Statistical Analysis Plan

VIVUS, Inc.

OB-403

Document Version: Final 1.1
Document Date: April 01, 2021

Table of Contents

Sponsor Name: Vivus, Inc.
Sponsor Protocol ID: OB-403

Approvals

The undersigned agree that all required reviews of this document are complete, and approve this Statistical Analysis Plan as final. Programming of the tables, figures and listings based upon the specifications within this document can proceed.

Approved by

Signature   Date

Printed Name/Title

Vivus Approval:

Printed Name/Title

Printed Name/Title

CONFIDENTIAL

Statistical Analysis Plan

Reviewers

The following reviews of the SAP were conducted:

Name and Title	Role	Version Last Reviewed	Company/ Organization
	Lead Statistician	Final 1.1	
	Peer Review Statistician	Final 1.1	
	Lead Programmer	Final 1.0	
	Medical Writer	Final 1.1	
	Sponsor	Final 1.1	Vivus
	Sponsor	Final 1.1	Vivus
	Sponsor	Final 1.1	Vivus
	QC Reviewer	Final 1.1	
	Biostatistician	Final 1.1	

Version History

Version #	Description of Changes	Version Date
Final 1.0	Original Version	01Sept2020
1.1	Updates to Sections 5.2 and 6	07Jan2021

Glossary of Abbreviations

Abbreviations pertain to the SAP only (not the TFLs).

Abbreviation	Term
AE	Adverse event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
BP	Blood Pressure
BMD	bone mineral density
BMC	bone mineral content
CANTAB	Cambridge Neuropsychological Test Automated Battery
CI	Confidence Interval
CRF	Case report form
C-SSRS	Columbia Suicide Severity Rating Scale
DXA	Dual-energy X-Ray Absorptiometry
ECG	Electrocardiogram
ET	Early Termination
HR	Heart Rate
LSD	protected Least Significant Difference
ITT	Intention-to-treat
IWQOL-Kids	Impact of Weight on Quality of Life-Kids
LOCF	Last observation carried forward
MAR	missing at random
MCMC	Markov chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intention-to-treat
MMRM	Mixed Effects Model with Repeated Measures
MNAR	missing not at random
PHQ-9	Patient Health Questionnaire-9
SAE	Serious Adverse Event
SD	Standard Deviation
SOC	System Organ Class
TEAEs	Treatment–Emergent AEs
TFLs	Tables, Figures and Listings

STATISTICAL ANALYSIS PLAN AMENDMENT 1

Not applicable.

1. Source Documents

The Statistical Analysis Plan (SAP) was written based on the following documentation:

Document	Date	Version
Protocol	24 Aug 2017	Original
Protocol Amendment #1	05 Oct 2018	Amendment # 1
Protocol Amendment #2	21 Jun 2019	Amendment # 2
<e>CRF	17 Sep 2019	3.0
DMC Charter	27 Feb 2020	1.0
DMC Charter Amendment	NA	NA

2. Protocol Details

2.1 Study Objectives

The primary objectives are to evaluate the safety and efficacy of VI-0521 for the treatment of obesity in adolescents. The secondary objective is to characterize changes in obesity-related risk factors.

2.2 Overall Study Design

In this multicenter, randomized, double-blind, placebo-controlled, parallel-design study, approximately 200 subjects will be enrolled at approximately 20 sites in the United States. Subjects will be randomly assigned in a 1:1:2 ratio to placebo, N= 50; mid-dose (PHEN/TPM 7.5 mg/46 mg), N= 50; or top-dose (PHEN/TPM 15 mg/92 mg), N =100, of VI-0521, to be taken orally once daily in the morning. Randomization will be stratified by age (12-14 vs 15-16 years old) and gender. The study will consist of a screening period of up to 28 days, followed by a 56-week treatment period.

Subjects will be instructed to follow a mild hypocaloric diet modification program representing a 500-calorie/day deficit and to implement a family-based lifestyle modification program for adolescents, as tolerated, throughout the study period. The lifestyle program will include physical activity, behavior change, and family support. The same lifestyle modification program, specific to this population, will be implemented across all sites. Study drug will be titrated according to the following schema.

Sponsor Name: Vivus, Inc.
Sponsor Protocol ID: OB-403

Group	Treatment Dosage for PHEN/TPM (mg)	Titration Dose for PHEN/TPM (mg)			
		Weeks 1-2	Weeks 3-4	Weeks 13-14	Weeks 15-16
Placebo	0/0	0/0	0/0	0/0	0/0
VI-0521 Mid	7.5/46	3.75/23	7.5/46	7.5/46	7.5/46
VI-0521 Top	15/92	3.75/23	7.5/46	11.25/69	15/92

Subjects who are unable to tolerate the assigned dose may be treated at a reduced dose level or may take a drug holiday as defined in the protocol. In addition, for growth monitoring, investigators will monitor rates of weight loss in treated subjects. For subjects with baseline BMI 95-98th percentile, reduce study drug dosage when BMI is < 85th percentile or when weight loss exceeds an average of 2 lbs (0.9 kg) per week. For subjects with baseline BMI ≥ 99th percentile, reduce study drug when weight loss exceeds an average of 2 lbs (0.9 kg)/week.

All subjects will return at approximately 4-week intervals for study assessments. All female subjects will undergo a pregnancy test at each visit. Subjects who discontinue study drug during the study will be encouraged to remain on study (off study drug) for continued follow-up by attending all remaining visits and have all study-related procedures performed, and to return at the 56-week time point for measurements and evaluations. For those who choose to completely withdraw from the study at any point, the end of study (Week 56) procedures should be completed.

