

Sequential Medicine, Ltd

SM-A-05

A Randomized, Double-Blind, Single-Dose, 4-Way Crossover Study to Assess the Efficacy and Safety of SM-1 (50-mg Diphenhydramine, 5-mg Delayed-Release Zolpidem, and 0.5-mg Delayed-Release Lorazepam) versus 2 Comparators (50-mg Diphenhydramine and 5-mg Delayed-Release Zolpidem; 50-mg Diphenhydramine and 0.5-mg Delayed-Release Lorazepam) and Placebo in a 5-Hour Phase Advance Model of Transient Insomnia

09 FEB 2018

Statistical Analysis Plan
Version 1.0

Prepared by:

PPD

*7551 Metro Center Drive, Suite 300
Austin, TX 78744*

Prepared by: _____

James Pottala
Lead Biostatistician, Biostatistics
PPD

Reviewed by: _____

Raymond Claus
Senior Reviewer, Biostatistics
PPD

Upon review of this document, including table, listing, and figure shells, the undersigned approves the statistical analysis plan. The analysis methods and data presentation are acceptable, and the table, listing, and figure production can begin.

Approved by: _____


Thomas Dahl
Sequential Medicine, Ltd

Date: 12/FEB/2018

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	4
1. INTRODUCTION	5
2. OBJECTIVES	5
2.1. PRIMARY OBJECTIVES	5
2.2. KEY SECONDARY OBJECTIVES	5
2.3. OTHER SECONDARY OBJECTIVES	5
2.4. SAFETY OBJECTIVE	6
2.5. OTHER OBJECTIVES	6
3. INVESTIGATIONAL PLAN	6
3.1. OVERALL STUDY DESIGN AND PLAN	6
3.2. STUDY ENDPOINTS	7
3.2.1. <i>Primary Efficacy Endpoints</i>	7
3.2.2. <i>Key Secondary Efficacy Endpoints</i>	7
3.2.3. <i>Other Secondary Efficacy Endpoints</i>	7
3.2.4. <i>Safety and Other Endpoints</i>	8
4. GENERAL STATISTICAL CONSIDERATIONS	8
4.1. SAMPLE SIZE	9
4.2. RANDOMIZATION, STRATIFICATION, AND BLINDING	9
4.3. ANALYSIS SET	10
5. SUBJECT DISPOSITION AND PROTOCOL DEVIATIONS	11
5.1. DISPOSITION	11
5.2. PROTOCOL DEVIATIONS	11
6. DEMOGRAPHICS AND BASELINE CHARACTERISTICS	11
6.1. DEMOGRAPHICS	11
6.2. ALCOHOL, TOBACCO, AND CAFFEINE USAGE	12
6.3. MEDICAL HISTORY	12
6.3.1. <i>General Medical History</i>	12
6.3.2. <i>Disease-Specific History</i>	12
6.4. INCLUSION AND EXCLUSION CRITERIA	12
7. TREATMENTS AND THERAPY	13
7.1. PRIOR AND CONCOMITANT THERAPY	13
7.2. STUDY TREATMENTS	13
7.2.1. <i>Treatment Compliance</i>	13
8. EFFICACY ANALYSIS	14
8.1. PRIMARY EFFICACY ENDPOINT	14
8.1.1. <i>Primary Analysis</i>	14
8.1.2. <i>Assumption Testing</i>	15
8.2. KEY SECONDARY EFFICACY ENDPOINTS	15

8.3.	OTHER SECONDARY EFFICACY ENDPOINTS	17
8.4.	OTHER ENDPOINTS	17
8.5.	FAMILY WISE ERROR RATE (FWER) CONTROL	18
9.	SAFETY ANALYSIS	18
9.1.	ADVERSE EVENTS.....	18
9.1.1.	<i>Incidence of Adverse Events</i>	19
9.1.2.	<i>Relationship of Adverse Events to Study Drug</i>	19
9.1.3.	<i>Severity of Adverse Event</i>	20
9.1.4.	<i>Serious Adverse Events</i>	20
9.1.5.	<i>Adverse Events Leading to Study Discontinuation</i>	21
9.1.6.	<i>Death</i>	21
9.2.	CLINICAL LABORATORY EVALUATIONS	21
9.2.1.	<i>Hematology</i>	21
9.2.2.	<i>Chemistry</i>	21
9.2.3.	<i>Urinalysis</i>	22
9.3.	VITAL SIGN MEASUREMENTS	22
9.4.	PHYSICAL EXAMINATION	22
9.5.	OTHER SAFETY DATA.....	22
10.	INTERIM ANALYSIS	23
11.	CHANGES IN PLANNED ANALYSIS.....	24
12.	REFERENCES.....	25
13.	APPENDICES.....	26
13.1.	SCHEDULE OF EVENTS.....	26
13.2.	SLEEP STUDY EVENTS BY DAY FOR TREATMENT PERIODS	29
13.3.	IMPUTATION FOR PARTIAL AND MISSING DATES	31
13.4.	SAMPLE SIZE RECALCULATION REPORT.....	33

List of Abbreviations

AE	adverse event
ANOVA	analysis of variance
BMI	body mass index
D+L	50-mg diphenhydramine and 0.5-mg delayed-release lorazepam
D+Z	50-mg diphenhydramine and 5-mg delayed-release zolpidem
DSST	Digit Symbol Substitution Test
eCRF	electronic case report form
ESS	Epworth Sleepiness Scale
FAS	full analysis set
FWER	family-wise error rate
IP	Investigational Product
IRB	institutional review board
KSS	Karolinska Sleepiness Scale
LPS	latency to persistent sleep
LSMEANS	least squares means
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	modified full analysis set
MSE	mean square error
NAW	number of awakenings
OSA	obstructive sleep apnea
PPS	per-protocol set
PSG	polysomnography, polysomnographic
PSQ	Post-Sleep Questionnaire
PT	preferred term
REM	rapid eye movement
SAE	serious adverse event
SM-1	50-mg diphenhydramine, 5-mg delayed-release zolpidem, and 0.5-mg delayed-release lorazepam
sNAW	subjective number of awakenings
SOC	system organ class
sQUALITY	subjective quality
sSOL	subjective sleep onset latency
sTST	subjective total sleep time
TEAE	treatment-emergent adverse event
TST	total sleep time
WASO	wakefulness after sleep onset

1. Introduction

SM-1 is a 3-component investigational drug that is intended to increase total sleep time. The formulation provides 50-mg immediate-release diphenhydramine, 5-mg delayed-release zolpidem, and 0.5 mg delayed-release lorazepam. This combination is a unique approach to addressing the problem of transient insomnia.

A Phase 2 clinical study of SM-1 was conducted in the target population. The primary endpoint was the comparison of total sleep time (TST) between SM-1 and placebo. This study showed that SM-1 had a significantly greater TST than placebo ($P<0.001$) as well as the 2-drug combination (zolpidem and lorazepam, $P=0.014$). SM-1 also reduced wakefulness after sleep onset (WASO) significantly compared with placebo ($P<0.001$). On waking, there were no residual drug effects as indicated by the Karolinska Sleepiness Scale (KSS), Digit Symbol Substitution Test (DSST), or predischARGE neurological examination results.

The current study will serve to confirm the efficacy, safety, and tolerability of SM-1 in subjects with transient insomnia. In addition, the inclusion of additional drug combination arms will serve to generate data to assess the contribution of zolpidem and lorazepam to the overall efficacy of the investigational drug.

2. Objectives

2.1. Primary Objectives

The primary objectives of this study are:

- To assess efficacy of SM-1 versus placebo as measured by polysomnographic (PSG)-defined TST
- To assess efficacy of SM-1 versus 2 additional combination products comprised of 2 of the 3 components of SM-1 (diphenhydramine plus zolpidem; and diphenhydramine plus lorazepam) as measured by PSG-defined TST

2.2. Key Secondary Objectives

The secondary objectives of this study are:

- To assess activity of SM-1 versus placebo on additional PSG measures of sleep induction and maintenance
- To assess SM-1 versus placebo on subject-reported efficacy measures from the Post-Sleep Questionnaire (PSQ)

2.3. Other Secondary Objectives

Other efficacy objectives include comparisons with the two 2-drug combination products.

- To assess activity of SM-1 versus the two 2-drug combination products on additional PSG measures of sleep induction and maintenance
- To assess SM-1 versus the two 2-drug combination products on subject reported efficacy measures from the PSQ.
- To assess the activity of the two 2-drug combination products versus placebo on PSG-defined TST, and additional PSG measures of sleep induction and maintenance

2.4. Safety Objective

The safety objective of this study is to assess safety of SM-1 in terms of adverse events (AEs) and morning measures of alertness (KSS and DSST).

