analysis. About 82% of randomized patients in the present study were completers as of last patient last visit on 20FEB2017. Hedeker and Gibbons indicated that (Hedeker & Gibbons, 1997) if a large percentage of patients complete the study, it is reasonable to classify patients into completers and dropouts.

Prior to selecting the completer/dropout pattern used above, exploratory analyses were conducted using data from previous SM-13496 schizophrenia studies (D1050229 and D1050233). The completion rates for these schizophrenia studies were 66% for Study D1050229, and 72% for Study D1050233, respectively. The drop out rate for D1002001 is 17.4%. These exploratory analyses involved the following sets of dropout patterns: 1) dropouts and completers, 2) early dropouts, late dropouts, and completers; and, 3) dropouts due to insufficient clinical response and AEs, dropouts due to other reasons, and completers. The overall conclusions from these exploratory analyses were consistent with those from the primary efficacy analyses from the respective studies.

The random effects pattern mixture model will include fixed terms for treatment, time (as a continuous variable), dropout pattern, and two-way interaction terms (i.e., treatment-by-time, dropout pattern-by-treatment, dropout pattern-by-time), and a three-way interaction term (dropout pattern-by-treatment-by-time). The model will also include subject-specific random effects for intercept and time. The response

variables will be the observed values of the efficacy variables over time, including the baseline value. Patients will be classified into two separate patterns: dropouts and completers. To illustrate, only two treatment groups are considered. Let yij represent the response value of the ith subject at time j (i.e., actual day of assessment), the following random effects pattern mixture model with an unstructured covariance matrix is fitted:

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 Time_{ij} + \beta_2 Treat_i + \beta_3 (Treat_i \times Time_{ij}) \\ &\quad + \beta_4 Drop_i + \beta_5 (Drop_i \times Time_{ij}) + \beta_6 (Drop_i \times Treat_i) + \beta_7 (Drop_i \times Treat_i \times Time_{ij}) \\ &\quad + v_{oi} + v_{1i} Time_{ij} + \varepsilon_{ij} \end{aligned} \quad (1)$$

Where:

Drop = 0 for completers, and 1 for dropouts
Treat = 0 for patients in the placebo group, and 1 for patients in the SM-13496 treatment group

The random intercept v0i and random time effect v1i are assumed to follow a bivariate normal distribution, and the model residuals εij are assumed to be independently distributed and follow a univariate normal distribution.

Based on the above specified random effects pattern mixture model, the parameter estimates for fixed effects (intercept, time, treatment, treatment-by-time, dropout pattern, dropout pattern-by-time, dropout pattern-by-treatment, and dropout pattern-by-treatment-by-time) as well as the covariance matrix for the parameter estimates will be obtained by using the SAS PROC MIXED procedure.

Substituting Drop = 0 and 1 in the above random effects pattern mixture model (1), the following regression models for completers and dropouts can be obtained. The model for completers is:

$$y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Treat_i + \beta_3 (Treat_i \times Time_{ij}) + v_{oi} + v_{1i} Time_{ij} + \varepsilon_{ij} \quad (2)$$

and the model for dropouts is:

$$\begin{aligned} y_{ij} &= (\beta_0 + \beta_4) + (\beta_1 + \beta_5) Time_{ij} + (\beta_2 + \beta_6) Treat_i \\ &\quad + (\beta_3 + \beta_7)(Treat_i \times Time_{ij}) + v_{oi} + v_{1i} Time_{ij} + \varepsilon_{ij} \end{aligned} \quad (3)$$

Using the estimates from model 1, the parameter estimates for models 2 and 3, i.e., $\beta_k^c = \beta_k$, $\beta_k^d = \beta_k + \beta_{4+k}$ (k = 0, 1,2,3; c=completers and d=dropouts), for the fixed effects (intercept, time, treatment, and treatment-by-time), as well as the corresponding estimates of the standard errors for these parameter estimates, will be derived for each dropout pattern. The SAS PROC IML procedure can be used for this purpose. In PROC IML, a design matrix is used to create the estimates for completers and dropouts, while an additional design matrix is used to create the overall parameter estimates (intercept, time, treatment, and treatment-by-time) that are a weighted average of the parameter estimates from each dropout pattern:

$$\hat{\bar{\beta}}_k = \hat{\pi}_k^c \hat{\beta}_k^c + \hat{\pi}_k^d \hat{\beta}_k^d \quad (4)$$

where k = 0, 1,2,3 denotes the four fixed effects. Weights $\hat{\pi}_k^c$ and $\hat{\pi}_k^d$ are determined based on the proportions of completers and dropouts in the placebo group for k = 0, 1 and in the SM-13496 treatment group for k = 2, 3. To obtain corresponding estimates of the standard errors for these overall estimates in PROC IML, the

following delta method (Hogan & Laird, 1997) will be used:

$$\hat{V}(\hat{\beta}_k) = (\hat{\pi}_k^c)^2 \hat{V}(\hat{\beta}_k^c) + (\hat{\pi}_k^d)^2 \hat{V}(\hat{\beta}_k^d) + \frac{\hat{\pi}_k^c \hat{\pi}_k^d}{N_k} (\hat{\beta}_k^c - \hat{\beta}_k^d)^2 \quad (5)$$

where Nk is the number of patients in the placebo group for k = 0, 1 and in the SM-13496 treatment group for k = 2, 3, and $\hat{V}(\hat{\beta}_k)$ denotes the estimate of the variance of $\hat{\beta}_k$ (i.e., the square of its estimated standard error).

Finally, the overall parameter estimates from the random effects pattern mixture model will be compared with the estimates from a random effects model that does not include dropout pattern, but will include fixed terms for treatment, time (as a continuous variable), and a treatment-by-time interaction. This model will also include subject-specific random effects for intercept and time. The random effect model is:

$$y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Treat_i + \beta_3 (Treat_i \times Time_{ij}) + v_{oi} + v_{1i} Time_{ij} + \varepsilon_{ij} \quad (6)$$

A comparison of the parameters between the above two models helps to assess any impact of missingness (Siddiqui, Hung, & O'Neil, 2009). If the results from these two models are consistent, it implies the MAR assumption is reasonable, because the random effects model without dropout pattern assumes MAR, which is the same assumption that MMRM requires.

15.1.4.3. Pattern Mixture Model with Placebo-based Multiple Imputation

A pattern-mixture model using a placebo-based multiple imputation method will be performed as a second sensitivity analysis to explore the robustness of the MMRM results for the primary and key secondary efficacy variables for the ITT population. This method was based on methodology proposed by Ratitch and O'Kelly (Ratitch & O'Kelly, 2011). The assumption that efficacy profiles of dropouts after discontinuation are similar to those of placebo patients is considered conservative because this methodology tends to minimize the difference between SM-13496 and placebo groups. If the results of this analysis are in line with the primary efficacy results, then it can be confidently concluded that the primary analysis results are robust.

Using this method, missing data after dropout will be imputed based on placebo group data using multiple imputation methodology for all patients (including active treatment patients).

The steps to implement this sensitivity analysis are as follows:

1. 1000 datasets will be generated where missing data at intermediate visit(s) will be imputed for each treatment group using non-missing data from all patients within the treatment group by a Monte Carlo Markov Chain (MCMC) imputation model using the MCMC statement in the SAS PROC MI procedure. As a result, each dataset will only have missing ending data, or a monotone missing data pattern.

2. For each dataset from Step 1, missing ending data will be imputed based on information from the placebo group. As a result, 1000 imputed complete datasets will be generated.

- Missing data at the first post-baseline visit will be imputed by a regression imputation model using the SAS PROC MI procedure with the REGRESSION option in the MONOTONE statement.
- The SAS PROC MI procedure will use data from all placebo patients as well as SM-13496 patients with missing data at the visit (i.e., only those that need imputation at the visit). Because patients

from SM-13496 groups without missing data at the visit are excluded from this step, they will not contribute to the estimation of the imputation model for the visit.

- Repeat for all other visits sequentially. Patients whose missing data were imputed at previous visits will contribute to the imputation for the next visit.
- The regression imputation model includes an intercept and the slopes of the measurements from all previous visits (Allison, 2010).

3. Analyze each imputed complete dataset

- The MMRM model used for the primary and key secondary efficacy variables will be utilized for each imputed dataset.

4. Combine estimates from the results of each MMRM model

- SAS MIANALYZE procedure will be used to combine the results.

The results of the placebo-based multiple imputation analysis will be compared with the MMRM results for the primary and key secondary efficacy variables to support the robustness of the MMRM results.

15.1.4.4. Primary Analysis on Per Protocol Population

The MMRM from the primary analysis will also be performed on the PP population to obtain additional information on the robustness of results across populations.

15.2. SECONDARY EFFICACY

The secondary efficacy analyses will be performed for the ITT population for all secondary variables, and the PP population where specified.

15.2.1. SECONDARY EFFICACY VARIABLES & DERIVATIONS

15.2.1.1. Clinical Global Impressions Scale – Bipolar Version - Severity of Illness (CGI-BP-S)

The change from baseline in CGI-BP-S depression score at Week 6 is a secondary efficacy variable. The CGI-BP-S depression score, measured at Screening, Baseline, Weeks 1 to 5, and Week 6/Early Term, is a single value, clinician-rated assessment of the subject's current severity of depression and ranges from 1= 'Normal, not ill' to 7= 'Very severely ill'.