Figure 1. Schematic Diagram of Study Design
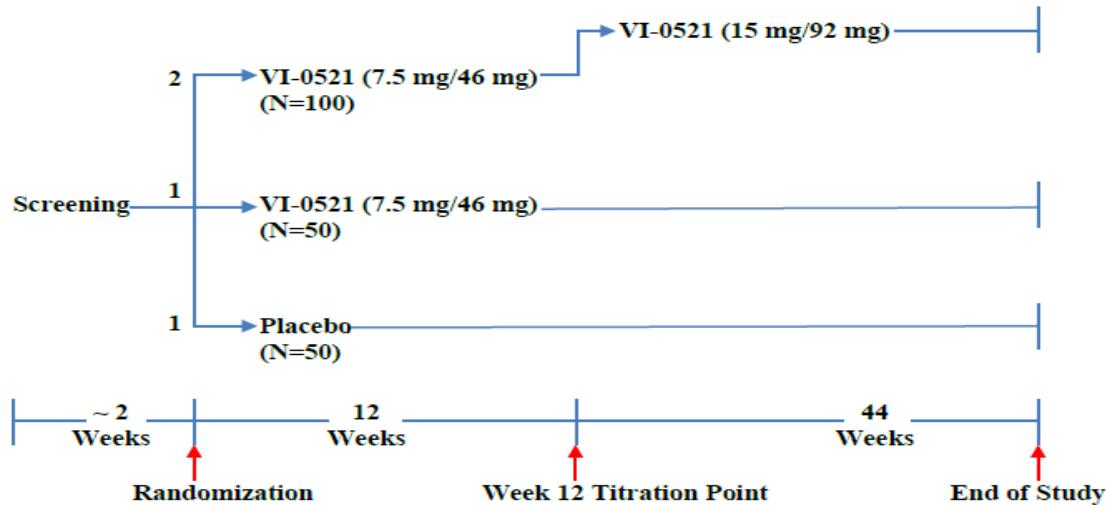
2.3 Sample Size and Power

In previous studies in adults, VI-0521 mid-dose (PHEN/TPM 7.5 mg/46 mg) resulted in a placebo-subtracted BMI reduction of 2.4 units with a standard error approximately 0.16 units and a within treatment standard deviation of 2.9. A very conservative estimate of the treatment difference between the mid-dose and placebo would be 2 units of BMI which represents more than 2 standard errors below what was observed before. If we enroll 200 subjects (50 placebo, 50 mid-dose, and 100 top-dose), we will have at least 90% power to detect a statistically significant difference between the top-dose (PHEN/TPM 15 mg/92 mg) and the placebo because we could assume that the top-dose will have a higher effect size than the mid-dose. This calculation assumes that there will be an approximately 30% dropout rate. This sample size will also provide approximately 80% power to detect a statistically significant difference between the mid-dose and placebo.

3. Efficacy and Safety Variables

3.1 Primary Efficacy Endpoints

- The mean percent change in BMI from baseline to end of study (Week 56).

3.2 Secondary Efficacy Endpoints

- Percent of subjects achieving a reduction ≥ 5%, ≥ 10% and ≥ 15% of baseline BMI at Week 56;

- Change from baseline in waist circumference at Week 56;

- Change from baseline in fasting insulin and Whole Body Insulin Sensitivity Index (Matsuda) at Week 56;

- Percent change from baseline in triglycerides and HDL-C at Week 56;

- Change from baseline in blood pressure at Week 56.

3.3 Exploratory Endpoints

- Evaluate effects of treatment on Impact of Weight on Quality of Life-Kids (IWQOL-Kids) questionnaire scores.

- Changes in various glycemic and lipid markers.

- Change in BMI Z-score.

3.4 Safety Endpoints

Safety will be assessed by evaluating adverse events (AEs)/serious adverse events (SAEs); vital signs, laboratory parameters (screening, periodically during the study and end of study); electrocardiograms; physical examinations; cognitive function tests using the Cambridge Neuropsychological Test Automated Battery (CANTAB). All subjects will be screened for the presence and severity of depression using the PHQ-

9: Modified for Teens and for suicidal/ideation using the Columbia Suicide Severity Rating Scale (C-SSRS) and follow up assessments will be done at each visit after treatment has been initiated. Bone age (X-ray of the hand and wrist) will be evaluated at baseline and end of study or early termination. Effect on bone mineral density and bone mineral content, as evaluated by Dual-energy X-Ray Absorptiometry (DXA) will be performed at baseline, end of study or early termination, at selected sites.

3.5 DXA (Sub Study)

A bone health sub-study will be conducted at selected sites in approximately 100 subjects (25 each on placebo and mid-dose, and 50 on top-dose) to assess the effect of VI-0521 administration on bone health using Dual X-ray Absorptiometry (DXA). DXA scans of the posterior-anterior (PA) spine (lumbar), and total body less head (TBLH) will be performed at baseline and at the end of study or early termination. Equipment and procedures used to obtain DXA data will be standardized as described in a separate document. Sites involved in DXA measurement will be trained on these procedures prior to performing scans on study subjects. Scans will be read at a central facility and the reader will be blinded.

The following enrollment criteria will apply:

1) Both male and female will be eligible to participate;
2) Subjects with a history of any non-traumatic fracture will not be eligible;
3) Subjects with juvenile osteoporosis at baseline will not be eligible; and
4) Subjects must meet manufacturer equipment specifications with regard to height and weight limitations.

4. Analysis populations

4.1 Randomized

This population will be comprised of all subjects who were initially randomized. This population will be used for summaries of subject disposition and baseline subject characteristics.