2.5. Other Objectives

Other objectives include the assessment of the effect of SM-1 and the two 2-drug combination products on sleep stage distribution.

3. Investigational Plan

3.1. Overall Study Design and Plan

This is a randomized, double-blind, single dose, 4-way crossover study to assess the efficacy and safety of SM-1 compared with placebo and two 2-drug combination products in subjects with a history of transient insomnia. Subjects will be exposed to a 5-hour phase advance model of transient insomnia with 8 hours (960 30-second epochs) of PSG recording as the primary efficacy assay. The primary efficacy endpoint is TST on an 8-hour PSG comparing SM-1 versus placebo, SM-1 versus D+Z, and SM-1 versus D+L (see Table 3-1 footnote for abbreviations).

The study has the following periods: screening (at least 8 and no more than 21 days), treatment and PSG period, and a follow-up safety phone call. There will be 84 subjects enrolled in 2 clinical centers in the study. The duration will be approximately 8 weeks, which includes 5 visits (screening period and 4 treatment periods) and 1-week (minimum 5-day) washout intervals between each treatment period.

After signing the informed consent, potential subjects will come in for screening visit (Visit 1) and specified procedures will be performed as in the schedule of events ([Appendix 13.1](#)). The subjects who satisfy the eligibility criteria will be issued a paper diary to record the time they go to bed and the time they get up during the screening period. Information recorded will be given to study personnel to calculate median habitual bedtime for each subject. At Visit 2, subjects will check-in at the study center 7 hours earlier than their median habitual bedtime. Once eligibility has been reconfirmed, subjects will be randomly assigned to 1 of 4 treatment sequences as shown in Table 3-1, and procedures will be performed per the schedule of events and according to the time of sleep study events relative to each subject's habitual bedtime ([Appendix 13.2](#)). Each treatment and PSG period visit will last approximately 11 to 12 hours.

Table 3-1 Randomized Treatment Sequences

	Treatment Period 1 Visit 2	Treatment Period 2 Visit 3	Treatment Period 3 Visit 4	Treatment Period 4 Visit 5
Sequence 1	SM-1	D+Z	D+L	Placebo
Sequence 2	D+Z	D+L	Placebo	SM-1
Sequence 3	D+L	Placebo	SM-1	D+Z
Sequence 4	Placebo	SM-1	D+Z	D+L

SM-1 = 50-mg diphenhydramine, 5-mg delayed-release zolpidem, and 0.5-mg delayed-release lorazepam;

D+Z = 50-mg diphenhydramine and 5-mg delayed-release zolpidem;

D+L = 50-mg diphenhydramine and 0.5-mg delayed-release lorazepam;
Placebo = identical in appearance to SM-1, D+Z, and D+L and has the same excipients, but no diphenhydramine, zolpidem, lorazepam, or delayed-release coating materials

Subjects will go to bed (“lights out”) 5 hours \pm 30 minutes before their median habitual bedtime. Approximately 90 minutes before lights out, baseline DSST measurements will be obtained. Approximately 60 minutes before lights out, PSG electrodes will be applied. Subjects will be administered their assigned treatment by study personnel 30 minutes before lights out and an oral cavity check will be performed to assure compliance with treatment. At the assigned lights out time, the subject will go to bed, PSG biocalibration will be performed, and PSG recording will begin. The recording will continue for 8 hours.

At the completion of the 8-hour PSG recording period, subjects will be awakened. Approximately 30 minutes after the end of PSG recording, subjects will complete the PSQ, KSS, and DSST. Before leaving the study center, subjects will undergo a brief discharge evaluation consisting of tandem gait, the Romberg test, and an assessment of vital signs and AEs. Subjects will remain at the study center until they are able to pass the discharge evaluation. Upon discharge, subjects will be told the time and date to return for the next overnight visit, dispensed a daily washout interval diary, and instructed to maintain normal sleep patterns.

Following a minimum 5 days washout, subjects will return to the study center and return their diary for review. The schedule of procedures for Visits 3, 4, and 5 will be the same as for Visit 2 except that the subjects will be given the next treatment in their randomization sequence (**Table 3-1**) and lights out will occur at the time established at Visit 2 \pm 15 minutes.

Within 7 days after completion of Treatment Period 4 (Visit 5), but at least 72 hours after the final dose, a follow-up safety phone call will be performed. Outcomes of any AEs will be discussed and recorded. If indicated, a visit to the study center for safety assessments will be scheduled.

3.2. Study Endpoints

3.2.1. Primary Efficacy Endpoints

The primary efficacy endpoints for this study are TST on an 8-hour PSG comparing SM-1 versus placebo, SM-1 versus D+Z, and SM-1 versus D+L.

3.2.2. Key Secondary Efficacy Endpoints

The comparison of SM-1 versus placebo for data collected from PSG recordings and PSQ responses will be performed on WASO, latency to persistent sleep (LPS), number of awakenings (NAW), subjective TST (sTST), subjective sleep onset latency (sSOL), and TST by quarters of the night.

3.2.3. Other Secondary Efficacy Endpoints

The comparison of SM-1 versus D+Z and SM-1 versus D+L for data collected from PSG recordings and PSQ responses will be performed on WASO, latency to persistent sleep (LPS), number of awakenings (NAW), subjective TST (sTST), subjective sleep onset latency (sSOL), and TST by quarters of the night.

The comparison of D+Z versus placebo and D+L versus placebo for data collected from PSG recordings and PSQ responses will be performed on TST, WASO, LPS, NAW, sTST, sSOL, and TST by quarters of the night.

3.2.4. Safety and Other Endpoints

Safety endpoints include the following:

- Incidence, severity, and relationship to treatment of AEs
- Karolinska Sleepiness Scale
- Digit Symbol Substitution Test
- Discharge evaluation (tandem gait and Romberg)

Other endpoints include percentage and number of minutes of sleep stages N1, N2, N3, and rapid eye movement (REM) and latency to REM sleep onset.

4. General Statistical Considerations

Continuous variables will be summarized using the mean, SD, median, minimum value, and maximum value. Categorical variables will be summarized using frequency counts and percentages. Data will be provided in by-subject data listings with center ID concatenated with subject ID.

Unless specified otherwise, all statistical analyses will be performed using a 2-sided hypothesis test at the overall 5% level of significance leading to 95% (2-sided) CIs. All p-values will be rounded to 4 decimal places. If a p-value is less than 0.0001, it will be reported as “<0.0001.” If a p-value is greater than 0.9999, it will be reported as “>0.9999.”

Data will be displayed in all listings sorted as indicated in the specifications. When count data are presented, the percent will be suppressed when the count is zero in order to draw attention to the non-zero counts. A row denoted “Missing” will be included in count tabulations where necessary to account for dropouts and missing values. The denominator for all percentages will be the number of subjects in that treatment with nonmissing values within the analysis set of interest, unless otherwise stated. Non-zero percentages will be rounded to 1 decimal place, except 100% will be displayed without any decimal places. Additional rounding rules are as follows:

- If the original value has 0 decimal places: mean, median, and CIs will have 1 decimal place and SD will have 2 decimal places.
- If the original value has 1 decimal place: mean, median, and CIs will have 2 decimal places and SD will have 3 decimal places.
- If the original value has 2 or more decimal places: mean, median, CIs, and SD will all have 3 decimal places.

Minimum and maximum will always have the same decimal places as the original measure, up to a maximum of 3 decimal places.

Study Day 1 is defined as date of first dose (i.e., date of study drug administration during Visit 2 = study Day 1); study day will be calculated as below:

- Study Day xx = date of measurement/observation – first dose date + 1; if date of measurement/observation is on or after first dose date.
- Study Day xx = date of measurement/observation – first dose date; if date of measurement/observation is prior to the first dose date.

The duration of study participation will be summarized for all subjects. If the date of the last visit on the Study Completion/Early Withdrawal page is missing, or if a subject is lost to follow-up, the latest available visit date will be used.

4.1. Sample Size

A total of 68 blinded subjects will be able to ensure 80% power to detect treatment differences on TST between SM-1 and all 3 controls (D+Z, D+L, and placebo) at a 2-sided 0.05 significance level with a fixed-sequence testing procedure using an effect size of at least 44 minutes between SM-1 and all three comparators.