The change from baseline will be calculated as the score at Week 6 – score at baseline.

A LOCF endpoint will also be derived for CGI-BP-S. For calculation details, please see section 6.3.

Additional efficacy ANCOVA analyses will be conducted based on change from baseline in CGI-BP-S depression score at Weeks 1 – 6, and Week 6 LOCF endpoint.

15.2.1.2. Sheehan Disability Scale (SDS)

The change from baseline in SDS total score at Week 6 (LOCF) endpoint is a secondary efficacy variable. The SDS is measured at Baseline and Week 6/Early Term. Three items are self-rated using an 11-point visual analog

scale ranging from 0 to 10 to assess disability across three domains: work/school, social life, and family life; the three items will be summarized individually in addition to the SDS total score. The SDS total score is calculated as the sum of the 3 items and ranges from 0 (Not at all) to 30 (Extremely). If one or more items are missing at a visit, as can occur when a subject opts out of the work/school item because it does not apply, the total score will be set to missing.

15.2.1.3. Young Mania Rating Scale (YMRS)

The change from baseline in YMRS to week 6 is a secondary efficacy variable. The YMRS is collected at Screening, Baseline, Weeks 1 to 5, and Week 6/Early Term. The YMRS is an 11-item clinician-rated instrument used to assess the severity of mania in patients with a diagnosis of bipolar disorder. The 11 items are: elevated mood, increased motor activity energy, sexual interest, sleep, irritability, speech (rate and amount), language-thought disorder, content, disruptive-aggressive behavior, appearance and insight. Seven items are rated on a 5-point scale, ranging from 0 to 4, and four items are rated on a 9-point scale, ranging from 0 to 8. Ratings are based on patient self-reporting, combined with clinician observation. The YMRS total score is the sum of the 11 individual items and ranges from 0 to 60. A higher score is associated with a greater severity of mania. If one or more items are missing at a visit, the total score will be set to missing.

The change from baseline will be calculated as the score at Week 6 – score at baseline.

A LOCF endpoint will also be derived for YMRS. For calculation details, please see section 6.3.

15.2.1.4. Hamiliton Rating Scale for Anxiety (HAM-A)

The change from baseline in HAM-A total score at Week 6 (LOCF) endpoint is a secondary efficacy variable. The HAM-A is used to quantify the severity of anxiety symptomatology. The HAM-A is a 14-item, rater administered questionnaire collected at Baseline and Week 6/Early Term. The 14 items are: anxious mood, tension, fears, insomnia, intellectual, depressed mood, somatic (muscular), somatic (sensory), cardiovascular symptoms, respiratory symptoms, gastrointestinal symptoms, genitourinary symptoms, autonomic symptoms, and behavior at interview. Each of the 14 items is rated on a 5-point scale, ranging from 0 (not present) to 4 (severe/disabling). The HAM-A total score is the sum of the 14 individual items and ranges from 0 to 56. A higher score is associated with a greater degree of anxiety. If one or more items are missing at a visit, the total score will be set to missing.

15.2.1.5. Montgomerg-Asberg Depression Rating Scale (MADRS)

The following measures based on the MADRS are designated as secondary efficacy variables:

- Change from baseline in the MADRS total score at each scheduled assessment (Weeks 1-6). Analysis will be analyzed using the MMRM method as specified for the primary endpoint analysis.
- Proportion of patients who achieve a treatment response, defined as ≥50% reduction from baseline MADRS total score at Week 6 (LOCF) endpoint. The MADRS total score percentage change will be defined as (value at post-baseline visit – baseline value) x 100 / (baseline value).
- Proportion of patients who achieve remission, defined as a MADRS total score of ≤12 at Week 6 (LOCF) endpoint.

15.2.2. MISSING DATA METHODS FOR SECONDARY EFFICACY VARIABLE(S)

Missing question data for questionnaires is accounted for in the variable descriptions. The methods used for analysis of these endpoints will address visit data that is missing.

15.2.3. Analysis of Secondary Efficacy Variables

15.2.3.1. Analysis of CGI-BP-S and YMRS

The change from baseline in CGI-BP-S depression score and YMRS total score will be evaluated using an MMRM model, utilizing the same procedures and reporting as indicated for the MADRS primary efficacy MMRM analysis.

Additionally, the LOCF ANCOVA sensitivity analysis performed for the primary endpoint will also be performed for both scores.

These analyses will be conducted on both the ITT population. The analysis for CGI-BP-S will be repeated on the PP population.

15.2.3.2. Analysis of SDS and HAM-A

The change from baseline in SDS total score and HAM-A total score will be evaluated using the same LOCF ANCOVA method that is used for the sensitivity analysis of the primary efficacy variable.

Additionally, an analysis of days lost and days underproductive will be conducted from the SDS scale. This will take the form of a categorical summary supported by a logistic regression comparison versus placebo.

15.2.3.3. Analysis of Proportion of patients achieving a treatment response

Patients having a 50% or greater improvement from baseline in MADRS total score will be defined as "responders". Patients having a negative percentage change indicate improvement in MADRS total score. The proportion will be evaluated using logistic regression. The responder indicator will be set to 1 if the percentage improvement is greater than or equal to 50%, set to 0 if less than 50%, or set to missing if the percentage is missing. The responder variable will be modeled using treatment and pooled study center as a categorical factor, and baseline MADRS total score as a covariate. If the model fails to converge, the model will be attempted with country (in place of pooled center) and baseline MADRS total score. If the model still does not converge then geographic region will be used (in place of country). Contrasts will be evaluated for each SM-13496 treatment group compared to placebo. Odds Ratios, 95% CIs and Wald chi-square p-values will be presented for each contrast. Only n and percent will be presented in cases where the number of patients meeting the above composite variable criteria is too small for an analysis to be performed.

A Cochran-Armitage trend test will also be conducted on this outcome across placebo and the two SM-13496 treatment groups. In addition, the number needed to treat (NNT) to attain an additional responder will be derived for each SM-13496 dose group as follows:

NNT = 1 / Risk Reduction

where Risk Reduction (RR) = (SM-13496 response rate – placebo response rate). NNT results will be provided in whole numbers with any fractional values rounded up to the nearest whole number. The 95% CIs will also be presented and are computed by taking the reciprocal of the 95% lower and upper bounds of the RR. The lower confidence limit will be rounded down to the largest integer value that is less than the computed estimate, and the upper confidence limit will be rounded up to the smallest integer value that is greater than the computed estimate.

Between-group effect size represented by Cohen's h will also be presented and will be derived as follows:

h = 2 x arcsine (sqrt (SM-13496 response rate)) - 2 x arcsine (sqrt (placebo response rate))

This analysis will be performed for the ITT population, and will be repeated by country. Similar responder classifications are defined for subjects with 20%, 30%, and 40% or greater improvement from baseline in MADRS total score. Subjects classified as 20%, 30% and 40% responders will be analyzed as in the previous paragraph, with the exception of the Cochran-Armitage test which is exclusive to the 50% analysis.

This analysis will be repeated for a subgroup of patients who also achieved response criteria on the CGI-BP-S scale. This is classed as a CGI-BP-S score <=3.

The time to response of MADRS will be analyzed using a time-to-event analysis. Patients failing to achieve a response during the study period will be censored at the day of completion/discontinuation from the study. Kaplan-Meier results will be displayed for the 25%, 50% and 75% quartiles. Additionally, the Kaplan-Meier estimates of the response rate for each treatment group will be displayed. A log-rank test will also be used to compare SM-13496 with placebo, and also compare the two SM-13496 doses. A Kaplan-Meier figure will also be produced to support the analysis.

The time-to-event analysis will also use a Cox proportional hazards model with fixed effects for treatment, geographic region and baseline MADRS total score as a continuous covariate. This model will be used to generate hazard ratios for the same comparisons as specified for the log-rank test.

15.2.3.4. Analysis of Proportion of patients achieving a treatment remission

Patients with a MADRS total score <=12 are classed as being in remission. The proportion of patients achieving a remission will be analyzed using the same methods as for the number of patients achieving a response, with the exception of the Cochran Armitage Test..

In summarizing the number of subjects achieving remission, the number of subjects achieving a MADRS total score <=8 and <=10 will also be presented.

15.2.4. SUBGROUP ANALYSIS METHODS

Subgroup analysis will be performed for the change from baseline in the MADRS total score and the CGI-BP-S (depression) score at Week 6. For details of the subgroups, please refer to section 7.5.

For all subgroups other than country, inferential analysis of treatment-by-subgroup interaction at Week 6 will be performed using the MMRM method on the ITT population. The model will include treatment, visit, baseline data, pooled study site, subgroup, treatment-by-subgroup, treatment-by-visit, visit-by-subgroup, and treatment-by-subgroup-by-visit interactions.

Analysis for country will be conducted using the MMRM method described above for the primary efficacy variable on the ITT population. The analysis will be repeated for each country using a separate model to create estimates in each case.

Safety parameters will be evaluated using a non-parametric rank ANCOVA test at week 6. This will use the same model logic as specified for the analyses using MMRM. Country will be evaluated separately, while other subgroups will be modeled with interaction terms included.