4.2 Safety

This population will be comprised of all subjects who were initially randomized and received at least one dose of study drug. This will be the primary population for all summaries of subject disposition and baseline characteristics, and safety analyses for purposes of regulatory submissions. The safety population is based on the treatments actually received.

Sponsor Name: Vivus, Inc.
Sponsor Protocol ID: OB-403

4.3 Intent-to-treat (ITT)

This population will be comprised of all subjects who were initially randomized and received at least one dose of study drug. This will be the primary population for all summaries of efficacy analyses for purposes of regulatory submissions. The ITT population is based on the randomized treatments.

4.4 Modified Intent-to-treat (mITT)

This population will be comprised of all randomized study subjects who receive study treatment and return for at least one post-randomization assessment of height and weight. This population will be used for the analysis of all efficacy variables for all other purposes, including but not limited to publications, presentations, and robustness of sensitivity of analyses.

4.5 Special Subpopulations

BMI change will be tabulated by age and gender. Additionally, further exploratory subgroup analyses of the primary efficacy endpoints may include evaluation by race and/or other subgroups deemed medically and/or scientifically important.

5. DATA Handling

5.1 Time Points and Visit Windows

Day 0 (Visit 2) is defined as the Baseline/Randomization visit, is also the first day of treatment. Relative days are calculated as (assessment date – Day 1 date). The day prior to Day 0 is Day -1.

The following visit windows defined in Table 1 will be used for the by-visit analyses of the primary endpoint, the secondary and other endpoints. All other analyses will use the nominal study visit as defined in the Study Schedule and eCRF.

If there are multiple visits (scheduled or unscheduled) within a visit window, the measurement closest to the target day of the visit will be used in the analysis. If the measurements are equally distant to the target day, then the later one will be used in the analysis. If both scheduled and unscheduled visits fall within the same visit window, the scheduled visit will be used for analysis.

Table 1   Definition of Visit Windows

Visit	Visit time (Week)	Acceptable visit window	Study Day Analysis Visit Window Range
1 Screening	Screening	-28 days from baseline	NA
2 Baseline[a]	Week 0	±3 days	Day -27 to Day 2
3 Treatment	Week 4	±1 week	Day 3 to Day 41(Week 6)
4 Treatment	Week 8	±1 week	Day 42 to Day 69(Week 10)
5 Treatment	Week 12	±1 week	Day 70 to Day 97(Week 14)
6 Treatment	Week 16	±1 week	Day 98 to Day 125(Week 18)
7 Treatment	Week 20	±1 week	Day 126 to Day 153(Week 22)
8 Treatment	Week 24	±1 week	Day 154 to Day 181(Week 26)
9 Treatment	Week 28	±1 week	Day 182 to Day 209(Week 30)
10 Treatment	Week 32	±1 week	Day 210 to Day 237(Week 34)
11 Treatment	Week 36	±1 week	Day 238 to Day 265(Week 38)
12 Treatment	Week 40	±1 week	Day 266 to Day 293(Week 42)
13 Treatment	Week 44	±1 week	Day 294 to Day 321(Week 46)
14 Treatment	Week 48	±1 week	Day 322 to Day 349(Week 50)
15 Treatment	Week 52	±1 week	Day 350 to Day 377(Week 54)
16 End of Study/ET[b]	Week 56/ET	±1 week	Day 378 to Day 398(Week 57)

[a] Baseline Visit (Visit 2) can be scheduled up to 4 weeks from Screening.
[b]. ET = Early Termination

5.2 Handling of Dropouts, Missing Data, and Outliers

For subjects who discontinue treatment prior to study completion, every attempt will be made to have them continue with clinic visits and study assessments. Particular attention will be given to collecting Week 56 assessments of weight and height, regardless of when subjects discontinued treatment.

Missing data will not be imputed for safety analyses. The safety evaluations will be performed on observed data only.

For the MMRM (mixed effects model with repeated measures) method used for the primary analysis on the primary endpoint, the retrieved dropouts will be used to impute missing data for subjects who discontinue the study prematurely. If there are no sufficient retrieved dropouts, then the wash-out imputation method will be applied.

The following sensitivity analyses will be considered to explore the impact of missing data on the conclusion of the primary analysis.

The first sensitivity analysis is using a multiple imputation method based on the monotonic missing pattern under the assumption of missing at random (MAR). The intermittent missing data will be imputed using multiple imputation MCMC (Markov chain Monte Carlo) procedure.

The second sensitivity analysis is also using multiple imputation, however, under the assumption of missing not at random (MNAR) to explore the validity of MAR using pattern-mixture model:

1) For subjects who discontinue study participation prior to Week 56 and do not have follow-up visit, the missing data will be imputed using the observed data from the subjects in the same arm who discontinue the study treatment but have the primary endpoint measurement in the follow up visit using a regression method. The intermittent missing data will be imputed using multiple imputation MCMC procedure. An ANCOVA (analysis of covariance) model using a similar mixed procedure (without the repeated measures) as the primary analysis will be applied to these multiple-imputed % change in BMI at Week 56 with treatment, baseline BMI value as a covariate, and age and gender as stratification factors. The results of ANCOVA analysis on the multiple imputed datasets will be combined and summarized.