The assumptions in the sample size analysis included a 2x2 crossover design (since the 4x4 crossover design collapses to a 2x2 when making a comparison between 2 treatment groups), minimum of 80% power, 2-sided alpha=0.05, standard deviation (defined as the square root of the within mean square error (MSE) from a repeated measures analysis of variance (ANOVA) model) of 88 minutes, and effect sizes of 120 minutes for SM-1 versus placebo and 44 minutes for SM-1 versus the 2-drug comparators with 0% loss to follow-up. To maintain the family-wise error rate (FWER) alpha=0.05, the 3 primary efficacy hypotheses will be tested using a fixed-sequence procedure; more details are provided in **Section 8.1.1**.

There is a sample size recalculation planned after approximately 50% of the subjects have completed the study. The details are provided in **Section 10**, and the sample size recalculation report is included in **Appendix 13.4**. The sample size recalculation allows increasing the sample size if the variability or loss to follow-up is greater than expected.

4.2. Randomization, Stratification, and Blinding

The study will be performed in a double-blind manner, with the subjects, investigator, and study center staff being blinded to the identity of study drug. All study drug will be supplied in identical packaging and will be identical in color, smell, taste, and appearance to enable double-blind conditions. Subjects will be randomly assigned to 1 of 4 treatment sequences shown in **Table 3-1** by a manual randomization process. When a center needs to randomly assign a subject, the lowest unused randomization number at that center will be selected from the randomization schedule. The kit labeled with this randomization number will include 4 bottles each labeled with the randomization number and a visit number (i.e., Visit 2 through Visit 5) to be dispensed at the corresponding visits. The randomization schedule will use a block size, which will not be revealed to the staff responsible for the conduct of the study. No stratification will be used in this study.

A subject will not be unblinded until the end of the study after database lock or if a medical treatment depends on knowing the study drug the subject received. If the blind needs to be broken because of a medical emergency, the investigator may unblind a subject's treatment allocation via a set of blind breaking envelopes located at each center. The investigator should make every effort to contact the medical monitor and/or sponsor prior to opening the envelope corresponding to the subject and visit, to reveal the actual treatment drug description. If a subject's treatment is unblinded, the sponsor must be notified immediately. Any subjects unblinded during the study will be summarized in a table and presented in a by-subject listing.

4.3. Analysis Set

The following analysis sets will be used in the statistical analyses:

All enrolled subjects: A subject is considered enrolled when he/she signs informed consent, passes the screening criteria, and is eligible for first dose of drug to be administered.

Full analysis set (FAS): The FAS will consist of all subjects who were randomly assigned to study drug at Visit 2 and have taken any study drug during the treatment period. All analyses using the FAS will group subjects according to randomized treatment. The FAS will be used as a *sensitivity* analysis for the primary, secondary, and other efficacy analyses.

Modified FAS (mFAS): The mFAS will consist of all subjects from the FAS that did not have their treatment codes potentially unblinded during the study. All analyses using the mFAS will group subjects according to randomized treatment. The mFAS will be used for the primary, secondary, and other efficacy analyses and will be the principal efficacy population.

Per-protocol set (PPS): The PPS will consist of all FAS subjects who fulfill all eligibility criteria, have taken all 4 doses of study drug, have not taken any prohibited medication, and have no significant protocol deviations that may affect the primary endpoint. Significant protocol deviations and whether each results in excluding a subject from the PPS will be defined in a Study Deviation Rules Document and will be finalized prior to database lock and study unblinding. All analyses using the PPS will group subjects according to actual treatment received. The PPS will only be used as a *sensitivity* analysis for the primary efficacy analysis.

Safety set: The safety set will consist of all subjects who received any study drug. All analyses using the safety set will group subjects according to actual treatment received. Each AE will be assigned to the actual treatment that was associated with the AE based on whether the start date of the AE was on or after the start date of the associated period, up through the day prior to the start date of the following period. The safety set will be used for all safety analyses.

A by-subject listing will be provided for any randomization errors (i.e., a subject was randomized to the incorrect sequence) or dispensing errors where a subject was randomized to the correct sequence but the treatments were administered in an incorrect order.

5. Subject Disposition and Protocol Deviations

5.1. Disposition

A summary of all subjects screened (i.e., signed informed consent) will include the number and percentage of subjects screened, randomized, and the reasons for screen failures by center as the following: inclusion/exclusion criteria, subject withdrawn, lost to (screening) follow-up, sponsor decision, AE, and other. A listing of screen-failed subjects will be provided as well.

The primary reason for study discontinuation will also be listed by subject, and summarized in a table for the mFAS and FAS populations by treatment sequence group and overall. The reason for study discontinuation may include any of the following: completed study, AE, death, lost to follow-up, physician decision, pregnancy, protocol violation, study terminated by sponsor, subject withdrawal, or other. Only 1 reason for study discontinuation per subject will be entered. The number of subjects in the mFAS, PPS, and SS populations and subjects completing all 4 treatment visits will be included.

5.2. Protocol Deviations

A deviation from the protocol is an unintended or unanticipated departure from the procedures or processes approved by the sponsor and the institutional review board (IRB)/independent ethics committee and agreed to by the investigator. The investigator may implement a deviation from the protocol to eliminate an immediate hazard to study subjects without prior IRB approval. A significant deviation occurs when there is nonadherence to the protocol by the patient or investigator that results in a significant, additional risk to the subject. Descriptions and categories of significant protocol deviations will be defined in the Study Protocol Deviation Document and finalized prior to database lock and study unblinding. The significant protocol deviations will be listed by subject and summarized in a table by categories for the mFAS and FAS by treatment sequence group and overall.

6. Demographics and Baseline Characteristics

6.1. Demographics

Baseline demographics will be listed by subject and summarized by treatment sequence group and overall for subjects in the mFAS and FAS. Demographic variables will include the following:

- Age at informed consent defined as the integer part of: $[(\text{date of informed consent} - \text{date of birth}) / 365.25]$
- Age group (greater than median vs less than median)
- Sex
- If female: fertility status (i.e., postmenopausal, potentially able to bear children, or sterile)
- Race (i.e., white, African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or other)
- Ethnicity (i.e., Hispanic or Latino, not Hispanic or Latino)
- Weight (kg)
- Height (cm)
- Body mass index (BMI) defined as: $\text{weight (kg)} / [\text{height (m)}]^2$

Continuous descriptive statistics will be presented for age, weight, height, and BMI. Frequency counts and percentages will be presented for age group, sex, fertility status, race, and ethnicity.

6.2. Alcohol, Tobacco, and Caffeine Usage

History of substance use for alcohol, caffeine, and tobacco/nicotine (type and amount) will be listed by subject. Alcohol, caffeine, and tobacco use history will all be categorized as current, former, or never. For alcohol, the average number of units per week will be recorded. A unit of alcohol is equivalent to 12 ounces of beer, 4 ounces of wine, or 1 ounce of liquor. The amount of caffeine will be recorded as mg/day. The average use of tobacco is measured in units per day; and if the subject is a current smoker, then whether they smoke during the sleep period will be entered in the electronic case report form (eCRF).

6.3. Medical History

6.3.1. General Medical History

Medical history and psychiatric history will be listed by subject and summarized for the FAS by treatment sequence group and overall. The verbatim terms will be coded by the primary system organ class (SOC) and preferred term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 or higher, and the summary will include the count and percentage of subjects with a significant history by SOC and PT. The listing will include start date, end date, and ongoing status.

6.3.2. Disease-Specific History

Subject sleep history will be assessed at screening to confirm subject eligibility for the study using the STOP-Bang (STOP: snoring, tiredness, observed apnea, and high blood pressure; Bang: BMI, age, neck circumference, and gender) questionnaire [**STOP-Bang 2012**]. The questionnaire consists of 8 dichotomous (yes/no) items related to the clinical features of obstructive sleep apnea (OSA). The total score ranges from 0 to 8 (based on the number of “yes” responses). The STOP-Bang responses will be listed by subject.

The Epworth Sleepiness Scale (ESS) [**Johns 1997**] will also be used at screening to measure a subject’s level of daytime sleepiness or their average sleep propensity in daily life. Subjects rate, on a 4-point scale (0 to 3), their usual chances of dozing off or falling asleep in 8 different situations or activities that most people engage in as part of their daily lives. The total ESS score is the sum of 8 item scores and can range between 0 and 24. The higher the score, the higher the person’s level of daytime sleepiness. If a subject fails to answer any 1 of the 8 items, then the total score will be missing. Responses to each question and the score totals will be listed by subject.