16. SAFETY OUTCOMES

All outputs for safety outcomes will be based on the Safety Population.

16.1.2. TEAEs Leading to Discontinuation of Study Medication

TEAEs leading to permanent discontinuation of study medication will be identified by using the 'action taken' variable collected on the eCRF.

For TEAEs leading to discontinuation of study medication, summaries of incidence rates (frequencies and percentages) by SOC and PT will be prepared.

16.1.3. Serious Adverse Events

Serious adverse events (SAEs) are those events recorded as "Serious" on the Adverse Events page of the eCRF. A summary of serious TEAEs by SOC and PT will be prepared. A listing will provide details of SAEs and indicate any drug relationship.

16.1.4. Adverse Events Leading to Death

TEAEs leading to Death are those events which are recorded as "Fatal" on the Adverse Events page of the eCRF. These events will be listed.

16.1.5. Adverse Events Of Special Interest

16.1.5.1. TEAEs relating to Extrapyramidal Symptoms

TEAEs relating to Extrapyramidal symptoms will be summarized by Preferred Term. These will be identified by using a list of specified terms from medical personnel. This summary will be repeated separated by country.

16.1.5.2. Metabolic TEAEs

Metabolic TEAEs will be summarized. These will be identified with by using a list of specified terms from medical personnel.

16.1.6. Subgroup Analysis of Adverse Events

A single summary of adverse events will be produced separated by the subgroups specified in section 7.5. Other adverse event summaries split by country are noted in the appropriate sections.

16.1.7. Additional AE Summaries

- Summary of AEs by combined terms
- Summary of AEs occurring in >=3% of subjects. This summary will be repeated separated by country.
- Summary of AEs by exposure at earliest onset.
- Summary of AEs by preferred term by descending frequency

There will be no statistical comparisons between the treatment groups for safety data, unless otherwise specified with the relevant section.

16.1. ADVERSE EVENTS

Adverse Events (AEs) will be coded using Medical Dictionary for Regulatory Activities (MedDRA) central coding dictionary, Version 19.1.

Treatment-emergent adverse events (TEAEs) are defined as AEs with a start date on or after the date of first dose through 7 days after study drug discontinuation (14 days for serious adverse events and deaths) or through the last double-blind dose date for patients continuing into the extension study. All AEs with an onset date after the consent date for the extension study but prior to the first dose in the extension study will be summarized separately in the extension study. Adverse events and serious adverse events (SAEs) with completely missing onset dates will be summarized as treatment emergent regardless of severity or relationship to study medication, unless the AE stop date occurs before the treatment start date.

See Appendix 2 for handling of partial dates for AEs. In the case where it is not possible to define an AE as treatment emergent or not, the AE will be classified by the worst case; i.e. treatment emergent.

An overall summary of patients within each of the categories described in the sub-section below, will be provided as specified in the templates.

Listings will include TEAEs and Non-TEAEs.

16.1.1. ALL TEAEs

Incidence of TEAEs will be presented by System Organ Class (SOC) and Preferred Term (PT) and also broken down further by maximum severity and relationship to study medication. This summary will be repeated separated by country.

16.1.1.1. Severity

Severity is classed as mild/ moderate/ severe (increasing severity). TEAEs starting after the first dose of study medication with a missing severity will be classified as not specified. If a subject reports a TEAE more than once within that SOC/ PT, the AE with the worst case severity will be used in the corresponding severity summaries. In determining maximum severity, response values will be ranked in order from minimum severity to maximum severity as values: 'not specified', 'mild', 'moderate', and 'severe'.

16.1.1.2. Relationship to Study Medication

Relationship, as indicated by the Investigator, is classed as related or not related. The categories of 'possible', 'probable', and 'definite' will be pooled to create the 'related' category. Similarly, 'not related' and 'unlikely' will be pooled to create the 'not related' designation. If a subject has more than one AE within an SOC and preferred term, the subject will be counted once according to the highest relationship. Adverse events with missing relationship will be classified as 'not specified'. In determining highest relationship, response values will be ranked in order from minimum relationship to maximum relationship as values: 'not specified', 'not related', and 'related'.

16.2. DEATHS

If any patients die during the study, the information will be presented in a data listing. Deaths will be recorded on the AE panel as an AE leading to death

16.3. LABORATORY EVALUATIONS

Results from the central laboratory will be included in the reporting of this study for Hematology, Blood Chemistry and Urinalysis. Endocrine parameters will be summarized separately. A list of laboratory assessments to be included in the outputs is included in Table 6 of the protocol.

A LOCF endpoint will also be derived for laboratory evaluations. For calculation details, please see section 6.3.

Presentations will use reported units.

Quantitative laboratory measurements reported as "< X", i.e. below the lower limit of quantification (BLQ), or "> X", i.e. above the upper limit of quantification (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as "< X" or "> X" in the listings.

The following summaries will be provided for laboratory data:

- Baseline and change from baseline by visit (for quantitative measurements)
- Incidence of abnormal values according to normal range criteria (Urinalysis parameters only)
- Shift from baseline according to normal range criteria (for quantitative measurements and categorical measurements)
- Summary of subjects meeting markedly abnormal criteria (for quantitative measurements and categorical measurements)
- Listing of patients meeting markedly abnormal criteria

Additionally, the change from baseline value at LOCF endpoint for selected laboratory parameters will be evaluated using a nonparametric rank ANCOVA. For each pairwise comparison versus placebo, values of the change from baseline at LOCF endpoint and baseline values will be ranked. A linear regression will be conducted on the change from baseline ranks with baseline value rank as independent variable to produce regression residuals. Using the values of the residuals as scores, Mantel-Haenszel row mean score tests will be produced for each SM-13496 group versus placebo (Stokes, Davis, & Koch, 1995). This will be conducted for prolactin, glucose, HbA1c, and lipid tests (HDL cholesterol, LDL cholesterol, total cholesterol, and triglycerides), glycoalbumin, and HOMA-IR. Prolactin will be assessed separately by sex.

The following parameters are presented split by fasting status and overall: HDL cholesterol, LDL cholesterol, total cholesterol, insulin, triglycerides, glucose and HOMA-IR. For laboratory variables split by fasting, change from baseline will only be calculated where the post-baseline fasting status matches the baseline fasting status.

16.3.1. LABORATORY SPECIFIC DERIVATIONS

A homeostasis model assessment of insulin resistance (HOMA-IR) (Matthews, Hosker, Rudenski, Naylor, Treacher, & Turner, 1985) parameter must also be calculated and included in summaries and rank ANCOVA analyses. The calculation of the parameter is as follows:

HOMA-IR = Glucose (mg/dL) x Insulin (mU/L) / 405

Also note that Prolactin data is considered a potential source of unblinding, so will not be available until after the study has unblinded.

16.3.2. LABORATORY REFERENCE RANGES AND MARKEDLY ABNORMAL CRITERIA

Quantitative laboratory measurements will be compared with the relevant laboratory reference ranges and categorized as:

- Low: Below the lower limit of the laboratory reference range.
- Normal: Within the laboratory reference range (upper and lower limit included).
- High: Above the upper limit of the laboratory reference range.

In addition to the high and low quantitative laboratory assignments (as identified by means of the laboratory reference ranges), markedly abnormal quantitative safety (and other) laboratory assessments will also be identified. Markedly Abnormal Post-baseline Laboratory Values (MAPLV) for selected laboratory parameters can be found in Table C:. The criterion of MAPLV is satisfied if a value falls into the markedly abnormal range. Patients will be represented in the count of a particular MAPLV if they have experienced that MAPLV at least once during the post-baseline double-blind phase, regardless of baseline value, up to and including LOCF endpoint. The number and percentage of patients with MAPLV will be presented by treatment group and for the combined SM-13496 group.

Table C: Criteria for Markedly Abnormal Post-Baseline Laboratory Values

Hematology Parameter	Markedly Abnormal Range
Hemoglobin	Male: ≤ 11.5 g/dL
	Female: ≤ 9.5 g/dL
Hematocrit	Male: ≤ 37%
	Female: ≤ 32%
WBC	≤2.8 x 10^9/L
	≥ 16 x 10^9/L
Neutrophils (percent)	≤ 15%
Eosinophils (percent)	≥ 10%
Platelet Count	≤ 75 x 10^9/L
	≥700 x 10^9/L
Clinical Chemistry Parameter	**Markedly Abnormal Range**
Alkaline Phosphatase	≥ 3 x ULN

ALT	≥ 3 x ULN
AST	≥ 3 x ULN
Total Bilirubin	≥ 2.0 mg/dL
Albumin	< 50% LLN
Glucose (fasting)	< 45 mg/dL
	>160 mg/dL
Sodium	< 130 mmol/L
	> 150 mmol/L
Potassium	< 3 mmol/L
	> 5.5 mmol/L
Chloride	< 90 mmol/L
	> 115 mmol/L
Calcium	< 8.4 mg/dL
	> 11.5 mg/dL
Blood Urea Nitrogen	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Creatine Phosphokinase	≥ 3 x ULN
LDH	≥ 3 x ULN
HbA1c	≥ 7.5%
Prolactin	≥ 5 x ULN
Total Cholesterol (fasting)	≥ 300 mg/dL
LDL Cholesterol (fasting)	≥ 200 mg/dL
Triglycerides (fasting)	≥ 300 mg/dL
Urinalysis Parameter	**Markedly Abnormal Range**
RBC	> 25 hpf
WBC	> 25 hpf

16.3.3. Subgroup Analysis of Laboratory Values

The following laboratory parameters will be evaluated for the subgroups specified in section 7.5: serum prolactin, blood glucose, HbA1c(NGSP), glycoalbumin (absolute and percentage), total cholesterol, LDL cholesterol, HDL cholesterol, HOMA-IR and triglycerides. These parameters will be separated by fasting status for the subgroup analyses, with the exception of prolactin, HbA1c and glycoalbumin.