2) Tipping point analyses: Subjects from the treatment arm who drop out the study will have their unobserved efficacy data imputed by the observed data from completers in the same arm using the multiple imputation method based on the monotonic missing pattern under the assumption of MAR with the resulting imputed values further worsened by an amount δ. Subjects who drop out the study from the control arm will be assumed to exhibit the same evolution of the disease as the completers in control arms and their values will be imputed by the multiple imputation method based on the monotonic missing pattern under the assumption of MAR above without the addition of δ. Sensitivity analysis may be performed for a range of δ to find a "tipping point" value of δ at which study conclusions start to change. When δ=0 the missing data are assumed to be MAR. When δ > 0, the missing data are assumed to be MNAR.

The third sensitivity analysis is the last observation carried forward (LOCF). For those subjects who discontinue study participation prior to Week 56, the last observed weight and height will be used to derive the change in BMI.

Similar analyses will be performed for the primary endpoint for the mITT population.

6. Statistical Methods

6.1 General Principles

All data processing, summarization and analyses will be performed using SAS Environment / Version 9.3 (or later) of the SAS® statistical software package.

The following principles will be applied to all TFLs unless otherwise stated:

Principle	Value
Significant tests	Two-sided and use a 5% significance level.
Treatment group labels and order presented	Placebo VI-0521 Mid VI-0521 Top
Tables	Data in summary tables presented by treatment group, assessment and visit (where applicable).
Listings	All data collected presented by treatment group, site, subject, and visit (where applicable), unless otherwise specified.
Descriptive summary statistics for continuous variables	Number of observations (N), mean, standard deviation (SD), minimum, median and maximum.
Descriptive summary statistics for categorical variables	Frequency counts and percentages [n (%)].
Denominator for percentages	Number of subjects in the pertinent analysis population, unless stated otherwise in table shells.
Include "Missing" as category	Demographics and Other Baseline Characteristics only
Display for 0 percentages	Blank
Display to one more decimal place than collected value	Mean Median Minimum Maximum
Display to two more decimal places than collected value	Standard Deviation Confidence Interval
Limit of precision for displays	3 decimal places
Date Format	DDMMMYYYY

6.2 Subject Disposition and Data Sets Analyzed

Subject disposition will be listed and summarized by treatment group and overall and will include the number and percentage of subjects:

- Screened;

- Randomized;

- Not randomized;

- Randomized and not treated;

- Treated;

- Included in each study population (Safety, ITT, mITT, DXA Substudy);

- Completed all study visits;

- Completed all study visits on study drug;

- Randomized but discontinue study drug and discontinue from all study visits and procedures;

- Randomized but discontinue study drug but continue with all study visits and procedures; and

- Randomized but discontinue from the study and returned for the Week 56 weight measurement.

For subjects who are not randomized, randomized subjects who discontinue study drug but continue in the study, and randomized subjects who discontinue study drug and study participation, a summary regarding reason of discontinuation will also be provided.

A summary of patient enrollment by site will also be provided by treatment group and overall for the Safety population.

6.3 Protocol Deviations

A protocol deviation is any change, divergence, or departure from the study design or procedures defined in the protocol. Protocol deviations for the study include, but are not limited to, the following:

- Failure to meet inclusion/ exclusion criteria;

- Non-compliance of protocol guidelines;

- Inability to adequately complete study assessments per protocol, etc.

Failure to comply with Good Clinical Practice (GCP) guidelines will also result in a protocol deviation. The Sponsor will determine if a protocol deviation will result in withdrawal of a subject.

All protocol deviations and deviation categories will be listed and summarized for the Safety population.

6.4 Demographics and Other Baseline Characteristics

Demographic and baseline characteristics will be listed and summarized by treatment group and overall for randomized, Safety and DXA substudy populations. Standard descriptive statistics will be presented for the continuous variables of:

- Age at screening (years)

- weight (kg)

- height (cm)

- body mass index (kg/m^2)

- Waist Circumference (cm)

- Fasting insulin

- Whole Body Insulin Sensitivity Index (Matsuda)

- Triglycerides

- HDL-C

- Systolic Blood Pressure (mmHg)

- Diastolic Blood Pressure (mmHg)

The total counts and percentages of subjects will be presented for the categorical variables of:

- Age at screening (12-14 years, 15-16 years)

- Gender

- Race

- Ethnicity

- BMI categories ≥95th to <99th, and ≥99th percentile)

No formal tests of statistical significance will be performed on the demographic and baseline data.

Other baseline measurements, such as vital signs, ECG, will be summarized by treatment group with the post-baseline measurements.

6.4.1 Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.3 (or a later version if updated during the study). All medical history will be listed, and the number and percentage of subjects with any medical history will be summarized for Safety population by system organ class (SOC) and preferred term (PT) for each treatment group and overall.

6.4.2 Prior and Concomitant Medications

Medications received prior to or concomitantly with study drug treatment will be coded by   using the WHO Drug Global Dictionary, Version September 2018 (or a later version if updated during the study), Anatomical Therapeutic Chemical (ATC) Classification codes.

Medications, other than study medication, taken within 30 days prior to screening will be considered as prior medications; while those taken during the study will be considered as concomitant medications.

If a medication cannot be classified as "prior" or "concomitant" after applying imputation rules for missing/incomplete dates, it will be classified as concomitant.

Prior medications and concomitant medications will be listed and summarized together for Safety population.

The number and percentage of subjects using each medication will be displayed together with the number and percentage of subjects using at least one medication within each therapeutic class (ATC-Level 2), chemical subgroup (ATC-Level 4), and preferred term.

6.5 Measurements of Treatment Compliance

Percentage compliance is calculated as:

100 * actual capsules taken/expected capsules taken

Where actual capsules taken is defined as (total number of capsules dispensed – total number of capsules returned) and expected taken is defined as the number of capsules prescribed per day times the number of days taken for study treatment.