6.4. Inclusion and Exclusion Criteria

The investigator will certify that the subject satisfies all eligibility criteria at screening and continues to satisfy all inclusion and exclusion criteria prior to dosing at Visit 2. A subject is considered enrolled when he/she signs informed consent, passes the screening criteria, and is eligible for first dose of drug to be administered. Inclusion and exclusion criteria that are violated will be presented in a by-subject listing.

7. Treatments and Therapy

7.1. Prior and Concomitant Therapy

The use of concomitant medications during this study is discouraged, unless required to treat AEs. The use of other concomitant medications should be approved by the investigator and sponsor (or designee) before subjects enroll in the study.

Any prior therapy within the 30 days before enrollment (i.e., Visit 2) or concomitant therapy taken by a subject during the course of the study will be documented, along with the reason for its use. Concomitant medications will include all prescription drugs, over-the-counter medications, herbal products, vitamins, minerals, and nutritional supplements. WHO Drug Dictionary version 01JUN2017 will be used to code all concomitant medications, which will be summarized in a table for PT by treatment group for the safety set. Partial dates for concomitant medication will be imputed as described in **Appendix 13.3**. A concomitant medication will be assigned to the treatment in 1 or more periods if the start date of that medication is on or after the start date of that period, up through the day prior to the start date of the following period (or 7 days after last dose, for the last period). Prior and concomitant medication data will be listed by subject for the safety set.

7.2. Study Treatments

The study includes the following 4 treatment arms that require each subject to take 1 pill orally from each of the arms during the 4 treatment periods, in a random sequence:

- SM-1 contains 50-mg diphenhydramine, 5-mg delayed-release zolpidem, and 0.5-mg delayed-release lorazepam
- D+Z contains 50-mg diphenhydramine and 5-mg delayed-release zolpidem
- D+L contains 50-mg diphenhydramine and 0.5-mg delayed-release lorazepam
- Placebo

7.2.1. Treatment Compliance

At Screening Visit 1, subjects who satisfy all eligibility criteria will be issued a pretreatment paper diary and instructed to record the time they went to bed with the intention of sleeping and the time they got up, during the screening period that lasts from 8 to 21 days. Each diary entry day will be recorded as collected, not expected, or missing data. The compliance rate will be calculated as the number of collected days divided by the number of days of collected + missing data multiplied by 100%. There must be at least 5 entries and at least a 70% compliance rate at Visit 2 for a subject to be eligible for randomization. The median habitual bedtime will be determined based on these pretreatment diary entries. The sleep diary is reissued between treatment visits.

Investigational product (IP) that is considered damaged will not be replaced and the data point for that particular visit will be missed for the subject. When treatment is administered during Visits 2, 3, 4, and 5, an oral cavity check will be performed to assure compliance with treatment. However, if the pill is vomited then the subject will not be given a replacement pill, and the subject will be required to take the next assigned treatment at the following visit. The PSG will not be conducted for visits in which the subject did not take IP. A minimum washout period of 5 days is required between treatment visits.

A table will summarize treatment compliance and washout period durations between visits by treatment group sequence and overall for the FAS; in addition, the diary compliance number of days collected, not expected, and missing with compliance rate will be reported. Subject listings will be provided for sleep diary data and for treatment administrations.

8. Efficacy Analysis

All efficacy summaries will present data by treatment group (i.e., SM-1, D+Z, D+L, and placebo) for the mFAS. The PPS will only be used for sensitivity analyses of the primary efficacy endpoint. The FAS will be used as a sensitivity analysis for the primary and secondary efficacy analyses. Subject level data will be provided in listings.

8.1. Primary Efficacy Endpoint

The primary endpoint is TST that will be determined at Visits 2, 3, 4, and 5 as the number of minutes the subject was in REM sleep + non-REM sleep during the 8-hour PSG recording, which will be scored at a central location blinded to treatment assignment. The analysis of the primary efficacy endpoint will be conducted using a repeated measures ANOVA model. The primary comparison will be SM-1 with placebo, followed by SM-1 with D+Z and SM-1 with D+L. Since missing data are expected to be minimal, no formal imputation methods will be used. Instead, missing data will be assumed to be missing at random, as handled by the Restricted Maximum Likelihood (REML) method in PROC MIXED.

8.1.1. Primary Analysis

The primary efficacy analysis will be conducted on the mFAS using a repeated measures ANOVA model with TST as the response variable. PROC MIXED in SAS[®] version 9.3 or higher will be used for the repeated measures ANOVA analysis and will include treatment, sequence, center, and period as fixed effects [Yarandi 2004]. The Kenward-Roger degrees of freedom approximation method will be used in the MODEL statement, which is less biased for small samples [Kenward and Roger 1997]. Restricted Maximum Likelihood method and the RANDOM intercept statement will be used to model the within-subject correlation as subjects nested in sequence.

There are 3 two-sided hypotheses in this primary efficacy analysis:

- Null hypothesis: $TST_{SM-1} - TST_{Placebo} = 0$; alternative hypothesis: $TST_{SM-1} - TST_{Placebo} \neq 0$
- Null hypothesis: $TST_{SM-1} - TST_{D+Z} = 0$; alternative hypothesis: $TST_{SM-1} - TST_{D+Z} \neq 0$
- Null hypothesis: $TST_{SM-1} - TST_{D+L} = 0$; alternative hypothesis: $TST_{SM-1} - TST_{D+L} \neq 0$

To control the FWER, all above hypotheses will be tested using the fixed-sequence procedure with two-sided $\alpha = 0.05$ (the details of the fixed-sequence procedure are provided in [Section 8.5](#)) [Dmitrienko and D'Agostino 2013]. The p-values from the ESTIMATE statement used to test each hypothesis and the 95% CIs will be reported; however, significant p-values will be indicated with an asterisk. The least squares means (LSMEANS) statement will be used to tabulate the mean estimate of TST and standard error for each treatment group.

The following sensitivity and subgroup analyses will be performed:

- Center will be removed from the model effectively pooling the centers

- Interaction between treatment and period will be added to the model to test 1st order carryover effects
- Use of PPS
- Subgroup by sex (male vs female)
- Subgroup by age group (median split)

8.1.2. Assumption Testing

The ANOVA model assumes residuals are normally distributed with constant variance. The univariate distribution of all the TST measurements will be tested for skewness. If skewness is greater than 2, then the responses will benefit from natural logarithm transformation; sSOL and LPS are expected to be logarithm transformed. For repeated measures ANOVA, the variance assumption is compound symmetry, or less restrictively that Huynh-Feldt condition of equal variances for all possible differences between repeated measures (i.e., sphericity) [Kuehl 2000]. The Huynh-Feldt condition will be tested using PROC GLM with the PRINTE option in the REPEATED statement. If sphericity is rejected, then these results should be interpreted with caution; the result from the sphericity test will be included in the table.

8.2. Key Secondary Efficacy Endpoints

The mFAS and FAS will be used for the key secondary efficacy endpoints, to compare SM-1 versus Placebo. The secondary endpoints are derived from the PSG recordings and PSQ assessments at Visits 2, 3, 4, and 5. The following PSG secondary endpoints will be included in the analyses: WASO, LPS, NAW, TST by quarters of the night; and PSQ secondary efficacy endpoints: sTST, subjective quality (sQUALITY), sSOL, and subjective number of awakenings (sNAW). Since the protocol was approved, sQUALITY and sNAW have been added as secondary efficacy endpoints.

The PSG is scored by dividing the 8-hour recordings into 960, 30-second epochs of various sleep stages. The following are derived from the PSG data transfer from the clinical sleep laboratory.

- WASO = total wake time in minutes from persistent sleep onset to lights on
- LPS = duration in minutes from lights off to the first epoch of 20 consecutive nonwake epochs
- NAW = number of awakening from the onset of persistent sleep until lights on. An awakening is defined as at least 2 consecutive epochs of wake. Individual awakenings must be separated by at least 1 epoch of stage N2, N3, or REM sleep
- TST1Q = duration in minutes of total REM sleep + non-REM sleep as Hour 1 + Hour 2
- TST2Q = duration in minutes of total REM sleep + non-REM sleep as Hour 3 + Hour 4
- TST3Q = duration in minutes of total REM sleep + non-REM sleep as Hour 5 + Hour 6
- TST4Q = duration in minutes of total REM sleep + non-REM sleep as Hour 7 + Hour 8

The PSQ is a self-assessment questionnaire that the centers enter into eCRF to assign the following secondary efficacy endpoints, which include the prefix “s” for being subjective:

- sTST = response to “How much time in total did you sleep last night?” converted to minutes
- sQUALITY = response to “How would you describe the quality of your sleep last night?”; 1=poor, 2=fair, 3=good, 4=excellent

- sSOL = response to “How long did it take you to fall asleep?” converted to minutes
- sNAW = response to “How many times did you wake up during the night?”