For these subgroup analyses, these selected parameters will be summarized by Observed and change from baseline at Week 6 (LOCF), including a nonparametric rank ANCOVA test. The same analysis methods will be performed as detailed in section 15.2.4.

16.4. ECG Evaluations

Results from the central [redacted] will be included in the reporting of this study.

A LOCF endpoint will also be derived for ECG. For calculation details, please see section 6.3.

The following ECG parameters will be reported for this study:

- Heart Rate (bpm)
- PR Interval (msec)
- QRS Interval (msec)
- RR Interval (msec)
- QT Interval (msec)
- QTcF Interval (msec)
- QTcB Interval (msec)
- Overall assessment of ECG:
  - Normal
  - Abnormal

The following summaries will be provided for ECG data:

- Observed baseline and change from baseline by visit (for quantitative measurements)
- Shift from baseline according to abnormal criteria (for overall assessment)
- Incidence of QTc Prolongation (see section 16.4.3)
- Summary of patients with abnormal ECG values (see section 16.4.2)
- Listing of patients with prolonged QTc results

16.4.1. ECG Specific Derivations

All required ECG parameters are available in the data provided for analysis. There are no ECG parameters to be derived.

16.4.2. ECG Abnormal Criteria

The individual ECG measurements would be noted as abnormal if exceeding the values indicated in Table D:.

Table D: Abnormal ECG Values by parameter

ECG Parameter	Abnormal
HR (bpm)	≥ 100 bpm
PR interval (msec)	≥ 210 msec
QRS interval (msec)	≥ 120 msec

QT interval (msec)	> 500 msec

16.4.3. ECG MARKEDLY ABNORMAL CRITERIA

Markedly abnormal quantitative ECG measurements will be identified in accordance with the following predefined markedly abnormal criteria:

- Absolute values for QTcB interval and QTcF will be classified as:

  - > 450 msec
  - > 480 msec
  - > 500 msec

- Change from Baseline for QTcB interval and QTcF will be classified as:

  - >=30 msec increase from baseline
  - >-60 msec increase from baseline

16.4.4. SUBGROUP ANALYSIS OF ECG VALUES

Observed baseline and change from baseline by visit analyses will be repeated by subgroups as defined in section 7.5 for QTcB interval and QTcF interval only. For the methodology please see section 15.2.4.

16.5. VITAL SIGNS

The following Vital Signs measurements will be reported for this study:

- Systolic Blood Pressure (mmHg)
- Diastolic Blood Pressure (mmHg)
- Pulse Rate (bpm)
- Temperature (°C)
- Weight (kg)
- BMI (kg/m^2)

The following summaries will be provided for vital signs data:

- Observed baseline and change from baseline by visit
- Incidence of markedly abnormal values
- Listing of patients meeting markedly abnormal post-baseline vital signs (MAPVS) criteria

Additionally, a rank ANCOVA (vs placebo) analysis will be performed for weight and BMI.

16.5.1. Vital Signs Specific Derivations

No additional vital signs parameters are required.

16.5.2. Vital Signs Markedly Abnormal Criteria

Markedly abnormal quantitative Vital Signs measurements will be identified in accordance with the following predefined markedly abnormal criteria.

Variable	Unit	Low	High
SBP	mmHg	≤ 90 mmHg AND change from baseline ≤ -20 mmHg	≥ 180 mmHg AND change from baseline ≥ 20 mmHg
DBP	mmHg	≤ 50 mmHg AND change from ≤ -15 mmHg	≥ 105 mmHg AND change from baseline ≥ 15 mmHg
Pulse rate	Bpm	≤ 50 bpm AND change from baseline ≤ -15 bpm	≥ 120 bpm AND change from baseline ≥ 15 bpm
Body temperature	°C	NA	≥ 38.3 °C AND change from baseline ≥ 0.8 °C
Weight	Kg	percentage change from baseline ≤ -7.0 %	percentage change from baseline ≥ 7.0 %

16.5.3. Subgroup Analysis of Vital Signs Values

Observed baseline and change from baseline by visit analyses will be repeated by subgroups as defined in section 7.5 for weight and BMI only. For the methodology please see section 15.2.4.

16.6. Physical Examination

Physical Examination results will be listed.

16.7. OTHER SAFETY ASSESSMENTS

16.7.1. DRUG-INDUCED EXTRAPYRAMIDAL SYMPTOM SCALE (DIEPSS)

The DIEPSS is a clinician-rated assessment of extrapyramidal symptoms induced by antipsychotics and consists of eight individual parameters; gait, bradykinesia, sialorrhea, muscle rigidity, tremor, akathisia, dystonia, and dyskinesia; and one global assessment; overall severity. The severity of each item is graded 0 (normal) to 4 (severe), and then a summation of all questions (excluding overall severity) provides the total score which is used for analysis. The investigator will assess the DIEPSS at Visits 2 to 8 and at discontinuation. The assessment at Visit 8 or at discontinuation should be performed within 72 hours after the final administration of the study drug and before the initiation of post treatment with antipsychotics or the study drug in the long-term study.

A LOCF endpoint will also be derived for DIEPSS. For calculation details, please see section 6.3.

16.7.2. MISSING DATA METHODS

Missing question data for questionnaires is accounted for in the variable descriptions. The methods used for analysis of this endpoint will address visit data that is missing. In the event that any component element of the Total Score (Excluding Overall Severity) is missing, then this total score will be set to missing.

16.7.3. ANALYSIS OF QUALITY OF LIFE VARIABLES

16.7.3.1. Analysis of DIEPSS

The change from baseline in the DIEPSS total score (excluding overall severity) and the individual DIEPSS scores will be analyzed using the MMRM method described above for the primary efficacy variable. The change from baseline in the DIEPSS total score will also be analyzed using the LOCF ANCOVA method described for the primary efficacy variable, as a sensitivity analysis.

16.7.3.2. Subgroup Analysis of DIEPSS

Observed baseline and change from baseline by visit analyses will be repeated by subgroups as defined in section 7.5 for DIEPSS scores (excluding overall severity). For the methodology please see section 15.2.4.

16.7.4. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

The C-SSRS is a tool designed to systematically assess and track suicidal adverse events (suicidal behavior and suicidal ideation) throughout the study. The C-SSRS can comprehensively identify suicidal events and limit the over-identification of suicidal behavior. The rater will assess the C-SSRS at Visits 1 to 8 and at discontinuation. The 'baseline/screening' version will be used at Visit 1 and the 'since last visit 'version at Visits 2 to 8 and at discontinuation. The assessment at Visit 8 or at discontinuation should be performed within 72 hours after the final administration of the study drug and before the initiation of post treatment with antipsychotics or the study drug in the long-term study.

Suicidal ideation is rated on a 6-point scale from 0='No ideation present'' to 5='Active ideation with plan and intent'. A score of 4 or 5 on this scale indicates serious suicidal ideation. Intensity of ideation is measured in

terms of frequency, duration, controllability, deterrents, and reasons for ideation. Each is measured with responses ranging from 0 or 1 to 5, representing approximately less to more. The ideation intensity total score is the sum of the five preceding items and can range from 3 to 25 for patients with suicidal ideation endorsed.

Suicidal behavior is collected as presence/absence of actual attempts, non-suicidal self-injurious behavior, interrupted attempts, aborted attempts, preparatory acts or behavior, and any suicidal behavior. In addition, the number of actual attempts, interrupted attempts, and aborted attempts is captured. Any attempt will be defined as suicidal behavior. The lethality associated with actual attempts is rated on a 6-point scale from 0='No physical damage or very minor physical damage' to 5='Death'. Potential lethality of attempts is rated on a 3-point scale from 0='Behavior not likely to result in injury' to 2='Behavior likely to result in death despite available medical care'.

A composite measure of suicidality measures the presence of any suicidal ideation or behavior. The number and percentage of subjects with suicidality as measured by the C-SSRS will be summarized and compared among the SM-13496 treatment groups versus placebo using Fisher's exact test, where suicidality is defined as having at least one occurrence of suicidal ideation or at least one occurrence of suicidal behavior. The baseline suicidality indicator will be set to 1 if the subject exhibits suicidality during the two weeks prior to the baseline visit, 0 if the subject does not exhibit suicidality during the two weeks prior to the baseline visit, and missing otherwise. The post-baseline suicidality indicator is similarly defined for each post-baseline weekly visit. This analysis will be performed at Week 6, LOCF endpoint, and overall (based on all post-baseline visits up to and including LOCF endpoint) by treatment group and combined SM-13496 group.