Percentage compliance will be summarized descriptively by treatment group for the Safety population.

The number and percentage of compliant subjects will be presented for the Safety population, where compliant is defined as percentage compliance between 80.0% and 120.0% inclusive. The following percentage compliance categories will also be presented:

- <80.0%
- 80% to 120%
- >120.0%

6.6 Efficacy

6.6.1 Primary Efficacy Analysis

The primary endpoint for this study is the mean % change in BMI from baseline to end of study (Week 56).

For subjects who discontinue treatment prior to trial completion, every attempt will be made to have them continue with clinic visits and study assessments. Particular attention will be given to collecting Week 56 assessments of weight and height, regardless of when subjects discontinued treatment. These retrieved dropouts will be used to impute missing data using multiple imputation method for the primary analysis for subjects who discontinue the study prematurely and are missing follow-up observations. This is based on the assumption that the group that best reflects what happened to the non-retrieved dropouts at week 56, are the retrieved dropouts. For patients who terminate the treatment early but decide to stay in the study and have all the data after being off treatment, those data will be used to impute the missing data off-treatment. For any patient who terminates the treatment and study prematurely and permanently, if the patient has the final measurement at Week 56, the measurement will be used for analyses at Week 56. The missing data between the last available measurement and Week 56 will be imputed by interpolation for analyses.

If there are no sufficient retrieved dropouts, the washout imputation method will be applied. This imputation method will only impute the time point at Week 56 for the

active treatment arms using multiple imputation regression with randomization strata and baseline BMI as the predictors. This approach ignores any post-baseline changes in the BMI values when predicting missing Week 56 values and will impute all patients similar to an average placebo patient. For patients in placebo arm, missing data at any time point will be imputed assuming MAR mechanism. The variables used as explanatory variables for imputation include Randomization strata and baseline BMI.

After such imputation, comparisons in the primary endpoint of change from the baseline BMI between treatment groups be assessed for the ITT population using a mixed effects model with repeated measures (MMRM) with factors of treatment, visit, treatment by visit interaction, baseline BMI value, age stratification, and gender stratification. Appropriate contrast will be applied for treatment comparisons at Week 56. The pairwise comparisons of interest are top-dose vs. placebo, and mid-dose vs. placebo, and top-dose vs. mid-dose. The primary null hypothesis will be that there is no treatment difference between any VI-0521 treatment groups and the placebo in the percent change from baseline to Week 56 in BMI. An appropriate contrast will be used for the comparisons at Week 56. The family-wise type 1 error for the comparisons will be controlled by Fisher's protected least significant difference (LSD) method at the 0.05 significance level: placebo, mid-dose, and top-dose will be first compared for overall difference in the percent change from baseline in BMI. Once the overall difference is significant at the 0.05 significance level, the above 3 pairwise comparisons will be conducted using Fisher's LSD method at the 0.05 significance levels. The order for comparisons of interest is top-dose vs. placebo, mid-dose vs. placebo, and top-dose vs. mid-dose. Due to the fact that only three treatments are compared, the above procedure strongly controls the family-wise type 1 error.

6.6.2 Secondary Efficacy Analysis

If both the mid- and top-dose are shown to be statistically significantly better than placebo for the primary endpoint using the Fisher's LSD procedure, then the secondary endpoints will be tested in a stepwise way to preserve the familywise type 1 errors.

Starting from top-dose of VI-0521 treatment, hypothesis tests for secondary efficacy endpoints will be based on null hypotheses that assume no *a priori* differences between placebo and top-dose treatment.

A closed or sequential testing procedure will be used for testing the key secondary efficacy variables. With this hierarchical procedure, the key secondary efficacy variables will only be tested if the primary efficacy analysis is statistically significant.

Analyses will be carried out for the key secondary variables.

Within the key secondary variables, the statistical significance level will be adjusted using the Hochberg method, controlling the familywise error rate at 5%. Correlation

analysis will be implemented among the key secondary variables to ensure the conditions required for Hochberg procedure are satisfied. All endpoints have to be statistically significant in favor of top-dose VI-0521 treatment compared to placebo, after the Hochberg adjustment, in order for the next set in the hierarchy to be tested. The testing is only stopped when all of the alpha is 'exhausted'. The sequential testing will stop at the first endpoint set where top-dose VI-0521 treatment does not demonstrate statistical superiority over placebo.

The above process will be repeated on mid-dose of VI-0521 treatment.

Percent of subjects achieving a reduction ≥ 5% from baseline in BMI at Week 56 will be analyzed using a logistic regression, with treatment, age and gender stratification as the main effect and baseline BMI value as a covariate at the 0.05 significance level. The adjusted odds ratios between the top-dose and placebo, between the mid-dose and placebo, and between the top-dose and mid-dose will be calculated together with their 95% confidence intervals. The p-values for the comparisons will also be generated.

The percent of subjects achieving a reduction ≥ 10% and ≥ 15% from baseline in BMI at Week 56 will be analyzed similarly.

Secondary efficacy endpoints that are continuous variables (change from baseline in waist circumference, fasting insulin, whole body insulin sensitivity index (Matsuda), triglycerides, HDL-C, and systolic and diastolic blood pressure) will be analyzed by a similar MMRM model as for the primary endpoint where the baseline BMI value will be replaced by the baseline value of the corresponding endpoint as a covariate.

The above analyses will be conducted for both the ITT and mITT populations.

6.6.3 Sensitivity Analysis

The MMRM method used for the analysis will be applied to the mITT population with the observed data without imputation. The following sensitivity analyses will also be considered to explore the impact of missing data on the conclusion of the primary analysis.