There are 11 key secondary endpoints including the 3-drug combination treatment (i.e., SM-1); however only 2 hypotheses will be tested using the fixed-sequence procedure to strongly control the FWER at 0.05. The difference in treatment groups with 95% CI and p-values will be reported for each endpoint. P-values that are significant based on the fixed-sequence procedure will be identified with an asterisk (see **Section 8.5**).

- Key Secondary Endpoints Controlled with FWER:
 - $WASO_{SM-1}$ versus $WASO_{Placebo}$
 - LPS_{SM-1} versus $LPS_{Placebo}$
- Key Secondary Endpoints Not Controlled with FWER:
 - $sTST_{SM-1}$ versus $sTST_{Placebo}$
 - $sQUALITY_{SM-1}$ versus $sQUALITY_{Placebo}$
 - $sSOL_{SM-1}$ versus $sSOL_{Placebo}$
 - NAW_{SM-1} versus $NAW_{Placebo}$
 - $sNAW_{SM-1}$ versus $sNAW_{Placebo}$
 - $TST1Q_{SM-1}$ versus $TST1Q_{Placebo}$
 - $TST2Q_{SM-1}$ versus $TST2Q_{Placebo}$
 - $TST3Q_{SM-1}$ versus $TST3Q_{Placebo}$
 - $TST4Q_{SM-1}$ versus $TST4Q_{Placebo}$

Each secondary endpoint will be tested for normality as described above (except NAW, sNAW, and sQUALITY) and the measurements transformed using natural logarithm as appropriate. If an endpoint includes a minimum measurement with zero minutes, then all the measurements for that endpoint will be increased by 1 minute prior to transformation. Sphericity will also be tested and reported for each of the secondary endpoints. These analyses will be performed using the same repeated measures ANOVA model as in the primary analysis, except for NAW, sNAW, and sQUALITY endpoints. Since missing data are expected to be minimal, no formal imputation methods will be used.

The NAW and sNAW endpoints will be modeled using PROC GENMOD with a Poisson distribution and a log link function in the MODEL statement. However, a zero-inflated Poisson (ZIP) distribution model will also be used to test if there is a preponderance of zero counts, i.e., more than expected under the Poisson distribution. The ZEROMODEL statement will be used with a logit link to test if the mixture model is required. If so, these results will be reported in addition to the standard Poisson model. The model goodness-of-fit will be reported as the deviance divided by the degrees of freedom, and if significant $p < 0.05$ overdispersion (i.e., the variance is significantly larger than the mean) exists then the results including deviance scaling will also be reported. The same fixed effects, LSMEANS and ESTIMATE statements included in the primary analysis will be used. The REPEATED statement will be used to induce a compound symmetry covariance structure.

The sQUALITY endpoint will be modeled using ordinal logistic regression in PROC GENMOD with a multinomial distribution and a cumulative logit link function. The proportional odds assumption will be tested and the p-value reported. The same fixed effects, LSMEANS and ESTIMATE statements included in the primary analysis will be used; however, odds ratio estimates with SE and 95% CI will be reported. The REPEATED statement will be used to induce a compound symmetry covariance structure.

8.3. Other Secondary Efficacy Endpoints

The mFAS and FAS will be used for the other secondary efficacy endpoints, to compare SM-1 versus D+Z and SM-1 versus D+L using the same methodology as described above for the following comparisons.

- Endpoint Comparisons:
 - $WASO_{SM-1}$ versus $WASO_{D+Z}$
 - $WASO_{SM-1}$ versus $WASO_{D+L}$
 - LPS_{SM-1} versus LPS_{D+Z}
 - LPS_{SM-1} versus LPS_{D+L}
 - $sTST_{SM-1}$ versus $sTST_{D+Z}$
 - $sTST_{SM-1}$ versus $sTST_{D+L}$
 - $sQUALITY_{SM-1}$ versus $sQUALITY_{D+Z}$
 - $sQUALITY_{SM-1}$ versus $sQUALITY_{D+L}$
 - $sSOL_{SM-1}$ versus $sSOL_{D+Z}$
 - $sSOL_{SM-1}$ versus $sSOL_{D+L}$
 - NAW_{SM-1} versus NAW_{D+Z}
 - NAW_{SM-1} versus NAW_{D+L}
 - $sNAW_{SM-1}$ versus $sNAW_{D+Z}$
 - $sNAW_{SM-1}$ versus $sNAW_{D+L}$
 - $TST1Q_{SM-1}$ versus $TST1Q_{D+Z}$
 - $TST1Q_{SM-1}$ versus $TST1Q_{D+L}$
 - $TST2Q_{SM-1}$ versus $TST2Q_{D+Z}$
 - $TST2Q_{SM-1}$ versus $TST2Q_{D+L}$
 - $TST3Q_{SM-1}$ versus $TST3Q_{D+Z}$
 - $TST3Q_{SM-1}$ versus $TST3Q_{D+L}$
 - $TST4Q_{SM-1}$ versus $TST4Q_{D+Z}$
 - $TST4Q_{SM-1}$ versus $TST4Q_{D+L}$

In addition, D+Z versus placebo and D+L versus placebo for data collected from PSG recordings and PSQ responses on TST, WASO, LPS, NAW, sTST, sSOL, and TST by quarters of the night will be analyzed. The ESTIMATE statement will be used to summarize a point estimate and 95% CI of the difference in treatment groups with p-values reported as described above.

8.4. Other Endpoints

The mFAS and FAS will be used for the other endpoints, which are derived from the PSG recordings at Visits 2, 3, 4, and 5. Other endpoints include percentage and duration in minutes of sleep stages N1, N2, N3, and REM and latency to REM sleep onset (defined as the number of non-REM sleep epochs from

lights off to the first epoch of REM sleep). These will be analyzed using the same repeated measures ANOVA models described above.

8.5. Family Wise Error Rate (FWER) Control

Only the 5 hypotheses listed in this section will be controlled at a FWER = 0.05 for this Phase III clinical trial. Only the 5 p-values from these hypotheses, when conducted in the mFAS population, have the possibility of being deemed statistically significant. All other p-values, e.g., in different populations, subgroups, sensitivity analyses, different endpoints, etc. will not be included in the FWER and thus any p-value <0.05 may have occurred by chance, since no multiple testing procedures will be applied to these other hypotheses.

The order of fixed-sequence testing is as follows:

- null hypothesis H1: $TST_{SM-1} - TST_{Placebo} = 0$;
 - alternative hypothesis K1: $TST_{SM-1} - TST_{Placebo} \neq 0$
- null hypothesis H2: $WASO_{SM-1} - WASO_{Placebo} = 0$
 - alternative hypothesis K2: $WASO_{SM-1} - WASO_{Placebo} \neq 0$
- null hypothesis H3: $TST_{SM-1} - TST_{D+Z} = 0$;
 - alternative hypothesis K3: $TST_{SM-1} - TST_{D+Z} \neq 0$
- null hypothesis H4: $TST_{SM-1} - TST_{D+L} = 0$;
 - alternative hypothesis K4: $TST_{SM-1} - TST_{D+L} \neq 0$
- null hypothesis H5: $LPS_{SM-1} - LPS_{Placebo} = 0$
 - alternative hypothesis K5: $LPS_{SM-1} - LPS_{Placebo} \neq 0$

Starting with H1, if the null hypothesis is rejected (i.e., p-value ≤ 0.05), then the next hypothesis in the sequence (H2) will be tested at the same two-sided alpha = 0.05. However, if H1 is accepted (i.e. p-value > 0.05), then all following null hypotheses are also accepted and testing stops. This procedure is repeated in order from H1 to H5, or until a single test fails to reject the null hypothesis.

Vertical box plot figures will be created for TST, WASO, and LPS. Each figure will include the raw data for the endpoint across the four treatment groups, i.e. SM-1, D+Z, D+L, and Placebo. The box plots will include details for mean, 25th percentile, median, 75th percentile, with whiskers up to 1.5 times the interquartile range with more extreme observations indicated as outliers.