An overall summary of post-baseline data across all visits up to and including LOCF endpoint will include the frequency and percentage of the following by treatment group and for the combined SM-13496 group:

- At least one suicidal ideation post-baseline
- No suicidal ideation post-baseline
- Emergence of suicidal ideation (no suicidal ideation at baseline, and any type of suicidal ideation post-baseline)
- Emergence of serious suicidal ideation (no suicidal ideation at baseline, and any serious suicidal ideation [ideation score of 4 or 5] post-baseline)
- Most severe type of ideation post-baseline
- Worsening of suicidal ideation (most severe suicidal ideation post-baseline was more severe than it was at baseline)
- At least one suicidal behavior post-baseline
- No suicidal behavior post-baseline
- Emergence of suicidal behavior (no suicidal behavior at baseline, and any type of suicidal behavior post-baseline)
- At least one actual attempt post-baseline
- At least one interrupted attempt post-baseline
- At least one aborted attempt post-baseline
- At least one preparatory act or behaviors post-baseline
- At least one instance of suicidality [any ideation or behavior] post-baseline
- No suicidality post-baseline

- Emergence of suicidality (no suicidality at baseline, and any suicidality post-baseline)
- Any completed suicides post-baseline

16.7.5. Treatment Emergent Mania

Treatment-emergent mania is observed as either of the following:

- A YMRS total score ≥16 at 2 or more consecutive assessments.
- Treatment emergent adverse events related to mania symptoms, as assessed by a medical reviewer.

Number and percentage of patients with treatment emergent mania will be summarized by treatment group. Additionally, a logistic regression analysis will be performed with effects for treatment group, pooled center, and baseline YMRS total score. Contrasts will be evaluated for each SM-13496 treatment group compared to placebo. Odds Ratios, 95% CIs and Wald chi-square p-values will be presented for each contrast.

If the model fails to converge, the model will be attempted replacing pooled center with country. If model still does not converge, use geographic region (Asia, Non-Asia). Japan, Malaysia, Philippines and Taiwan are classified as belonging to Asia while Ukraine, Russia, Lithuania and Slovakia are classified as Non-Asian countries. Only n and percent will be presented in cases where the number of subjects meeting the above composite variable criteria is too small for an analysis to be performed.

16.8. Pharmacokinetic Data

Samples for pharmacokinetic analysis are drawn at Week 3, Week 4 and Week 6/ Early Term. The data will be summarized, separated by timepoint and SM-13496 dose at time of sample collection.

Samples for pharmacokinetic analysis are drawn at Week 3, Week 4 and Week 6/ Early Term. For purposes of analysis, windows will be used based on scheduled visit and collection timepoint to select samples for summarization (see Table 12). All samples collected within these windows will subsequently be summarized. Samples collected outside the specified windows will be included only in a data listing. Samples collected from early terminated subjects should not be included in summaries and figures.

For blood samples, time in hours since last dose will be summarized by most recent assigned dose level and country (overall and each country) to assess whether samples were collected according to protocol-defined intervals. For inclusion at a given visit and collection timepoint, subjects will be required to have at least 5 days of exposure at the assigned dose level prior to sample collection.

Blood concentrations of SM-13496 and selected quantifiable metabolites (which may include, but are not limited to, ID-11614, ID-14283, and ID-14326) will be summarized at each visit by most recent assigned dose level. For inclusion at a given visit and collection timepoint, subjects will be required to have at least 5 days of exposure at the assigned dose level prior to sample collection. SM-13496 treatment group (actual dose) mean pharmacokinetic sample concentrations and respective standard deviations will be plotted by visit and dose and country (overall and each country) using box plot style in the Safety population. Concentrations will also be plotted by visit, dose and country using a box plot style based on the Safety population.

Table 12 - Pharmacokinetic Sample Windows to Select Data for Summarization

Visit	Sample Windows
Week 3	Within 13-15 hours post-dose
Week 4	Within 13-15 hours post-dose
Week 6	Within 13-15 hours post-dose

17. DATA NOT SUMMARIZED OR PRESENTED

The other variables and/or domains not summarized or presented are:

- Comments
- MINI screening questionnaire

These domains and/or variables will not be summarized or presented, but will be available in the clinical study database, SDTM and/or ADaM datasets.

18. REFERENCES

Allison, P. (2010). Missing Data. In P. Marsden, & J. Wright, *Handbook of Survery Research* (pp. 631-657). Bingley, UK: Emerald Group Publishing Ltd.

Dobson, A. J. (2002). *An introduction to generalized linear models, 2nd ed.* Chapman & Hall/CRC.

EMA. (n.d.). *Guideline on Missing Data in Confirmatory Clinical Trials.* Retrieved December 15, 2011, from European Medicines Agency: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf

Hedeker, D., & Gibbons, R. (1997). Application of random-effects pattern mixture models for missing data in longitudinal studies. *Psychological Methods*, 64-78.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 800-2.

Hogan, J., & Laird, N. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 239-258.

Little, R., & Yau, L. (1996). Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs. *Biometrics*, vol 52, 1324-1333.

Matthews, D., Hosker, J., Rudenski, A., Naylor, B., Treacher, D., & Turner, R. (1985). Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 412-419.

O'Kelly, M., & Ratitch, B. (2014). *Clinical Trials with Missing Data - A Guide for Practitioners.* Wiley.

Ratitch, B., & O'Kelly, M. (2011). Implementation of pattern-mixture models using standard SAS/STAT procedures. *SUGI*.

Siddiqui, O., Hung, H., & O'Neil, R. (2009). MMRM vs LOCF: A comprehensive comparison based on simulation study and 25 NDA data sets. *Journal of Biopharmaceutical Statistics*, 227-246.

Stokes, M., Davis, C., & Koch, G. (1995). *Categorical Data Analysis using the SAS system.* Cary, NC: SAS Institute Inc.

APPENDIX 1. PROGRAMMING CONVENTIONS FOR OUTPUTS

OUTPUT CONVENTIONS

Outputs will be presented according to the following conventions, which follow Sumitomo output conventions but use information from previous SM-13496 study layout:

1. ABBREVIATIONS

- ASCII American standard code for information interchange file format
- CGM Computer graphics metafile
- ODS Output Delivery System
- RTF Rich text file format

2. INTRODUCTION

This document applies to standards used for outputting tables, listings and figures. It is intended to provide specifications to guide the statistician or statistical programmer in setting up specifications for programming tables, listings and figures. These standards should be used in the absence of customer specific standards.

3. OUTPUT FILE NAMING CONVENTIONS

File names should only consist of lowercase letters, digits (0 to 9) and dashes. A period should only be used to indicate a separator between the file name and the extension. No spaces, other special characters or punctuation marks are permitted. As dashes are not permitted in SAS program names, they will be replaced by underscores.

Output files should be in RTF format.

The program, program log and output file name should reflect the type and number of the statistical output. If this is not possible, then the output name should be at least as descriptive as possible. A prefix can be used to distinguish between a Table, Listing and Figure document ('t' for table, 'l' for listing and 'f' for figure). If there is only 1 digit in the number of the table, listing or figure in the place where 2 digits are possible, a leading zero should be added in the file name to make sorting consistent with the sequence (eg t14-3-01-1.RTF)

4. PAPER SIZE, ORIENTATION AND MARGINS

- The size of paper will be A4.
- The page orientation should preferably be landscape, but portrait is also permitted.

- Margins should provide at least 3 centimeters of white space all around the page, regardless of the paper size.
- The number of columns per page (linesize) should be 132 for A4.
- The number of rows per page (pagesize) should be 46 for A4.

5. FONTS

The font type 'Times New Roman should be used as a default for tables and listings, with a font size of 8. The font color should be black. No bolding, underlining italics or subscripting should be permitted. Try to avoid using super-scripts, unless absolutely necessary. Single spacing should be used for all text.
Figures should have a default font of "Times Roman", "Helvetica", or "Courier New" .
This can be achieved by using the following options in SAS:

```
goptions
gunit = pct
cback = white
colors = (black)
hby = 2.4
ftext = "TimesRoman"
htext = 2.5
device = cgmof97l
gaccess = gsasfile;
filename gsasfile "....cgm";
```

6. HEADER INFORMATION

Headers should be defined as follows:

- The header should be placed at the top of the page (same place on each page) regardless of the size or orientation of the table or listing
- The customer name and protocol number should appear in row 1, left-aligned
- The output identification number should appear in row 2, centered
- The output title should start in row 3, centered
- The output population should appear in row 4, centered. The population should be spelled out in full, e.g. Intention-to-Treat in preference to ITT.
- Row 5 should be a continuous row of underscores ('_') (the number of underscores should equal the linesize)
- Row 6 should be a blank line
- Mixed case should be used for titles
- The output titles should be designed so that they are arranged consistently through all outputs. For example, content (eg Vital Signs) followed by metric (eg Change from Baseline) e.g. Vital Signs – Change from Baseline.

- Titles should not contain quotation marks or footnote references
- The column headings should be underlined with a row of underscores ('_')
- Column headings spanning more than one column should be underlined and have underscores on either side of the title and should be centered
- Column headings containing numbers should be centered
- Column headings should be in proper case
- In general, the population count should appear in the column header in the form "(N=XXX)"
- "Statistic" should be the column header over n, Mean, SE, n (%) etc.
- As a rule, all columns should have column headings.
- The header of the outputs will be in a non-selectable format for all pages after page one.