The first sensitivity analysis is using a multiple imputation method based on the monotonic missing pattern under the assumption of MAR. The intermittent missing data will be imputed using multiple imputation MCMC procedure.

The second sensitivity analysis is also using multiple imputation, however, under the assumption of MNAR to explore the validity of MAR using pattern-mixture model:

1) For subjects who discontinue study participation prior to Week 56 and do not have follow-up visit, the missing data will be imputed using the observed data from the subjects in the same arm who discontinue the study treatment but have the primary endpoint measurement in the follow up visit using a regression method. The

intermittent missing data will be imputed using multiple imputation MCMC procedure. An ANCOVA model using a similar mixed procedure (without the repeated measures) as the primary analysis will be applied to these multiple-imputed % change in BMI at Week 56 with treatment, baseline BMI value as a covariate, and age and gender as stratification factors. The results of ANCOVA analysis on the multiple imputed datasets will be combined and summarized.

2) Tipping point analyses: An additional sensitivity analysis using a 2-way tipping-point strategy will be conducted on the primary endpoint to explore the influence of missing data from active treatment and placebo arms on the overall conclusion from statistical inference. In this approach, a wide spectrum of assumptions regarding the magnitude of missingness (from less conservative to more conservative) is proposed for replacing missing data. All subjects who drop out the study will have their unobserved efficacy data imputed by the observed data from completers in the same arm using the multiple imputation method based on the monotonic missing pattern under the assumption of MAR with the resulting imputed values further worsened by an amount δ. Scenarios where dropouts on active arms have worse outcomes than dropouts on placebo will be included. The analysis finds a 'tipping' point from among these assumptions under which the study conclusions shift from being favorable to the active treatments to being unfavorable. After such a tipping point is determined, clinical judgment can be applied as to the plausibility of the assumptions underlying this tipping point. The tipping point can be identified while the result is no longer statistically significant.

The third sensitivity analysis is the LOCF. For those subjects who discontinue study participation prior to Week 56, the last observed weight and height will be used to derive the change in BMI.

Similar analyses will be performed for the primary endpoint for the mITT population.

6.6.4 Subgroup Analysis

BMI change will be tabulated by age and gender. Additionally, further exploratory subgroup analyses of the primary efficacy endpoints may include evaluation by race and/or other subgroups deemed medically and/or scientifically important.

6.6.5 Exploratory Analysis

IWQOL-Kids will be analyzed by a similar MMRM model as for the primary endpoint where the baseline BMI value will be replaced by the baseline value of the corresponding endpoint as a covariate to evaluate effects of treatment on Impact of Weight on Quality of Life-Kids (IWQOL-Kids) questionnaire scores.

Change from baseline will be presented for various glycemic, lipid markers and BMI Z-score parameters with numerical measures using descriptive statistics. BMI Z-score will be calculated as $[(BMI/M)^L-1]/(L*S)$ where L, M and S values are available in percentile data files.

6.7 Safety

All safety analyses will be done for the Safety population. Safety data will be summarized for all treatment groups.

Safety will be assessed by an evaluation of adverse events (each study visit); laboratory parameters (screening, periodically during the study and end of study); electrocardiograms; physical examinations (screening, end of study); PHQ-9: Modified for Teens, C-SSRS, and vital signs (at each study visit). Descriptive statistics will be generated for the questionnaire data.

6.7.1 CANTAB

Cognitive testing will include CANTAB Paired Associates Learning (PAL), Pattern Recognition Memory (PRM) and Spatial Span (SSP) and will be collected at Screening, Baseline, Week 16 and Week 56 for each treatment assignment. The tasks will evaluate effects on episodic memory, visual pattern recognition memory and working memory capacity, respectively. The key outcome measures to be included in the analysis are described in Table 2.

For each key CANTAB outcome measure, descriptive summary statistics (n, mean, median, standard deviation, minimum and maximum) for change from baseline will be reported for visits at Week 16 and Week 56 across all treatment groups (Placebo, Mid-dose and Top-dose of VI-0521) and stratification factors for age and gender. Line graphs of mean change from baseline to Week 16 and Week 56 by treatment groups and stratification factors will also be produced.

Mixed effects models with repeated measures will be used to generate least squares means and standard errors for change from baseline to Weeks 16 and 56 for each treatment group, controlling for stratification factors age and gender and baseline performance for each CANTAB outcome measure.

The standardized mean difference (effect size) between placebo and each treatment group will be calculated using the least squares mean change from baseline to Week 16 and Week 56 estimates and the pooled standard deviation of change across both treatment and placebo groups.

Table 2 CANTAB Key outcome measures

CANTAB Task	Key Outcome Measure (code)	Description	Cognitive Domain / Direction of Effect
PAL	Total Errors Adjusted (PALTEA)	The number of times the subject chose the incorrect box for a stimulus on assessment problems, plus an adjustment for the estimated number of errors they would have made on any problems, attempts and recalls they did not reach. This measure allows you to compare performance on errors made across all subjects regardless of those who terminated early versus those completing the final stage of the task	Episodic memory/ Lower is better
PAL	First Attempt Memory Score (PALFAMS)	The number of times a subject chose the correct box on their first attempt when recalling the pattern locations. Calculated across all assessed trials	Episodic memory/ Higher is better
PRM	Percent Correct Immediate (PRMPCI)	The number of correct patterns selected by the subject in the immediate forced-choice condition, expressed as a percentage	Visual pattern recognition memory/ Higher is better
PRM	Percent Correct Delayed (PRMPCD)	The number of correct patterns selected by the subject in the delayed forced-choice condition, expressed as a percentage	Visual pattern recognition memory/ Higher is better
SSP	Forward Span Length (SSPFSL)	The longest sequence of boxes successfully recalled by the subject. Applicable to Forward variants only	Working Memory Capacity/ Higher is better
SSP	Reverse Span Length (SSPRSL)	The longest sequence of boxes successfully recalled by the subject. Applicable to Reverse variants only	Working Memory Capacity/

6.7.2 PHQ-9

The PHQ-9 questionnaire will be completed at every study visit.