9. Safety Analysis

Safety and tolerability will be assessed in terms of the incidence, severity, and relationship to treatment of AEs and by morning measures of alertness (KSS and DSST). In addition, safety assessments will include alcohol screening, urine pregnancy testing for women of childbearing potential, vital sign measurements, and pre-discharge evaluations. A safety follow-up phone call is performed as Visit 6, 3 to 7 days after the final dose. All safety analysis will be conducted on the safety set, and summaries will be presented by treatment group and overall.

9.1. Adverse Events

An AE is defined as any untoward medical occurrence in a subject enrolled into this study regardless of its causal relationship to study drug. A treatment-emergent AE (TEAE) is defined as any event not

present before exposure to study drug or any event already present that worsens in either intensity or frequency after exposure to study drug.

A treatment-emergent AE is defined as an AE that meets any of the following conditions:

- Begins on or after Study Day 1 up to 7 days after the last dose of study drug
- Begins before Study Day 1 and worsens in severity or frequency on or after Study Day 1 up to 7 days after the last dose of study drug

Adverse events with unknown onset dates or unknown end dates will be counted as having occurred during the investigational drug period (i.e., as TEAEs) unless the event resolves before Study Day 1. Partial dates will be imputed as given in **Appendix 13.3**. A TEAE will be assigned to the treatment in 1 period if the start date of that TEAE is on or after the start date of that period, up through the day prior to the start date of the following period (or 7 days after last dose, for the last period).

All AEs will be followed to adequate resolution. The MedDRA Version 20.0 or higher will be used to code all AEs. However, only TEAEs will be presented in AE tables, according to SOC and PT. Any AEs that occur prior to Study Day 1 without worsening in severity should be recorded on the medical history eCRF and will be summarized as part of medical history.

The overall summary of TEAEs by SOC and PT will be presented using counts and percentages for the following categories:

- Any TEAE
- Any study drug related TEAE
- Any severe TEAE
- Any serious TEAE
- Any study drug related serious TEAE
- Any TEAE leading to study discontinuation
- Any TEAE leading to death (listing only)

9.1.1. Incidence of Adverse Events

The total number of TEAEs will be summarized. Also the number of subjects with at least 1 TEAE will be grouped by SOC and PT. If a subject reports the same PT multiple times, then that PT will only be counted once per subject. As with the PT, if a subject reports multiple TEAEs within the same SOC, then that SOC will only be counted once per subject. For tables showing incidence by SOC and PT, PTs will be sorted within SOC in descending order of incidence in the total group. All AEs will be presented in a by-subject listing; an asterisk will be used to indicate any AE that is not a TEAE.

9.1.2. Relationship of Adverse Events to Study Drug

Treatment-emergent AEs will be summarized by the total number of events and relationship. Treatment-emergent AEs will also be summarized by the number of subjects with at least 1 TEAE and will be grouped by SOC, PT, and relationship to study drug. The relationships will be collected as the possibility that study drug caused the event. The possible relationships are “not related,” “possibly related,” “probably related,” and “definitely related.” A treatment-related TEAE is an AE with any

relation to study drug other than “not related.” In the TEAE relationship table, if a subject reports multiple occurrences of the same AE, only the most closely related occurrence will be presented for a given SOC or PT. Treatment-emergent AEs that are missing relationship will be presented in the summary table as “possibly related” but will be presented in the data listing with a missing relationship. The relationship or association of the study drug in causing or contributing to the AE will be characterized by the investigator using the following classification and criteria:

<u>Not Related:</u>	This relationship suggests that there is no association between the study drug and the reported event.
<u>Possibly Related:</u>	This relationship suggests that treatment with the study drug caused or contributed to the AE, i.e., the event follows a reasonable temporal sequence from the time of drug administration or follows a known response pattern to the study drug, but could also have been produced by other factors.
<u>Probably Related:</u>	This relationship suggests that a reasonable temporal sequence of the event with drug administration exists and, based upon the known pharmacological action of the drug, known or previously reported adverse reactions to the drug or class of drugs, or judgment based on the investigator’s clinical experience, the association of the event with the study drug seems likely. The event disappears or decreases on cessation or reduction of the dose of study drug.
<u>Definitely Related:</u>	This relationship suggests that a definite causal relationship exists between drug administration and the AE, and other conditions (concurrent illness, progression/expression of disease state, or concurrent medication reaction) do not appear to explain the event. The event reappears or worsens if the study drug is readministered.

9.1.3. Severity of Adverse Event

Treatment-emergent AEs will be summarized by the total number of events and severity. Treatment-emergent AEs will also be summarized by the number of subjects with at least 1 event and will be grouped by SOC, PT, and severity. In the TEAE severity table, if a subject reported multiple occurrences of the same TEAE, only the most severe will be presented for a given SOC or PT. Treatment-emergent AEs that are missing severity will be presented on tables as “severe” but will be presented in the data listing with a missing severity. Treatment-emergent AEs will be classified by the investigator as follows:

<u>Mild:</u>	These events require minimal or no treatment and do not interfere with the subject’s daily activities.
<u>Moderate:</u>	These events result in a low level of inconvenience or concern with the therapeutic measures. Moderate events may cause some interference with normal functioning.
<u>Severe:</u>	These events interrupt a subject’s usual daily activity and may require systemic drug therapy or other treatment. Severe events are usually incapacitating.

9.1.4. Serious Adverse Events

The seriousness of an AE should be assessed by the investigator independently from the severity of the AE. A serious AE is defined as any untoward medical occurrence that at any dose results in death, is immediately life threatening, requires inpatient hospitalization or prolongation of existing inpatient

hospitalization, results in significant disability/incapacity, or is a congenital anomaly/birth defect. Important medical events that may not result in death, be life threatening, or require hospitalization may be considered serious AEs when, based upon medical judgment, they may jeopardize the subject or may require medical or surgical intervention to prevent one of the outcomes listed above.

The total incidence of serious AEs and study drug related serious AEs will be presented in separate tables. Serious AEs and study drug related serious AEs will also be summarized by the number of subjects with at least 1 event and will be grouped by SOC and PT. If a subject reported multiple occurrences of the same serious AE, only 1 event will be counted for a given SOC or PT. Serious AEs will also be listed separately.

9.1.5. Adverse Events Leading to Study Discontinuation

All TEAE data collected with a “Yes” to question “Caused study discontinuation?” will be summarized in a table and presented in a listing. There is not an option to discontinue treatment and remain in the study; therefore, treatment discontinuation and study discontinuation are equivalent.

9.1.6. Death

All subject deaths during this study will be collected and presented in a listing. The information that is presented includes date of death, autopsy performed (yes/no), death certificate completed (yes/no), AE number, and relationship to study drug of the SAE that led to death.

9.2. Clinical Laboratory Evaluations

Clinical laboratory assessments will be conducted at screening only. These will include hematology, chemistry, electrolytes, liver function tests, and renal function parameters. The actual values for all laboratory measurements will be listed by subject and will include any scheduled and unscheduled visits.

Laboratory assessments will be performed by a central laboratory. All listings will be based on the SI units provided by the central laboratory. Each subject’s hematology and chemistry values will be flagged as “low” or “high” relative to the normal ranges of the central laboratory or as “unknown” if no result is available. Each subject’s urinalysis values will be flagged as “positive,” “negative,” or if no value is available, “unknown.”

9.2.1. Hematology

The following laboratory tests will be included at the screening assessment: red blood cell count, total white blood cell count (including basophils, eosinophils, lymphocytes, monocytes, and neutrophils), platelet count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, and mean corpuscular hemoglobin concentration.

9.2.2. Chemistry

The following laboratory tests will be included at the screening assessment: sodium, potassium, bicarbonate, chloride, calcium, blood urea nitrogen, random glucose, alkaline phosphatase, creatinine, alanine aminotransferase, aspartate aminotransferase, gamma glutamyltransferase, and total and direct bilirubin.

9.2.3. Urinalysis

Urine drug screen test is only conducted at the screening visit. If the test was performed, the date and result (positive/negative) of test will be included in a separate by-subject listing.

9.3. Vital Sign Measurements

Vital signs will be measured at screening and check-in and before discharge from the study center at Visits 2, 3, 4, and 5. Vital sign measurements will include pulse, systolic and diastolic blood pressure, and respiration rate. All vital signs will be measured after the subject has been supine for 5 minutes. Vital sign actual measurements and change from baseline measurements will be summarized by treatment group for scheduled visits; however, listings will also include any unscheduled visits.