7. TABLE AND LISTING OUTPUT CONVENTIONS

General:

- The first row in the body of the table or listing should be blank
- The left hand column should start in column 1. No indenting or centering of the output should occur.
- Rounding should be done with the SAS function ROUND.
- Numbers in tables should be rounded, not truncated.
- Alphanumeric output should be left aligned.
- Numbers should be decimal point aligned.
- Whole numbers should be right aligned.
- Text values should be left aligned.
- The first letter of a text entry should be capitalized
- Listings of adverse events, concomitant medications, medical histories etc. should be sorted in chronological order, with earliest adverse event, medication or history coming first.
- Placebo should appear first in tables with treatments as columns
- In general, only present totals (across treatment groups) at baseline/randomization, and do not present them post randomization, unless the customer specifically requests it.

- If possible, include 100% frequencies in the table shell, so that it is clear what the denominator is for percentage calculations.
- All listing outputs should be sorted (preferably by Treatment, Site Number and Subject Number).
- Exponentiation will be expressed using a double asterisk, i.e., mm3 will be written as mm**3.
- All variables that are output in the CRF (which have data present) should appear in the listings, along with all derived data appearing in the corresponding tables
- The width of the entire output should match the linesize

Univariate Statistics:

- Statistics should be presented in the same order across tables (i.e., n, Mean, SD, Median, Minimum, Maximum)
- Table statistics should line up under the N part of the (N=XXX) in the table header. All decimal points should line up. If the minimum and maximum are output on one line as Minimum, Maximum then the comma should line up with the decimal point. It is accepted that using the Times New Roman font there will be limitations to this, though it will be followed as far as is possible.
- If the original data has N decimal places, then the summary statistics should have the following decimal places:
- Minimum and maximum: N
- Mean, median and CV%: N + 1
- SD: N + 2

Frequencies and percentages (n and %):

- Percent values should be reported inside parentheses, with one space between the count and the left parenthesis of the percentage. Parentheses should be justified to accept a maximum of 100.0 as a value and padded with blank space if the percent is less than 100.0. An example is given below:

77 (100.0%)

50 (64.9%)

0 (0.0%)

- Percentages will be reported to one decimal place, except percents <100.0% but >99.9% will be presented as '>99.9%' (e.g., 99.99% is presented as >99.9%); and percents < 0.1% will be presented as '<0.1%' (e.g., 0.08% is presented as <0.1%). Rounding will be applied after the <0.1% and >99.9% rule.

Eg (<0.1%)

(6.8%)

(>99.9%)

- Percentages may be reported to 0 decimal places as appropriate (for example, where the denominator is relatively small).
- Where counts are zero, percentages will not appear in the output.

Confidence Intervals:

- As a rule confidence intervals are output to one place more than the raw data, and standard deviations and standard errors to two places more than the raw data
- Confidence intervals should be justified so that parentheses displayed on consecutive lines of a table "line up".
- Boundary values of confidence intervals should be separated by a comma.
- Boundary values should be padded as necessary to accept negative values and to allow alignment of the decimal place.
- An example is given below:

(-0.12, -0.10)
(9.54, 12.91)

P-values:

- P-values should be reported to three decimal places, except values <1.000 but >0.999 will be presented as '>0.999' (e.g., 0.9998 is presented as >0.999); and values <0.001 will be presented as '<0.001' (e.g., 0.0009 is presented as <0.001). Rounding will be applied after the <0.001 and >0.999 rule

Ratios:

- Ratios should be reported to one more decimal place than the original data.

Spacing:

- There must be a minimum of 1 blank space between columns (preferably 2)

Denominators:

- If a different count other than the population count is used for a denominator (within the table) to calculate percentages, there should be a row in the table that identifies that number "n".
- Alternatively, a footnote should be included in each table with percentages to indicate the denominator for percentages.

Missing values:

- A "0" should be used to indicate a zero frequency.
- A blank will be used to indicate missing data in an end-of-text table or subject listing.
- In the case that only a single value is used in the creation of summary statistics (i.e. n=1) then only n and the mean will be presented.

8. FIGURE OUTPUT CONVENTIONS

- Figures should be provided in RTF files using the SAS Output Delivery System (ODS), as Computer Graphics Metafile (CGM) formatted graphical output generated by SAS.
- The CGM file itself should contain the title or footer.
- The image should be clear and of high quality when viewed in the Word document, and when printed.
- In general, boxes around the figures should be used.

9. FOOTNOTE INFORMATION

Footers should be defined as follows:

- A continuous line of underscores ('_') will follow the body of the table or listing prior to any footnotes at the bottom of the page
- Footnotes will only appear on the last page of the output.
- Table footnotes should be defined using compute statements in the proc report, and should appear directly after the body of the table.
- The program path and name and version number (if applicable) should appear as footnote 1 at the bottom of the page
- The date/time stamp should appear as footnote 2 at the bottom of the page

- Footnotes should be left-aligned.
- Footnotes should be in sentence case.
- Superscripts are used for footnotes linking to the body of the table.
- The choice of footnote symbols should be consistent. E.g. if you have the footnote "# indicates last observation carried forward" for one table, the same symbol and footnote should indicate LOCF for all tables.
- If text wraps across more than one line (for a note), the first letter for all lines of text after the first one will be indented to align beneath the first letter of the text in the first line.
- The page identification in the format Page X of Y (where Y is the total number of pages for the output) will appear at the top of the page, right aligned
- Common notes from table to table should appear in the same order.
- The symbols should appear in the same order as what they are defined in the table or listing, from left to right.

10. PROGRAMMING INSTRUCTIONS

Programming instructions must appear in blue font at the end of each table or listing shell. Programming instructions, where necessary, should follow the table or listing shells in blue font, beginning with the words "Programming Note" followed by a colon. These include notes on the output, reminders of how to handle missing values, repeat shells for similar tables etc.

DATES & TIMES

Depending on data available, dates and times will take the form yyyy-mm-ddThh:mm:ss.

SPELLING FORMAT

English US

PRESENTATION OF TREATMENT GROUPS

For outputs, treatment groups will be represented as follows and in that order:

Treatment Group	For Tables and Graphs	For Listings (include if different to tables)
Placebo	Placebo	Placebo

Treatment Group	For Tables and Graphs	For Listings (include if different to tables)
SM-13496 20-60mg	SM-13496 20-60mg	SM-13496 20-60mg
SM-13496 80-120mg	SM-13496 80-120mg	SM-13496 80-120mg
Combined SM-13496	Combined SM-13496	-
Screen Failure	-	Not Randomized

PRESENTATION OF VISITS

For outputs, visits will be represented as follows and in that order:

Long Name (default)	Short Name
Screening	Scr
Baseline	BL
Week 1	W1
Week 2	W2
Week 3	W3
Week 4	W4
Week 5	W5
Week 6	W6
Week 6 (LOCF)	W6LOCF

LISTINGS

All listings will be ordered by the following (unless otherwise indicated in the template):

- randomized treatment group (or treatment received if it's a safety output), first by active dose [by ascending dose group] and then control/ placebo
- center-subject ID,
- visit (where applicable),
- date (where applicable),
- For listings where non-randomized patients are included, these will appear in a category after the randomized treatment groups labeled 'Not Randomized'.

APPENDIX 2. PARTIAL DATE CONVENTIONS

Imputed dates will NOT be presented in the listings.

ALGORITHMS OF PARTIAL DATES FOR AEs AND MEDICATIONS

Partial dates for AEs and medications will be imputed as follows:

Adverse event or prior/concomitant medication start date (references to year and month refer to the year and month of the start date, respectively):

1. If year and month are known, and it is the month of the first dose date and the stop date is prior to the first dose date, use the first day of the month.
2. If year and month are known, and it is the month of the first dose date and the stop date is the same as or later than the first dose date, use the first dose date.
3. If year and month are known, and it is the month following the first dose date, use the first day of the month.
4. If year and month are known and it is any month prior to first dose date, use the first day of the month.
5. If only year is known, and it is previous to the year of the first dose date, use June 30th of that year. If it is the same as first dose date year, assume it is the first dose date. If it is later than the first dose date year, assume it is the first day of the year.
6. If the start date is completely missing, use the first dose date.
7. Should any of the previous start dates created be after a complete stop date provided, use the stop date instead of the date that would otherwise be created.
8. Otherwise, if start date is unknown leave as missing.

Adverse event or prior/concomitant medication stop date (references to year and month refer to the year and month of the stop date, respectively):

1. If year and month are known and study medication stopped during that month, use the stop date of study medication.
2. If year and month are known and study medication stopped after that month, use the last day of the month.
3. If year and month are known and study medication stopped prior to that month, use the first day of the month.
4. If only year is known, and it is the same as last dose date year, assume it is the last dose date. If it precedes the last dose date year, assume it is the last day of the year. If it is later than the last dose date year, assume it is the first day of the year.
5. If stop date is unknown and continuing is checked, set to date of last double-blind dose.
6. Should any of the previous stop dates created come before a start date, either a complete date or an imputed one, use the start date instead of the date that would otherwise be created.
7. Otherwise, if stop date is unknown leave as missing.