The mean change in PHQ-9 score from baseline to Week 56, along with the percentage of subjects with at least 1 PHQ-9 score indicative of depression of moderate or greater severity (PHQ-9 score of 10 or more), the percentage of subjects with at least 1 PHQ-9 score indicative of major depression (Kroenke), and the percentage of subjects with at least 1 positive answer to question 9 relating to suicidal ideation will be presented by treatment group. A listing of all subjects presenting with any of the outcomes listed above, or a PHQ-9 score of 15 or more will be provided.

6.7.3 C-SSRS

For each item of the C-SSRS, counts and percentages will be presented by treatment group and in total summarizing the number of yes responses at each study visit. In addition, a composite measure will be summarized at each study visit that includes the following C-SSRS items: Active suicidal ideation with some intent to act without specific plan, Active suicidal ideal with specific plan and intent, Actual attempt, Engaged in non-suicidal non-injurious behavior, Interrupted attempt, Aborted attempt, Preparatory acts or behavior, and Suicidal behavior. The baseline distribution of responses to the C-SSRS will also be presented.

6.7.4 Hand and Wrist X-Ray

Changes of bone age from baseline to Week 56 in hand and wrist X-ray will be evaluated. Differences between treatment groups will be evaluated using methods similar to those used to evaluate other continuous variables.

6.7.5 DXA

In the subset of subjects treated at sites where DXA scans are being done, mean changes from baseline to Week 56 in bone mineral density (BMD) and bone mineral content (BMC)-Z scores will be evaluated. The mean change in BMD and BMC will be summarized descriptively as a continuous variable.

6.7.6 Adverse Events

All adverse events (AEs) recorded on the eCRF will be coded using the MedDRA dictionary Version 21.1 (or a later version if updated during the study) and classified as either pre-treat AEs or treatment – emergent AEs (TEAEs) as follows:

- Pre-treat AEs are events that start prior to the date of first dose of study treatment.

- TEAEs are events with start date on or after the date and time of first dose of study treatment and up to 28 days after date and time of last dose of study treatment or events with start date and time prior to the date and time of first dose of study treatment whose severity worsens on or after the date and time of first dose of study treatment.

All AE data will be listed by treatment group. Treatment-emergence status will be flagged in the listing. In addition, corresponding listings of serious AEs (SAEs), AEs leading to discontinuation of study treatment and AEs resulting in death will be produced.

Summary tables of TEAEs by treatment group will be produced for the Safety population.

The severity of all AEs is recorded as mild, moderate, or severe. If severity is missing for a TEAE, it will be considered severe only in the overall category in the summary tables.

The relationship between an AE and study treatment is assessed as Related or Not Related by the investigator.

An overview table will summarize the number and percentage of subjects with at least one of the following TEAEs, where subjects with more than one TEAE in a particular category are counted only once in that category:

- Any TEAE
- Any TEAE by severity (mild, moderate, severe)
- Drug-related TEAE
- TEAE leading to dose reduction
- Drug-related TEAE leading to dose reduction
- TEAE leading to study treatment discontinuation
- Treatment-emergent SAE
- Drug-related Treatment-emergent SAE
- Treatment-emergent SAE leading to death;
- Drug-related Treatment-emergent SAE leading to death
- Treatment-emergent SAE leading to treatment discontinuation

The number and percentage of subjects reporting each TEAE will be summarized by System Organ Class (SOC) and Preferred Term (PT) for the Safety population. Tables will be sorted alphabetically by SOC. PTs will be sorted by descending overall total. The following summaries will be produced:

- TEAEs by SOC and PT
- Drug-related TEAEs by SOC and PT
- TEAEs by maximum severity by SOC and PT
- TEAEs related to treatment by maximum severity by SOC and PT
- TEAEs causing discontinuation from treatment by SOC and PT
- TEAEs related to treatment causing discontinuation from treatment by SOC and PT
- SAEs by SOC and PT

In the above summaries, subjects with more than one AE within a particular SOC are counted only once for that SOC. Similarly, subjects with more than one AE within a particular PT are counted only once for that PT. For summaries by maximum severity, subjects with multiple AEs within a particular SOC or PT will be counted under the category of their most severe AE within that SOC or PT.

No statistical comparisons of AEs between treatment groups will be performed.

Sponsor Name: Vivus, Inc.
Sponsor Protocol ID: OB-403

6.7.7 Laboratory Evaluations

A summary of observed values and change from baseline will be presented for all laboratory parameters with numerical measures using descriptive statistics. Shift tables displaying low-normal-high at baseline versus low-normal-high at end of study in a 3-by-3 contingency table will be provided. For selected laboratory parameters, scatter plots of baseline versus Week 56 results, will be produced by treatment group.

A laboratory value that is above or below normal range will be considered an abnormal value. For selected laboratory parameters, threshold limits of clinical concern will be defined as multiplicative factors of the normal ranges. The list of multiplicative factors for each laboratory parameter will be included in the Statistical Analysis Plan. The frequency and percentage of subjects with laboratory results above or below the normal range and threshold limits at each scheduled assessment or any time during the treatment will be summarized by treatment group.