9.4. Physical Examination

A brief physical examination will be performed at screening and should include evaluation of height (cm); body weight (kg); appearance; skin; head and neck; eyes, ears, nose, and throat; chest and lungs; heart; abdomen; and extremities. Each appearance area result will be recorded as normal, abnormal, or not done. The subject's height and weight measured at screening will be used to calculate BMI used for the STOP-Bang questionnaire and to ensure that the subject meets the BMI entry criterion. Physical examination findings will be provided in by-subject data listings.

9.5. Other Safety Data

The safety set will be used for additional safety endpoints, and summaries will be presented by treatment group and overall. Data will also be presented in a by-subject data listing for the below endpoints.

The KSS is completed after waking at Visit 2 through Visit 5. The KSS [[Akerstedt et al 2014](#)] is self-reported 9-point Likert instrument on which subjects rate their sleepiness as: "extremely alert" (score = 1), "alert" (score = 3), "neither alert nor sleepy" (score = 5), "sleepy-but no difficulty remaining awake" (score = 7), and "extremely sleepy/ fighting sleep" (score = 9). The steps in between have a scale value but no verbal label. The KSS responses will be analyzed as a continuous variable using descriptive statistics and a repeated measures ANOVA model. Due to anticipated variation in the duration between Lights On and completing the KSS test, the descriptive summary will subgroup the results by subjects that completed the test in <30 minutes versus ≥ 30 minutes. In addition, the number of minutes between Lights On and the KSS test will be included as a covariate in a separate repeated measures ANOVA model as a sensitivity analysis.

The DSST will be completed prior to lights out and after waking at each treatment period Visit 2 through Visit 5. DSST consists of 9 digit-symbol pairs, followed by a list (rows) of digits. Under each digit is an open field where the subject is asked to write down on paper the corresponding symbol as quickly as possible. The key outcome parameter for the DSST is the number of correct responses in 90 seconds. The descriptive summary will include after waking measurements following PSG assessment. In addition, a repeated measures ANOVA model will be used to model the correct number of responses after waking as the dependent variable. Due to anticipated variation in the duration between Lights On and completing the DSST test, the descriptive summary will subgroup the results by subjects that completed the test in <30 minutes versus ≥ 30 minutes. In addition, the number of minutes between Lights On and

the DSST test will be included as a covariate in a separate repeated measures ANOVA model as a sensitivity analysis.

Before being discharged from the study center after waking at Visits 2 through Visit 5, the subjects must pass the tandem gait and the Romberg test to assess whether they may be safely discharged. Failure to pass either test will be considered an AE, and subjects will remain at the study center until they are able to pass. Test results will be summarized for both tests.

At the screening visit and upon check-in to the study center at Visits 2, 3, 4, and 5, a urine or serum pregnancy test will be performed for female subjects who are of childbearing potential. Data will be listed by subject only.

At the screening visit and upon check-in to the study center at Visits 2, 3, 4, and 5, subjects will undergo an alcohol breath test. Any subject who tests positive for alcohol will not be administered study drug. Data will be listed by subject only.

10. Interim Analysis

After approximately 50% of the subjects have completed the study, the sample size to achieve at least 80% power for the primary endpoint analysis will be recalculated based on the current blinded data. These data will have undergone routine data cleaning, but the data will not be expected to be 100% cleaned nor will the data be frozen or locked, and the blind will not be revealed. A data extraction will be coordinated by the team when it's anticipated that 50% of the subjects have completed the study (i.e., final safety follow-up call).

The treatment group will remain blinded and will not be included in the repeated measures ANOVA model, thereby pooling the variance across treatment groups. The mFAS will be used with TST as the response variable. PROC MIXED in SAS® version 9.3 or higher will only include period as a fixed effect. Restricted Maximum Likelihood method and the REPEATED statement will be used to model the within-subject correlation as a homogenous compound symmetry matrix. The total variance will be partitioned into within-subject and between-subject components using the COVTEST option in the PROC MIXED statement. The square root of the within-subject variance will be entered into the Sample Size Recalculation Report given in [Appendix 13.4](#); and the sample size required to maintain 80% power for the primary efficacy analysis. The revised power estimate maintaining 68 subjects using the revised variability estimate will also be provided.

To determine the amount of missing information in the data, the number of subjects will be multiplied by 4 to determine the number of expected TST measurements. The percentage of missing data will be calculated as $[1 - (\text{actual number of TST measurements} / \text{expected number of TST measurements})] * 100\%$. For example, if 34 subjects completed the study at the time of the data import from the clinical sleep laboratory, and 3 subjects had missed their last visit and 1 subject didn't return after their initial visit, then 6 TST measurements would be missing and the percent of missing data would be calculated as $[1 - (130/136)] * 100\% = 4.4\%$ missing data. The calculated percentage of missing data will be entered into the Sample Size Recalculation Report.

Any additional considerations regarding the Sample Size Recalculation recommendation will be provided in the “Additional Comments” field of the report, and an overall recommendation to increase the total blinded sample size to a specific count or to continue the trial without modifying the sample size will be recorded in the report.

11. Changes in Planned Analysis

Since the protocol was approved, sQUALITY and sNAW based on the PSQ have been added as secondary efficacy endpoints with analyses described in [Section 8.2](#) and [Section 8.3](#).

A sensitivity analysis was removed from the primary endpoint, i.e., the effects of sex and age group (median split) were not added to the model as fixed factors for a covariate adjustment. However, the separate subgroup analyses with age group (median split) and sex were conducted.

Sleep history was not summarized, but was listed.

12. References

Akerstedt T, Anund A, Axelsson J, et al. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *J Sleep Res.* 2014;23(3):240-52.

Dmitrienko A, D'Agostino R Sr. Traditional multiplicity adjustment methods in clinical trials. *Stat Med.* 2013;32(29):5172-218.

Johns MW. Epworth Sleepiness Scale (ESS) [Internet]. Lyon, France: Mapi Research Trust; 1990. [updated 1997; cited 30 November 2017]. Available from: <https://eprovide.mapi-trust.org/instruments/epworth-sleepiness-scale>

Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997;53(3):983-97.

Kuehl RO. Design of experiments: statistical principles of research design and analysis, 2nd ed. Pacific Grove: Duxbury; 2000.

The official STOP-Bang website [Internet]. Toronto, Canada: Toronto Western Hospital, University Health Network, University of Toronto; 2012. STOP-Bang Questionnaire; 2012 [cited 30 November 2017] [2 screens]. Available from: <http://www.stopbang.ca/osa/screening.php>

Yarandi HN. Crossover designs and proc mixed in SAS. Paper SD04. Proceedings of the SouthEast SAS Users Group, Nashville, TN; 2004.

13. Appendices

13.1. Schedule of Events

	Screening ^a	Treatment and Polysomnography Periods ^b						Follow-Up Safety Phone Call ^b
		Period 1 Visit 2 Day 1	Period 2 Visit 3 Day 8	Period 3 Visit 4 Day 15	Period 4 Visit 5 Day 22	Period 5 Visit 6 3-7 days post dose		
Study Visit	Visit 1							
Study Day	Day -21 to Day -8							
Assessments								
Informed consent	X							
Demographics	X							
Medical/psychiatric/medication history	X							
Sleep history ^c	X							
Epworth Sleepiness Scale	X							
Brief physical examination ^d	X							
Vital sign measurements ^e	X	X	X	X	X			
Urine pregnancy test (females of childbearing potential)	X	X	X	X	X			
Urine drug screen	X							
Alcohol breath test	X	X	X	X	X			
Clinical laboratory assessments ^f	X							
Inclusion/exclusion criteria	X	X						
Paper sleep diary issued ^g	X	X	X	X	X			
Diary return and review		X	X	X	X			
Study center check-in for overnight visit ^h		X	X	X	X			

	Screening ^a	Treatment and Polysomnography Periods ^b						Follow-Up Safety Phone Call ^b
		Period 1 Visit 2 Day 1	Period 2 Visit 3 Day 8	Period 3 Visit 4 Day 15	Period 4 Visit 5 Day 22	Period 5 Visit 6 3-7 days post dose		
Study Visit	Visit 1							
Study Day	Day -21 to Day -8							
Assessments								
Adverse events	X	X	X	X	X	X ⁱ		
Concomitant medications	X	X	X	X	X			
Randomization ⁱ		X						
Treatment administration		X	X	X	X			
Polysomnography		X	X	X	X			
Post-Sleep Questionnaire		X	X	X	X			
Karolinska Sleepiness Scale		X	X	X	X			
Digit Symbol Substitution Test training	X							
Digit Symbol Substitution Test		X	X	X	X			
Predischarge evaluation ^k		X	X	X	X			

a Potential subjects will be evaluated during a screening period lasting at least 8 and no more than 21 days before the first day of the first treatment period.