ALGORITHM FOR PARTIAL DATES IN MEDICAL HISTORY

Partial dates for medical history (including psychiatric history) will be imputed as follows:

Medical history start date (references to year and month refer to the year and month of the start date, respectively):

1. If year and month are known, and it is the year or month previous to Screening, use the last day of the month.
2. If year and month are known, and it is the month of Screening, use Screening date – 1.
3. If only year is known, and it is previous to the year of Screening, use June 30th of that year.
4. If only year is known, and it is the year of Screening, use Screening date – 1.

APPENDIX 3. SAS CODE

Treatment: TRTPN
Analysis visit: AVISIT

1. Mixed Model Repeated Measures (MMRM) Analysis
Note: Ensure that input dataset is sorted by USUBJID AVISIT prior to initiating the MIXED procedure.

```
ods output estimates=outset  /*Output from estimate statement*/
           diffs=outdiff     /*Treatment Group Comparison*/
           R=outr    /*R matrix */
           lsmeans=outlsmean;/*LSmeans of each treatment group*/

PROC MIXED;
  CLASS SUBJID POOLSITE TRTPN AVISIT;
  MODEL CFB = BASE POOLSITE TRTPN AVISIT TRTPN*AVISIT/DDFM=KR;
  REPEATED AVISIT /TYPE=UN SUBJECT=SUBJID R;
  LSMEANS TRTPN* AVISIT /PDIFF CL;
  ESTIMATE 'LINEAR TREND TEST FOR DOSE RESPONSE'
  TRTPN  0 1 -1 TRTPN*AVISIT 0 0 0 0 0 0 0  0 0 0 0 0 0 1
  0 0 0 0 0 0 -1/CL;
  ESTIMATE 'DOSE-RESPONSE 20-60 vs 80-120' TRTPN -1 1 0 TRTPN*AVISIT
  0 0 0 0 0 0 -1  0 0 0 0 0 0 1  0 0 0 0 0 0 0/CL;
RUN;
```

a. Mixed Model Repeated Measures (MMRM) Analysis for the case above model fails to converge

```
PROC MIXED empirical;
```

```
      CLASS SUBJID POOLSITE TRTPN AVISIT;
      MODEL TOTCFB = TOTBL POOLSITE TRTPN AVISIT TRTPN*AVISIT;
      REPEATED AVISIT /TYPE=XXX  SUBJECT=SUBJID R;
      LSMEANS TRTPN*AVISIT /PDIFF CL;
      ESTIMATE 'LINEAR TREND TEST FOR DOSE RESPONSE'
      TRTPN  0 1 -1 TRTPN*AVISIT 0 0 0 0 0 0 0  0 0 0 0 0 0 1
      0 0 0 0 0 0 -1/CL;
      ESTIMATE 'DOSE-RESPONSE 20 vs 80-160' TRTPN -1 1 0 TRTPN*
      AVISIT 0 0 0 0 0 0 -1  0 0 0 0 0 0 1  0 0 0 0 0 0 0/CL;
RUN;
```

XXX: application order
Heterogeneous Toeplitz (TYPE=TOEPH)
Heterogeneous first-order autoregressive (TYPE=ARH(1))
Toeplitz (TYPE=TOEP)

2. Analysis of Covariance

```
PROC MIXED;
      CLASS TRTPN POOLSITE;
      MODEL CFB = TRTPN POOLSITE BASE;
      LSMEANS TRTPN / PDIFF CL;
      ESTIMATE 'LINEAR TREND' TRTPN 0 1 -1 /CL;
      ESTIMATE 'DOSE-RESPONSE 20 vs 80-160' TRTPN -1 1 0/CL;
RUN;
```

3. Sensitivity Analysis

a. Random Effects Pattern Model

```
If TRT01PN=1 then DOSE1=1 else DOSE1=0;
If TRTO1PN=2 then DOSE2=1 else DOSE2=0;
If COMPLFL="Y" then DROP=1 else DROP=0;

PROC MIXED;
      CLASS USUBJID;
      MODEL AVAL = QSDAY DOSE1 DOSE2 QSDAY*DOSE1 QSDAY*DOSE2
                   DROP DROP*QSDAY DROP*DOSE1 DROP*DOSE2
                   DROP*DOSE1*QSDAY DROP*DOSE2*QSDAY / SOLUTION COVB
                   DDFM=KR;
      RANDOM INTERCEPT DAY/SUB=USUBKID TYPE=UN G;
RUN;

PROC IML;
USE VCOV1;
READ all into vcovm;
USE estb;
READ all into estbm;

MEANV = ESTBM[,1];
```

```
VARB  = VCOVM[1:12,1:12];
A = { 1 0 0 0 0 0  0 0 0 0 0 0,   /* Completer - control intercept      */
      0 1 0 0 0 0  0 0 0 0 0 0,   /* Completer - control slope          */
      0 0 1 0 0 0  0 0 0 0 0 0,   /* Completer - dose1   intercept diff */
      0 0 0 1 0 0  0 0 0 0 0 0,   /* Completer - dose2   intercept diff */
      0 0 0 0 1 0  0 0 0 0 0 0,   /* Completer - dose1   slope     diff */
      0 0 0 0 0 1  0 0 0 0 0 0,   /* Completer - dose2   slope     diff */

      1 0 0 0 0 0  1 0 0 0 0 0,   /* Dropout   - control intercept      */
      0 1 0 0 0 0  0 1 0 0 0 0,   /* Dropout   - control slope          */
      0 0 1 0 0 0  0 0 1 0 0 0,   /* Dropout   - dose1   intercept diff */
      0 0 0 1 0 0  0 0 0 1 0 0,   /* Dropout   - dose2   intercept diff */
      0 0 0 0 1 0  0 0 0 0 1 0,   /* Dropout   - dose1   slope     diff */
      0 0 0 0 0 1  0 0 0 0 0 1 }; /* Dropout   - dose2   slope     diff */

meancd = A*meanv;
varccd = A*varb*T(A);

*COMPLETERS;
meanc = meancd[1:6,];
serrc = sqrt(vecdiag(varccd[1:6,1:6]));
zvalc = meanc/serrc;
lowec=meanc-1.96*serrc;
uppec=meanc+1.96*serrc;
probc = 2*(1-PROBNORM(abs(zvalc)));

**DROPOUTS;
meand = meancd[7:12,];
serrd = sqrt(vecdiag(varccd[7:12,7:12]));
zvald = meand/serrd;
lowed=meand-1.96*serrd;
upped=meand+1.96*serrd;
probd = 2*(1-PROBNORM(abs(zvald)));


/* indicator matrix to average over 2 patterns */
B = { 1 0 0 0 0 0  1 0 0 0 0 0,
      0 1 0 0 0 0  0 1 0 0 0 0,
      0 0 1 0 0 0  0 0 1 0 0 0,
      0 0 0 1 0 0  0 0 0 1 0 0,
      0 0 0 0 1 0  0 0 0 0 1 0,
      0 0 0 0 0 1  0 0 0 0 0 1 };

/* derivative matrix for average with respect to proportion
   (for application of delta method)                        */
 C = { 1 0 0 0 0 0  -1 0 0 0 0 0,
       0 1 0 0 0 0  0 -1 0 0 0 0,
       0 0 1 0 0 0  0 0 -1 0 0 0,
       0 0 0 1 0 0  0 0 0 -1 0 0,
       0 0 0 0 1 0  0 0 0 0 -1 0,
       0 0 0 0 0 1  0 0 0 0 0 -1 };


propcd  = J(12,12,0);
```

```
propcd[1,1] =  &grpn11/(&grpn11+&grpn21); /* Completer - control intercept
*/
propcd[2,2] =   &grpn11/(&grpn11+&grpn21); /* Completer - control slope
*/
propcd[3,3] =  &grpn12/(&grpn12+&grpn22); /* Completer - dose1   intercept
diff */
propcd[4,4] =  &grpn13/(&grpn13+&grpn23); /* Completer - dose2   intercept
diff */
propcd[5,5] =   &grpn12/(&grpn12+&grpn22); /* Completer - dose1     slope
diff */
propcd[6,6] =   &grpn13/(&grpn13+&grpn23); /* Completer - dose2     slope
diff */

propcd[7,7] =    &grpn21/(&grpn11+&grpn21);      /* Dropout      - control
intercept      */
propcd[8,8] =   &grpn21/(&grpn11+&grpn21);    /* Dropout    - control slope
*/
propcd[9,9] =    &grpn22/(&grpn12+&grpn22);      /* Dropout      - dose1
intercept diff */
propcd[10,10] =     &grpn23/(&grpn13+&grpn23);  /* Dropout      - dose2
intercept diff */
propcd[11,11] =  &grpn22/(&grpn12+&grpn22); /* Dropout    - dose1    slope
diff */
propcd[12,12] =  &grpn23/(&grpn13+&grpn23); /* Dropout    - dose2    slope
diff */

propcdd  = J(6,6,0);
propcdd[1,1] =  &grpn11*&grpn21/((&grpn11+&grpn21)**3);
propcdd[2,2] =  &grpn11*&grpn21/((&grpn11+&grpn21)**3);
propcdd[3,3] =  &grpn12*&grpn22/((&grpn12+&grpn22)**3);
propcdd[4,4] =  &grpn13*&grpn23/((&grpn13+&grpn23)**3);
propcdd[5,5] =  &grpn12*&grpn22/((&grpn12+&grpn22)**3);
propcdd[6,6] =  &grpn13*&grpn23/((&grpn13+&grpn23)**3);

meanpm =  B*propcd*meancd;
varcpm =  B*propcd*varccd*propcd*T(B)
          + propcdd*C*meancd*T(meancd)*T(C);
serrpm = sqrt(vecdiag(varcpm));
zvalpm = meanpm/serrpm;
lowepm=meanpm-1.96*serrpm;
uppepm=meanpm+1.96*serrpm;
probpm = 2*(1-PROBNORM(abs(zvalpm)));

CREATE outdata var {meanc serrc lowec uppec probc meand serrd lowed upped
probd meanpm serrpm lowepm uppepm probpm};
      append;
        close outdata;
QUIT;

*Model repeated without drop out pattern:

   PROC MIXED;
      CLASS USUBJID;
      MODEL AVAL = QSDAY DOSE1 DOSE2 DAY*DOSE1 DAY*DOSE2 /
```

```
                    SOLUTION COVB DDFM = KR;
      RANDOM INTERCEPT DAY /SUB=USUBJID TYPE=UN G;
    RUN;
```

b. Multiple Imputation

```
NOTE: Placebo based imputation

PROC MI
VAR SITEGR1N SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5 SCORE_6
SCORE_7;
BY TRT01PN;
MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE;
RUN;

%MACRO MI_IMP(INDATA=, ANVIS=, SEED=, VAR1=, VAR2=, OUTDATA=);
    DATA EFF_MONO_IMP EFF_MONO_REST;
        SET &INDATA;
        IF TRT01PN IN (2,3) AND LASTVIS >=&ANVIS THEN OUTPUT
EFF_MONO_REST;
        ELSE OUTPUT EFF_MONO_IMP;
    RUN;

    PROC SORT DATA= EFF_MONO_IMP;
        BY _IMPUTATION_;
    RUN;

    PROC MI DATA=EFF_MONO_IMP OUT=EFF_REG_IMP NIMPUTE=1 SEED=&SEED
MINIMUM=0 MAXIMUM=60;
        VAR &VAR1 &VAR2;
        BY _IMPUTATION_;
        MONOTONE REGRESSION(&VAR2= &VAR1 / DETAILS);
    RUN;

    DATA &OUTDATA;
        SET EFF_MONO_REST EFF_REG_IMP;
    RUN;

    PROC DATASETS;
        DELETE EFF_MONO_IMP EFF_MONO_REST EFF_REG_IMP;
    RUN;
%MEND MI_IMP;
```

```
%MI_IMP(INDATA=EFF_MONO, ANVIS=2, SEED=445566, VAR1=SCORE_1 SCORE_2,
VAR2=SCORE_3, OUTDATA=EFF_IMP1);
%MI_IMP(INDATA=EFF_IMP1, ANVIS=3, SEED=445566, VAR1=SCORE_1 SCORE_2
SCORE_3, VAR2=SCORE_4, OUTDATA=EFF_IMP2);
%MI_IMP(INDATA=EFF_IMP2, ANVIS=4, SEED=445566, VAR1=SCORE_1 SCORE_2
SCORE_3 SCORE_4, VAR2=SCORE_5, OUTDATA=EFF_IMP3);
%MI_IMP(INDATA=EFF_IMP3, ANVIS=5, SEED=445566, VAR1=SCORE_1 SCORE_2
SCORE_3 SCORE_4 SCORE_5, VAR2=SCORE_6, OUTDATA=EFF_IMP4);
%MI_IMP(INDATA=EFF_IMP4, ANVIS=6, SEED=445566, VAR1=SCORE_1 SCORE_2
SCORE_3 SCORE_4 SCORE_5 SCORE_6, VAR2=SCORE_7, OUTDATA=EFF_IMP5);

DATA MI_EFF;
    SET EFF_IMP5;
    AVISITN=1;
    DIFF=SCORE_2-SCORE_1;
    OUTPUT;
    AVISITN=2;
    DIFF=SCORE_3-SCORE_1;
    OUTPUT;
    AVISITN=3;
    DIFF=SCORE_4-SCORE_1;
    OUTPUT;
    AVISITN=4;
    DIFF=SCORE_5-SCORE_1;
    OUTPUT;
    AVISITN=5;
    DIFF=SCORE_6-SCORE_1;
    OUTPUT;
    AVISITN=6;
    DIFF=SCORE_7-SCORE_1;
    OUTPUT;
RUN;

ods listing close;
PROC MIXED;
    CLASS USUBJID TRT01PN AVISITN SITEGR1N;
    MODEL      DIFF=SCORE_1      TRT01PN      AVISITN      SITEGR1N
TR01PN*AVISITN/DDFM=KR SOLUTION CL;
    REPEATED AVISITN /SUBJECT=USUBJID TYPE=UN;
```

```
    LSMEANS TRT01PN*AVISITN / DIFF=CONTROL("1", "6")diff=control("1"
"6") CL;
    BY _imputation_;
    RUN;

PROC MIANALYZE;
    CLASS TRT01PN;
    BY _TRT01PN;
    MODELEFFECTS TRT01PN;
run;

proc mianalyze parms=ls_mean;
    CLASS TRT01PN;
      BY AVISITN;
    MODELEFFECTS TRT01PN;
RUN;
```

4. MADRS 50% Responders Logistic Model

Note: Placebo will be reference
Note: All SM-13496 groups will be included in a model

```
PROC LOGISTIC DESCENDING;
      CLASS TRTPN POOLSITE / PARAM=REF;
      MODEL RESP50 = TRTPN POOLSITE BASE;
      CONTRAST 'LUR 20-60 vs. PLACEBO' TRTPN 1 0/ESTIMATE=EXP;
      CONTRAST 'LUR 80-120 vs. PLACEBO' TRTPN 0 1/ESTIMATE=EXP;
RUN;
```

5. Ranked ANCOVA (Selected Safety Parameters)

```
PROC RANK OUT= OUT1 NPLUS1 TIES = MEAN;
      VAR VSCFB VSBASE;
      RANKS RKVSCFB RKVSBASE;
RUN;

PROC GLM DATA = OUT1;
      MODEL RKVSCFB = RKVSBASE;
      OUTPUT OUT=RESIDUAL R=RESID;
RUN;

PROC FREQ DATA = RESIDUAL;
      TABLES TRTPN*RESID / NOPRINT CMH2;
RUN;
```

6. NNT and 95% CI

```
pi0a: responder rate of SM-13496 group
pi1: responder rate of placebo group
delta_ua : upper limit of NNT before taking the reciprocal
delta_la : lower limit of NNT before taking the reciprocal

nnt = round(1/(pi0a - pi1), 0.01);
delta_la = pi0a - pi1 - 1.96*sqrt(pi1*(1-pi1)/PCOUNT + pi0a*(1-
pi0a)/COUNT);
delta_ua = pi0a - pi1 + 1.96*sqrt(pi1*(1-pi1)/PCOUNT + pi0a*(1-
pi0a)/COUNT);

if pi0a>=pi1 then do;
   lower_ba = min(delta_ua, 1);
   upper_ba = max(delta_la, 0.000001);
   end;
else if pi0a<pi1 then do;
   lower_ba = min(delta_ua, -0.000001);
   upper_ba = max(delta_la, -1);
   end;

lcia = floor(round(1/lower_ba));
ucia = ceil(round(1/upper_ba));


* CA test */
Placebo group = 0
SM-13496 group = 1

ODS OUTPUT TrendTest=__TREND;
proc freq data=_madrs;
  table TRTPN *resp50/cmh trend;
run;

/* Between group effect size */
data bes01 ;
      pi0a = COUNT/COUNT ; ***a = treatment 1***;
      pi1 = _PCOUNT/PCOUNT ;    ***placebo***;
      rslt0a=2*(arsin(sqrt(pi0a)));
      rslt1 =2*(arsin(sqrt(pi1)));
      h = abs(rslt0a - rslt1) ;
      ch = put ( h , 5.2 );
run ;
```

7. Hochberg procedure

```
data a;
   input Test$ Raw_P @@;
   datalines;
test01  0.28282    test02  0.30688    test03  0.71022
```

```
;
proc multtest inpvalues=a hoc;
run;
```

APPENDIX 4. MEDICAL REVIEW OF TERMS

EPS Terms	Metabolic Terms	Combined Terms	Treatment Emergent Mania Terms
Muscle rigidity	Blood cholesterol increased	Parkinsonism : Muscle rigidity, Bradykinesia, Parkinsonism, Tremor, Psychomotor retardation	Bipolar I disorder
Trismus	Blood glucose increased		Hypomania
Bradykinesia	Blood insulin increased		Mania
Dystonia	Blood triglycerides increased		
Parkinsonism	Fructosamine increased		
Tremor	Glucose urine present	Dystonia: Dystonia, Trismus	
Psychomotor retardation	High density lipoprotein decreased		
	Low density lipoprotein increased		
	Weight increased		
	Glucose tolerance impaired	Somnolence: Sedation , Somnolence	
	Dyslipidaemia		
	Hyperinsulinaemia		
	Hyperlipidaemia		