Table 3: Clinical Laboratory Tests

Fasting blood chemistry	Hematology	Other
• albumin • alkaline phosphatase • ALT • AST • GGT • bicarbonate • blood urea nitrogen • serum calcium • serum chloride • serum sodium • creatinine (and estimated creatinine clearance) • glucose • lactate dehydrogenase • serum phosphorus • serum potassium • total and direct bilirubin • total protein • uric acid	• hemoglobin • hematocrit • red blood cell count • red blood cell indices • total white blood cell count • white blood cell differential (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) • platelet count	• thyroid stimulating hormone
	Lipid panel	**Urinalysis**
	• total cholesterol • LDL-C • HDL-C • triglycerides	• midstream urinalysis with reflex microscopic evaluation • pregnancy test (all female subjects)
		Urine Drug Screen
		• cannabinoids • amphetamines • cocaine • barbiturates • benzodiazepine • opiates
	Glycemic testing	**Serology**
	• HbA1c • insulin • glucose	• HBsAg • HCV • HIV

All laboratory data will be reported in International System of Units (SI) and Conventional units. Out-of-reference-range values will be flagged as high (H) or low (L) in the listings.

For analysis purposes, values preceded by a "<" or a ">" sign (i.e. those below or above the limits of quantification) will be considered equal to the lower or upper limit of quantification, respectively.

For analysis by visit, analysis windowing as described in Section 5.1 will be utilized for each scheduled visit such that unscheduled visits will also be considered. For each laboratory analyte, the baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of treatment.

6.7.8 Vital Signs

The following vital signs will be listed and summarized by treatment group and visit.

- Systolic and diastolic blood pressure (mmHg)

- Heart rate (breaths/min)

- Weight (kg)

- respiration rate (breaths/min)

- body temperature (°C or °F)

Vital signs data and changes from baseline in vital signs will be summarized by visit using standard descriptive statistics for the Safety population. The baseline value will be defined as last scheduled or unscheduled value collected prior to the first dose of study treatment. For post-baseline, only data from scheduled visits will be included in the summary tables.

6.7.9 Electrocardiograms

The following quantitative ECG measurements will be listed and summarized by treatment group and visit for the Safety population:

- heart rate (bpm)

- RR interval (msec)

- PR interval (msec)

- QRS interval (msec)

- QT interval (msec)

An overall clinically significant Investigator assessment of ECG will be provided (categories "Yes", "No").

6.7.10   Physical Examination

Abnormalities identified from physical examination are recorded in the eCRF will be listed.

For each physical examination body system, the number and percentage of subjects with abnormalities at baseline and post-baseline will be summarized by treatment group for the Safety population.

6.8 Interim Analysis

No interim analysis is planned.

6.9 Data Monitoring Committee

An independent Data Monitoring Committee (DMC) with multidisciplinary representation will be established to evaluate accumulating study data on a periodic basis and to assess the ongoing safety of the study for the subjects enrolled and to be enrolled. As a result, following each data review, the DMC will make a recommendation to the sponsor regarding continuation, revision, or termination of the study. Details related to DMC responsibilities, authorities, and procedures will be documented in the DMC charter, which will be finalized by the DMC prior to the first DMC data review meeting.

7. Changes in Planned Analysis

None

8. Data Issues

None

9. References

1   ICH. *Statistical Principles for Clinical Trials*, Guideline E9, 1998. Available at http://www.emea.eu.int/pdfs/human/ich/036396en.pdf

2   CPMP. *Points to Consider on Missing Data*. EMEA: London, 2001. Available at http://www.emea.eu.int/pdfs/human/ewp/177699EN.pdf

3   Phillips A and Haudiquet V. *ICH E9 guideline "Statistical principles for clinical trials": a case study*. Statistics in Medicine 2003; 22:1-11

4   Brown D J. *ICH E9 guideline "Statistical principles for clinical trials": a case study. Response to A. Phillips and V. Haudiquet.* Statistics in Medicine 2003; 22:13-17

5   Phillips A, Ebbutt A, France L, Morgan D, Ireson M, Struthers L and Heimann G. *Issues in applying recent CPMP "Points to Consider" and FDA guidance documents with biostatistical implications.* Pharmaceutical Statistics 2003; 2:241-251

6   Senn S. *Statistical Issues in Drug Development*. John Wiley & sons (Chichester), 1997.

7   Chow S-C and Liu J-P. *Design and Analysis of Clinical Trials: Concepts and Methodologies*. John Wiley & sons (New York), 1998.

8   Brown H and Prescott R. *Applied Mixed Models in Medicine*. John Wiley & sons (Chichester), 1999.

9   Fairclough D L. *Design and Analysis of Quality of Life Studies in Clinical Trials.* Chapman & Hall/CRC, 2002.

10   Green S, Benedetti J and Crowley J. *Clinical Trials in Oncology (2nd edition)*. Chapman & Hall/CRC, 2002.

11   McEntegart D. *Forced randomization when using interactive voice response systems.* Applied Clinical Trials October 2003; 50-58.

12   CPMP. *Points to Consider on Adjustment for Baseline Covaraites*. EMEA: London, 2003

13   ICH. *ICH E3 Guideline: Structure and Content of Clinical Study Reports Questions & Answers*, 2012. Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_QAs_R1_Step4.pdf

14   Modified z-scores in the CDC growth charts http://www.cdc.gov/nccdphp/dnpa/growthcharts/resources/BIV-cutoffs.pdf

15   The PHQ Validity of a Brief Depression Severity Measure http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/pdf/jgi_01114.pdf