b Example study days are shown for a 1-week washout interval. Washout intervals must be no less than 5 days. Actual study days for Treatment Periods 2, 3 and 4 and the follow-up safety phone call will depend on the length of interdose washout intervals.

c Sleep history will be assessed using the STOP-Bang questionnaire, which includes 4 subjective items (STOP: snoring, tiredness, observed apnea, and high blood pressure) and 4 demographics items (Bang: body mass index, age, neck circumference, and gender).

d A brief physical examination will include evaluation of height (meters); body weight (kg); appearance; skin; head and neck; eyes, ears, nose, and throat; chest and lungs; heart; abdomen; and extremities. Each subject's height and weight measured at screening will be used to calculate their body mass index.

e Vital sign measurements will include pulse, blood pressure, and respiration rate. All vital signs will be measured after the subject has been supine for 5 minutes. On polysomnography recording days, vital signs will be measured at check-in approximately 7 hours earlier than each subject's median habitual bedtime and again before discharge from the study center.

f Clinical laboratory assessments will include hematology, chemistry, electrolytes, liver function tests, and renal function parameters.

g Subjects are to record the time they go to bed with the intention of sleeping and the time they get up in a pretreatment diary for a minimum of 7 days during the screening period, with at least 5 entries completed over the 7 days ≥70% compliance). Information recorded in the diary will be communicated to study personnel no later than 24 hours before check-in at Visit 2 to allow the calculation of each subject's median habitual bedtime. Subjects will return their paper diary to study personnel at each visit during the treatment and PSG period. Upon discharge from the study center at Visits 2, 3, and 4, subjects will be dispensed a new daily washout interval diary to record the time they went to bed with the intention of sleeping and the time they got up. Subjects will be instructed to maintain their normal sleep patterns.

- h Before Visit 2, subjects will be contacted by study personnel and told the time and date to present for their first overnight visit. Appendix 13.2 describes the timing of study procedures upon check in to the study center. Upon discharge from the study center at Visits 2, 3, and 4, subjects will be contacted the time and date to return for the next overnight visit.
- i Subjects will be contacted by phone for a safety followup within the 7 days after completion of Treatment Period 4 (Visit 5) but at least 72 hours after administration of the final dose of study drug. Subjects who discontinue study drug or withdraw from the study prematurely will be contacted within 24 hours of the final dose of study drug. Outcomes of any adverse events will be discussed and recorded. If indicated, a visit to the study center for safety assessments will be scheduled.
- j Once eligibility has been reconfirmed at Visit 2, subjects will be randomly assigned to 1 of 4 treatment sequences. Subjects will be crossed-over to their next treatment at each visit based on their treatment sequence randomization.
- k Before leaving the study center at Visits 2, 3, 4, and 5, subjects will undergo a brief discharge evaluation consisting of a gait and the Romberg test.

13.2. Sleep Study Events by Day for Treatment Periods

Treatment Period Day	Time Relative to Habitual Bedtime	Time Relative to Lights Out/Waking	Procedure ^a
Day 1	Approximately 7 hours before	Approximately 120 minutes before lights out	<ul style="list-style-type: none"> Subjects report to the study center Alcohol breathalyzer administered Urine pregnancy testing performed for females of childbearing potential Return and review diary Pretreatment assessment of adverse events Update concomitant medications and therapies Vital signs measured
	Approximately 6.5 hours before	Approximately 90 minutes before lights out	<ul style="list-style-type: none"> Baseline Digit Symbol Substitution Test measurement obtained Subjects provided a light, low-fat snack (e.g., fruit and crackers)
	Approximately 6 hours before	Approximately 60 minutes before lights out	<ul style="list-style-type: none"> PSG electrodes applied to the subject and machine calibrated
	5.5 hours before	30 minutes before lights out	<ul style="list-style-type: none"> Subjects administered dose of study drug Oral cavity check performed administered
Day 2	Period 1: 5 hours ± 30 minutes before Periods 2, 3, and 4: Same as Period 1 ± 15 minutes before	Lights out	<ul style="list-style-type: none"> Subjects go to bed PSG biocalibration Begin PSG recording
	—	Waking (8 hours after lights out)	<ul style="list-style-type: none"> End PSG recording Subjects awakened and allowed a bathroom visit
	—	30 minutes after waking (8.5 hours after lights out)	<ul style="list-style-type: none"> Subjects complete the morning assessments in the following order: <ol style="list-style-type: none"> Post-Sleep Questionnaire Karolinska Sleepiness Scale Digit Symbol Substitution Test

			<ul style="list-style-type: none">• Subjects served a standard breakfast• Before leaving the study center, subjects undergo a predischARGE evaluation consisting of tandem gait and the Romberg test, as well as an assessment of vital signs and adverse events to determine whether they may be safely discharged from the study center.• Subjects who pass this evaluation will be discharged from the study center. Failure will be considered an adverse event, and subjects will remain at the study center until able to pass.
--	--	--	---

Abbreviation: PSG, polysomnography

a Procedures must be conducted in the order they appear in the bulleted list.

13.3. Imputation for Partial and Missing Dates

Adverse Event

A TEAE will be assigned to the treatment in 1 period if the start date of that TEAE is on or after the start date of that period, up through the day prior to the start date of the following period (or 7 days after last dose, for the last period).

- If onset date is completely missing:
 - If TEAE end date is missing or on/after the date of first dose, then onset date is set to date of first dose.
 - If TEAE end day is missing, but the month and year of the TEAE end date is \geq the month and year of the first dose date, then the onset date is set to the 1st day of the month and year.
 - If TEAE end month is missing, but the year of the TEAE end date is \geq the year of the first dose date, then the onset date is set to January 1.
 - If TEAE end date is prior to the date of the first dose, then onset date is not imputed.
- If onset year is present, but month is missing:
 - If year = year of first dose, then set onset month and day to month and day of first dose.
 - If year < year of first dose, then set onset month and day to December 31st.
 - If year > year of first dose, then set onset month and day to January 1st.
- If month and year are present, but day is missing:
 - If year = year of first dose and
 - If month = month of first dose, then set day to day of first dose date.
 - If month < month of first dose, then set day to last day of month.
 - If month > month of first dose, then set day to 1st day of month.
 - If year < year of first dose, then set day to last day of month.
 - If year > year of first dose, then set day to 1st day of month.
- For all other cases, set onset date to date of first dose.

Concomitant Medications

A concomitant medication will be assigned to the treatment in 1 or more periods if the start date of that medication is on or after the start date of that period, up through the day prior to the start date of the following period (or 7 days after last dose, for the last period).

- If start date is completely missing or year is missing, then start date will not be imputed.
- If start year is present and month and day are missing, then set start month and start day to January 1.
- If start year and day are present and month is missing, then set start month to January.
- If start year and start month are present and start day is missing, then set start day to 1st day of month.

- If end date is completely missing or year is missing, then end date will not be imputed.
- If end year is present and month and day are missing, then set end month and day to December 31.
- If end year and day are present and month is missing, then set end month to December.
- If end year and end month are present and end day is missing, then set end day to last day of the month.

13.4. Sample Size Recalculation Report

PPD Biostatistics review of sample size for the Sequential Medicine SM-A-05 Phase 3 trial, with data extracted on [_____].

The original assumption included the square root of the within mean square error from the ANOVA model to be 88 minutes with an effect size of 44 minutes between SM-1 and D+Z. However, the blinded analysis estimated the variability at [_____minutes], which resulted in a required sample size of [_____] to maintain a minimum of 80% power for the primary efficacy endpoint. The current power under the new variability estimate with 68 subjects is [_____%].

The original assumption did not include any loss of power due to subjects lost to follow-up or missed visits. The number of data points expected in the analysis were 4 visits times 68 subjects completed the study = 272 pieces of information. However, [_____ %] of the data points were missing in the blinded analysis from the subjects up to the time of the sample size recalculation. While these missing data have already inflated the variance in the ANOVA model, the assumption is made that this missing data rate will remain similar at the end of the study.

As a result, the recommendation is:

- To continue the trial unmodified.
- To continue the trial and increase the total blinded sample size to [_____subjects]

Additional Comments: