A Phase 3 Study to Evaluate the Safety, Tolerability, and Efficacy of Naltrexone for use in Conjunction with Buprenorphine in Adults with Opioid Use Disorder Transitioning from Buprenorphine Maintenance Prior to First Dose of Vivitrol®

Unique Protocol ID:	ALK6428-A302
NCT Number:	NCT02696434
Date of Statistical Analysis Plan:	29 November 2017

STATISTICAL ANALYSIS PLAN

ALK6428-A302

Study Title:	A Phase 3 Study to Evaluate the Safety, Tolerability, and Efficacy of Naltrexone for use in Conjunction with Buprenorphine in Adults with Opioid Use Disorder Transitioning from Buprenorphine Maintenance Prior to First Dose of VIVITROL®
Document Status:	Final Version, 1.0
Document Date:	29 November 2017
Based on:	Protocol Amendment 3, dated 14 June 2017
Sponsor:	Alkermes, Inc. 852 Winter Street Waltham, MA 02451 USA

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

The following abbreviations are used in the statistical analysis plan.

Abbreviation or Term	Explanation or Definition
AE	Adverse event
ANOVA	Analysis of variance
ATC	Anatomical Therapeutic Chemical [classification system]
BUP	Buprenorphine
BMI	Body mass index
COWS	Clinical Opiate Withdrawal Scale
CI	Confidence interval
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	Electrocardiogram
EOT	End of treatment
ET	Early termination
HAM-D	Hamilton Rating Scale for Depression
MedDRA	Medical Dictionary for Regulatory Activities
MINI	Mini-International Neuropsychiatric Interview
MMSE	Mini-Mental State Examination
MOTYB	Months of the Year Backward
NTX	Naltrexone
PBO	Placebo
PBO-N	Placebo for naltrexone
PCS	Potentially clinically significant
PGART	Patient Global Assessment and Response to Therapy
QSUI	Quantitative Substance Use Inventory
QTcB	QT interval corrected using Bazett's formula
QTcF	QT interval corrected using Fridericia's formula
SAE	Serious adverse event
SAP	Statistical analysis plan
SMAST	Short Michigan Alcohol Screening Test
SOWS	Subjective Opiate Withdrawal Scale

Abbreviation or Term	Explanation or Definition
SD	Standard deviation
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal
VAS	Visual Analogue Scale
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical methods to be used for analyses and data presentation for reporting efficacy and safety results for study ALK6428-A302. This document has been prepared based on Alkermes ALK6428-A302 Study Protocol Amendment 3 (dated 14 June, 2017) [1].

1.1. Study Objective(s)

1.1.1. Primary Objective

- To evaluate the efficacy of oral naltrexone used in conjunction with buprenorphine (BUP) in adults with Opioid Use Disorder transitioning from BUP maintenance prior to the first dose of VIVITROL®

1.1.2. Secondary Objective

- To determine the safety and tolerability of oral naltrexone used in conjunction with BUP in adults with Opioid Use Disorder transitioning from BUP maintenance prior to the first dose of VIVITROL®

1.2. Summary of Study Design

This Phase 3, randomized, double-blind, placebo-controlled, parallel group study will evaluate a dosing schedule for active vs placebo oral naltrexone co-administered with BUP in BUP-dependent individuals prior to first dose of VIVITROL. Eligible subjects will be randomized in a 1:1 ratio to one of two treatment groups (naltrexone + BUP or placebo naltrexone [PBO-N] + BUP) for induction onto VIVITROL, stratified according to low (<8 mg/day) vs high (8 mg/day) BUP maintenance dose at the time of initiation of the BUP Lead-in Period. Approximately 92 subjects are planned to be randomized; 46 subjects per group.

The study participation for each subject will last approximately 9 weeks; up to 3 weeks for screening (Days -26 to -6; during which time subjects will remain under the care of their original BUP provider), approximately 2 weeks for the BUP Lead-in Period and the Treatment Period (Days -5 to 7), 4 to 5 days for VIVITROL induction and post-VIVITROL monitoring, and a 4-week outpatient follow-up period.

This study includes:

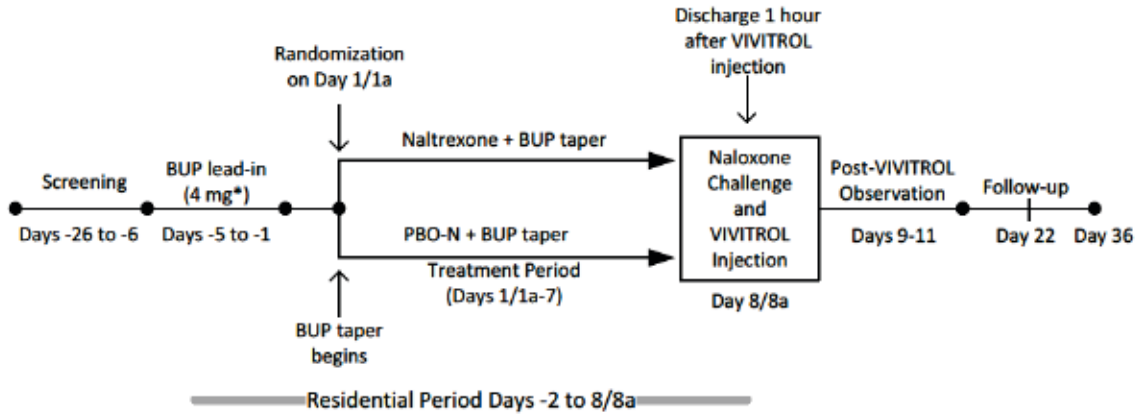
- BUP Lead-in Period: Outpatient Days -5 through -3; residential Days -2 and -1 (option for earlier residential admission at the study clinician's discretion)
- Treatment Period: Transitional dosing with oral naltrexone or PBO-N in conjunction with BUP taper; residential Days 1[1] to 7

[1] Subjects who do not qualify for randomization on Day 1 will receive Day -1 BUP dosing and repeat Day 1 assessments and procedures on the following day, Day 1a.

- VIVITROL Induction and Post-VIVITROL Observation Period
  - A naloxone challenge and administration of VIVITROL on Day 8[2] prior to discharge
  - Post-VIVITROL outpatient monitoring (Days 9-11)

A schematic of the study design is provided in Figure 1. Please refer to the final study protocol for the schedule of procedures/assessments and additional details on study conduct.

Figure 1: Study Design Schematic



*Subjects maintained on <4 mg at Day -5 will continue on their current dose until the Treatment Period taper calls for further decrease to establish a consistent daily dose prior to transitional dosing with naltrexone (see Protocol Section 8.2.1).

Evaluation of withdrawal symptoms will occur throughout the BUP Lead-in Period, the Treatment Period, and the VIVITROL Induction and Post-VIVITROL Observation Period (Days -5 through 11). On Days -2 through 8/8a, AM and PM vital signs, Clinical Opiate Withdrawal Scale (COWS)/ Subjective Opiate Withdrawal Scale (SOWS), and Visual Analogue Scale (VAS) will be measured. On Day 8/8a, PM vital signs, COWS/SOWS and VAS will be measured prior to discharge. In addition to the AM and PM assessments described above, clinical testing sessions to monitor opioid withdrawal effects in response to study drug will occur on Days 1 through 7. Also, subjects will receive psychoeducational counseling starting on Day -5 and throughout the Treatment Period.

Up to two additional inpatient visits may be conducted (Days 1a and 8a), as needed, for subjects who do not meet criteria for proceeding with procedures on Day 1 and Day 8.

[2] Subjects who do not qualify to receive VIVITROL on Day 8 will receive Day 7 study drug (naltrexone/PBO-N). Pre- and postdose COWS will be assessed for clinical safety purposes only. Day 8 assessments and procedures will be repeated the following day (Day 8a).

1.3. Criteria for Evaluation

Efficacy:

Efficacy will be assessed via the following:

- Administration of VIVITROL
- Opioid withdrawal using the COWS score
- Opioid withdrawal using the SOWS score
- Desire for opioids using a VAS
- Patient Global Assessment of Response to Therapy (PGART)

Exploratory Endpoints:

The following exploratory assessments will be evaluated:

- Pupil diameter (measured during clinical sessions)
- Quantitative Substance Use Inventory (QSUI)
- Hamilton Rating Scale for Depression (HAM-D)
- Brief Assessments of Cognition (BAC) Symbol Coding Test
- Wechsler Memory Scale-III Spatial Span (WMS-III SS) Test
- Continuous Performance Test (CPT)
- Test of Attentional Performance (TAP)

Safety and Tolerability:

Safety and tolerability will be assessed via the following:

- Adverse events (AEs)
- Vital signs and arterial oxygen saturation (measured during clinical testing sessions)
- Laboratory test results
- Electrocardiogram (ECG) parameters (uncorrected QT, QT interval corrected for heart rate using the Fridericia formula [QTcF], QT interval corrected for heart rate using the Bazett formula [QTcB], HR, RR, and QRS intervals)
- Columbia-Suicide Severity Rating Scale (C-SSRS) scores

2. SAMPLE SIZE CONSIDERATIONS

The primary efficacy endpoint of the study is the proportion of subjects who received and tolerated VIVITROL injection on Day 8/8a as demonstrated by mild opioid withdrawal symptoms (COWS ≤12 or SOWS ≤10) following VIVITROL administration.

Assuming the proportion of subjects receiving and tolerating VIVITROL administration is 90% in Group 1 and 60% in Group 2, a sample size of approximately 46 subjects per treatment group will provide at least 90% power to detect a statistically significant difference between the two treatment groups at 5% level of significance in a 2-sided test.

3. DATA ANALYSES

3.1. General Statistical Methodology

The safety and efficacy endpoints will be summarized by treatment group and overall as described below. In general, summary statistics (n, mean, standard deviation [SD], median, minimum, and maximum for continuous variables, and number and percentage of subjects in each category for categorical variables) will be provided by treatment group. All summary tables will be based on observed data, and missing values will not be imputed unless otherwise indicated. Measurements collected from unscheduled visits or repeated assessments will not be included in the by-visit summary tables or figures, but will be included in the derivation of the last post-baseline value during treatment, the analyses for the PCS post-baseline values, and subject listings. Source data for the summary tables and statistical analyses will be presented as by-subject data listings.

Baseline value is defined as the last non-missing assessment prior to the first dose of study drug.

All statistical tests and confidence intervals (CIs), unless stated otherwise, will be 2-sided and will be set at an alpha level of 0.05.

All the efficacy analyses will be summarized by the planned treatment assignment, and all the safety analyses will be summarized by actual treatment.

The data will be summarized for the following two study periods as appropriate.

- Treatment Period, which is defined as the first dose start date to the date of end of treatment (EOT). For the efficacy analysis, EOT is defined as study Day 7 or the day of early termination (ET), if a subject discontinues before Day 7. For the safety analysis, EOT is defined as Day 8/8a or the day of ET, if a subject discontinues before Day 8/8a.
- Follow-up Period, which is defined as the period between the date of the EOT visit plus 1 day and the last study visit.

3.2. Study Populations

3.2.1. Safety Population

The Safety Population will include all randomized subjects who receive at least one dose of naltrexone or PBO-N. The Safety Population will be used for both the safety and efficacy analyses.

3.2.2. Disposition of Subjects

Subjects who initiate screening but are not randomized will be considered screen failures. A randomized subject who is eligible to receive the naloxone challenge will be considered a treatment completer. Subjects completing all visits during the Follow-up Period will be categorized as completing the study.

Subject disposition will be summarized by treatment group and overall for all randomized subjects, and includes the following:

- Subjects who were randomized
- Subjects in the Safety Population
- Subjects who were treatment completers (subjects eligible to receive the naloxone challenge)
- Subjects who received the naloxone challenge on Day 8/8a
- Subjects who received the VIVITROL injection
- Subjects who completed the treatment (subject status in disposition CRF page)
- Subjects who completed the study
- Subjects who discontinued the study during the Treatment Period
- Subjects who discontinued the study during the Follow-up Period
- Reasons for discontinuation

Percentages of disposition are based on the subjects in the Safety Population. A listing of disposition will be provided for all subjects.

3.2.3. Protocol Deviations

Subjects with major protocol deviations will be summarized by treatment group and overall, along with supportive listings, for each of the following categories:

- Did not meet the inclusion/exclusion criteria
- Received prohibited medications
- Failed to adhere to the protocol:
  - Failure to remain in secure inpatient residential areas for the duration of the residential period.
  - Missed more than a single day of study medication during the lead-in period, Days -5 to -3
  - Missed more than a single day of study medication during the residential treatment period
  - Failure to perform pre-VIVITROL naloxone challenge as indicated and subject receives VIVITROL.
  - Failure to comply with study protocol regarding subjects who do not transition to VIVITROL (see section 8.3.3 for detailed instructions regarding reinstatement of BUP/Nx treatment and notification of/referral back to BUP provider). This includes (1) failure to refer those subjects who have transitioned from BUP but have not received VIVITROL to a BUP provider; and (2) failure to reinitiate BUP in these subjects prior to discharge.

- Randomization or dosing error
- Other major deviations (such as duplicate subject, or failure to discontinue dosing if gating COWS is not done or is out of range)

Duplicate subjects include subjects who are enrolled multiple times in the same study. During the conduct of this study, data (CRF and source) will be reviewed to identify duplicate subjects and sites will be contacted to investigate and confirm. Upon confirmation, duplicate subjects will be prohibited from further participation in the trial and will be discontinued from the study.

By enrolling in the study a second time, these subjects go against the spirit of the protocol which is to assess the efficacy, safety and tolerability of the ALKS 6428 regimen in a patient population that is presumably naïve to the regimen. The data collected on these subjects during their second enrollment are therefore unreliable, as their experience during their first enrollment may have impacted their outcome in the second enrollment. Because data from the second enrollment could potentially misinform and adversely influence the interpretation of the study, these data will be included in the safety analyses but excluded from the efficacy analyses.

3.3. Demographic and Baseline Characteristics

Demographics and baseline characteristics such as gender, age, race, ethnicity, weight, height and body mass index (BMI), will be summarized by treatment group and overall for the Safety Population. Other baseline characteristics such as BUP maintenance dose (<8 mg/day vs 8 mg/day), alcohol breath test results, Mini-Mental State Examination (MMSE) score, VAS/ COWS/SOWS scores, pupil diameter, HAM-D total score, Short Michigan Alcohol Screening Teset (SMAST) score, BAC Symbol Coding test score, COWA task score, WMS-III SS test score, CPT score, and TAP score will be summarized by treatment group and overall for the Safety Population.

Medical history will be summarized by treatment group and overall for the Safety Population.

Demographic, baseline characteristics and medical history listings will be provided for all subjects.

3.4. Prior, Concomitant and Ancillary Medications

Prior medications are defined as medications taken prior to the first dose of study drug. Concomitant medications are defined as medications taken on or after the first dose of study drug. All medications as documented by the Investigator will be coded using the World Health Organization Anatomical Therapeutic Chemical (WHO-ATC) drug dictionary Enhanced version March 2016. Subjects also receive standing doses of the ancillary medications (defined in Protocol Section 8.9.2) to treat symptoms of withdrawal.

Prior and concomitant medications, including ancillary medications, will be summarized by treatment group and overall for the Safety Population. For the summary tables, if a subject has taken a medication more than once, the subject will be counted only once for the medication.

All prior and concomitant medications, including ancillary medications will be listed.

3.5. Treatment Adherence Rate and Extent of Exposure

3.5.1. Treatment Adherence Rate

Study drug administration will be either directly observed (oral naltrexone/BUP and PBO-N/BUP) or administered (VIVITROL injections) by clinical study staff. Therefore, the treatment adherence rate will be derived directly using the following formula:

Adherence rate = number of days with study drug taken ÷ number of days in attended in Treatment Period ×100.

Any deviations or treatment non-adherence issues will be recorded by the person observing or administering the treatment. Treatment adherence rate will be summarized for the Safety Population by treatment group and overall for the Treatment Period (Days 1 –7). In addition, the number and percentage of subjects who completed one day, two days and up to seven days of treatment will be summarized and tabulated.

3.5.2. Extent of Exposure

Exposure (days) to study drug is defined as the number of days from the date the first dose of study drug was taken to the date the last dose was taken, inclusive (ie, last dose date – first dose date + 1) during the Treatment Period (Days 1 –7). Exposure to study drug will be summarized by treatment group and overall for the Safety Population using descriptive statistics (number of subjects, mean, SD, minimum, median, and maximum).

3.6. Efficacy Analyses

3.6.1. General Considerations

Efficacy analyses will be performed using the Safety Population.

All statistical tests will be 2-sided with a type I error rate of 5%, unless otherwise specified. All confidence intervals will be 2-sided 95% confidence intervals.

3.6.2. Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of subjects who received and tolerated VIVITROL injection on Day 8/8a as demonstrated by mild opioid withdrawal symptoms (COWS ≤12 or SOWS≤10) following VIVITROL administration.

A two-part naloxone challenge will be administered on Day 8. Subjects with a negative naloxone challenge (ie, not physically dependent on opioids) will then receive a VIVITROL injection. Subjects who have a positive naloxone challenge on Day 8 (ie, still physically dependent on opioids) will repeat Day 7 dosing and assessments and then take the challenge again the next day (Day 8a).

Subjects in the Safety Population, who meet the following two criteria, will be included in calculating the proportion of subjects who received and tolerated a VIVITROL injection:

1. Subject receives a VIVITROL injection on Day 8/8a.

2. After the VIVITROL injection on Day 8/8a, the subject's opioid withdrawal symptoms are mild (COWS≤12 or SOWS ≤10). The 1-hour post-dose VIVITROL COWS/SOWS assessment on Day 8/8a will be used for this evaluation.

The numerator is the number of subjects in the Safety Population who meet the two criteria above, and the denominator is the number of subjects in the Safety Population. The primary efficacy endpoint will be analyzed using a logistic regression model with treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day versus 8 mg/day) as factors.

The proportions, difference in proportions (NTX/BUP vs PBO-N/BUP), P-value and odds ratio with 95% CI will be summarized.

In addition, the COWS and SOWS scores 1 hour post-VIVITROL on Day 8/8a will be summarized (n, mean, standard deviation, median, minimum, maximum, COWS category [≤12 or >12], and SOWS category [≤10 or >10]) by treatment group and overall for the Safety Population. The number of subjects who receive a VIVITROL injection will also be summarized.

3.6.3. Secondary Efficacy Endpoints

The secondary efficacy endpoints are defined as follows:

- Mean scores for "desire for opioids" VAS during the Treatment Period (Days 1/1a-7); VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11); and the entire study (Days 1/1a-End of Study)
- Number of days with COWS peak score ≤12 during Treatment Period prior to the VIVITROL injection (Days 1/1a-7)
- Number of post-VIVITROL days (Days 9-11) in which subjects in each group demonstrate mild (COWS ≤12) opioid withdrawal
- Mean peak COWS scores during the Treatment Period (Days 1/1a-7); and VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11)
- Area under the curve (AUC) for COWS scores during the Treatment Period (Days 1/1a-7); and VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11)

The mean VAS score during the Treatment Period (Days 1/1a-7) for each subject in the Safety Population will be derived as:

$$\text{Mean VAS Score} = \frac{\sum \text{ VAS Score}}{\text{\# of VAS assessments during Days 1/1a through 7}}$$

The mean VAS score during VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11), and the entire study (Days 1/1a-End of Study) will be derived using the method similar as the mean VAS score during the Treatment Period (Days 1/1a-7).

The VAS score and change from baseline during the Treatment Period, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Periods (Days 1/1a-11) will be summarized by treatment group and overall for the Safety Population, where baseline VAS score is defined as the most recent visit prior to Day 1.

The COWS will be administered 4-6 times per day during the Treatment Period (Day 1/1a – Day 7). The COWS score is the total score for all the item scores. On each day during the Treatment Period, the highest COWS score among the 4-6 COWS administered will be identified as the daily peak COWS score. The number of days during the Treatment Period will be counted where the daily peak COWS score is ≤12. A negative binomial model will be used for the analysis of the secondary endpoint, number of days with COWS peak score ≤12 during Treatment Period prior to the VIVITROL injection (Days 1/1a-7). The model will include the number of days with COWS peak score ≤12 as the response variable, the treatment group and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as predictors in the analysis. The odds ratio, standard error, 95% CI for the odds ratio, and P-value will be summarized.

Number of days with COWS peak score ≤12 during the Treatment Period will be summarized (n, mean, standard deviation, median, minimum and maximum) by treatment group and overall for the Safety Population.

The number of post-VIVITROL days (Days 9-11) with COWS peak score ≤12 will be analyzed. A negative binomial model will be used, with the number of post-VIVITROL days (Days 9-11) with COWS peak score ≤12 as the response variable. The model will include the treatment group and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as predictors in the analysis. The odds ratio, standard error, 95% CI for the odds ratio, and P-value will be summarized.

The daily peak COWS score for each subject is derived based on the algorithm described above for each treatment day. The mean peak COWS score during the Treatment Period (Days 1/1a-7) for each subject is derived as:

$$\text{Mean Peak COWS Score} = \frac{\sum \text{Daily Peak COWS Score}}{\text{Number of Days with daily Peak COWS score}}$$

The daily peak COWS score during VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11), and the entire study (Days 1/1a-End of Study) will be derived using the method similar as daily peak COWS score during the Treatment Period (Days 1/1a-7).

The total AUC COWS score for each subject is derived based on the actual time (in minutes) the COWS is administered on each day during the Treatment Period (Days 1/1a-7) by using the linear trapezoidal rule. The total AUC COWS score during the treatment period (Days 1/1a-7) is the summation of daily AUC COWS scores during Day 1/1a through Day 7. Then, the normalized AUC COWS score for each subject during Day 1/1a through Day 7 is derived as:

$$\text{Normalized AUC COWS Score} = \frac{\sum \text{Daily AUC COWS Score}}{\text{Number of Days with daily AUC COWS score}}$$

The normalized AUC COWS score during VIVITROL Induction and Post-VIVITROL Observation Period (Days 8/8a-11), and the entire study (Days 1/1a-End of Study) will be derived using the method similar as the normalized AUC COWS score during the Treatment Period (Days 1/1a-7).

The mean peak COWS score, normalized AUC COWS score and mean VAS score endpoints will be analyzed using an analysis of covariance (ANCOVA) model. The model will include

treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as factors, and corresponding baseline value as a covariate. The least squares (LS) mean, standard error (SE), and LS mean difference between (NTX/BUP vs. PBO-N/BUP) along with the 95% CI, will be summarized.

The mean peak COWS score, normalized AUC COWS score and mean VAS score during the Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Period (Days 1/1a-11) will be summarized by treatment group and overall for the Safety Population. A data listing for mean peak COWS scores, normalized AUC COWS scores, and mean VAS scores will be provided.

In addition, daily peak COWS score, daily average COWS score, daily average SOWS score and daily average VAS score during the Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Period (Days 1/1a-11) will be summarized (n, mean, standard deviation, median, minimum and maximum) by treatment group and overall for the Safety Population. Data listings for COWS and SOWS scores at each study visit assessment, including the item scores, will also be provided.

3.6.4. Exploratory Efficacy Endpoints

The exploratory endpoints are defined as follows:

- Number and proportion of subjects in PGART response category at the end of the Post-VIVITROL Observation Period (Day 11)
- Change from baseline in pupil diameter following study drug administration during the Treatment Period (Days 1/1a-7)
- Change in frequency of substance use from screening to Day 36 assessed via the QSUI
- AUC for SOWS scores during the Treatment Period and VIVITROL Induction and Post-VIVITROL Observation Period (Days 1/1a-11)
- Change from baseline in standardized T scores for cognitive assessments
    - The BAC Symbol Coding test,
    - COWA task,
    - WMS-III SS test
    - CPT
    - TAP

The PGART will be used to assess treatment satisfaction. The PGART consists of one item: "How would you rate your response to the medication that you received during the study?" It is measured on a 5-point Likert scale (0 = Poor, 1 = Fair, 2 = Good, 3 = Very Good, 4 = Excellent).

The PGART response will be categorized as a binary variable (< 2 and ≥ 2). The categorized PGART response for Day 11 will be analyzed using the logistic regression model with treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as factors.

The descriptive statistics for the PGART response at each assessment visit will also be summarized by treatment group and overall for the Safety Population. A data listing for PGART at each study visit assessment will also be provided.

The mean change from baseline in pupil diameter for each day during the treatment period (Days 1/1a to 7) will be analyzed using an ANCOVA model. The model will include treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day versus 8 mg/day) as factors, and corresponding baseline value as a covariate. The least squares (LS) mean, standard error (SE), and LS mean difference between (NTX/BUP vs PBO-N/BUP) along with the 95% CI, will be summarized. In addition, the pupil diameter will be descriptively summarized by treatment group and overall at each assessment visit for the Safety Population. A data listing at each study visit assessment will be provided.

The change in frequency of substance use from screening to Day 36 will be assessed using the Quantitative Substance Use Inventory (QSUI) Question 5a (the number of days using any illicit drugs). The baseline value of Question 5a will be subtracted from the Day 36 value of Question 5a. This change in number of days using any illicit drugs (Question 5a) at Day 36 will be analyzed using an ANCOVA model with treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as factors, and corresponding baseline value as a covariate. The least squares (LS) mean, standard error (SE), and LS mean difference between (NTX/BUP vs. PBO-N/BUP) along with the 95% CI, will be summarized. In addition, each question of the QSUI scale will be descriptively summarized by treatment group and overall at each assessment visit for the Safety Population. A data listing at each study visit assessment will be provided.

The AUC for SOWS scores during the Treatment Period and VIVITROL Induction and Post-VIVITROL Observation Period (Days 1/1a-11) will be derived and analyzed using the ANOVA model similar to the method used for the total AUC for COWS scores.

Standardized T scores will be calculated for each of the cognitive parameters at each visit assessment. The change in standardized T score from baseline to Day 22 and Day 36 will be analyzed using an ANCOVA model respectively, with treatment and randomization stratification factor of BUP maintenance dose (<8 mg/day vs 8 mg/day) as factors, and corresponding baseline value as a covariate. In addition, each parameter will be descriptively summarized at each visit assessment by treatment group and overall for the Safety Population. A data listing at each study visit assessment will be provided.

The following graphical displays will be presented by treatment group:

- Plot of Mean (SE) Daily Peak COWS Score during the Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Period.
- Plot of Mean (SE) Daily Average COWS Score during the Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Period.
- Plot of Mean (SE) Daily Average SOWS Score during the Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-VIVITROL Observation Period.
- Plot of Mean (SE) Daily Average VAS score during Treatment Period prior to the VIVITROL injection, during the VIVITROL Induction Period, and during the Post-

VIVITROL Observation Period.

3.7. Safety Analyses

The safety analysis will be carried out using the Safety Population. Safety assessments will be summarized using descriptive statistics along with supportive listings based on observed values. Listings will be provided for all safety parameters.

3.7.1. Adverse Events

Adverse events will be coded by System Organ Class and Preferred Term using the Medical Dictionary for Regulatory Activities (MedDRA®) Version 19.1. The verbatim term will be included in the AE listings.

An AE will be considered as treatment emergent AE (TEAE) if the event starts or worsens on or after the date of first dose of study drug. For the determination of the TEAEs during the Treatment Period, AEs with the greatest severity before the baseline will be used as the benchmark for comparison with the AEs occurring during the Treatment Period. The Treatment Period is defined as the first dose start date to date of EOT, inclusive. The TEAE will be summarized by treatment group and overall for Safety Population.

Newly-emergent AEs (NEAEs), defined as AEs that started or worsened after the date of the EOT visit, will also be summarized. The AEs with the greatest severity on or before the date of the EOT visit will be used as the benchmark for the comparison of the AEs occurring during the Follow-up Period. The Follow-up Period is defined as the period between the date of the EOT visit plus 1 day and the last study visit, inclusive.

All AEs will be listed by subject but only TEAEs, serious AEs (SAEs), AEs leading to study discontinuation, and NEAEs will be included in the summary tables. Summary tables will be provided for the Treatment Period and Follow-up Period by treatment group and overall for the Safety Population. The AE leading to study discontinuation and SAEs leading to death, will be summarized in the period when the discontinuation or death occurred.

An overview table, including number of subjects with TEAEs, AEs leading to study discontinuation, SAEs and study drug related TEAEs will be provided.

The following summary tables will be produced for the Treatment Period for the Safety Population:

- TEAEs by System Organ Class and Preferred Term
- TEAEs experienced by ≥5% of subjects
- TEAEs by System Organ Class, Preferred Term, and Severity
- Drug-Related TEAEs by System Organ Class, and Preferred Term

In addition, the AE overview table and AEs by System Organ Class and Preferred Term for NEAEs during the Follow-up Period will be provided.

All AE tables will be sorted by System Organ Class and then Preferred Term in decreasing frequency based on the "all subjects" column.

A subject having the same AE (as determined by the coded MedDRA preferred term) more than once will be counted only once in the number and percentage of subjects' calculation for that AE. Similarly, if a subject had more than one AE in a System Organ Class, the subject will be counted only once in the total number of subjects with an AE for that System Organ Class. If a subject had the same AE on multiple occasions, the highest severity (severe > moderate > mild) recorded for the event will be presented in the AE by severity summary. Similarly, if a subject had the same AE on multiple occasions, the closest relationship (related > not related, where related includes definitely related, probably related, and possibly related; and not related includes probably not related and definitely not related) recorded for the event will be presented in the AE by relationship summary.

All AEs will be included in the listings. The number and percentage of subjects with at least one SAE will be summarized by System Organ Class and Preferred Term for the Safety Population. Adverse events leading to discontinuation will be summarized similarly. Supporting listings of SAEs and AEs leading to study discontinuation will be provided. Subjects who died during the study will also be listed.

3.7.2. Clinical Laboratory Parameters

Laboratory parameters will be presented in conventional units. Only scheduled laboratory parameters will be included in the summaries. All laboratory data, including those collected at unscheduled visits, will be included in the listings. Laboratory results (baseline and change from baseline) for chemistry and hematology parameters for each visit during the entire study will be summarized by treatment group and overall for the Safety Population.

For urinalysis, the number and percentage of subjects experiencing abnormalities at any post-baseline visit will be summarized. In addition, the number and percentage of subjects with values considered potentially clinically significant (PCS) occurring at any post-baseline visit for selected parameters will be summarized by treatment group and overall for the Safety Population.

Clinical laboratory test values, scheduled or unscheduled, will be considered PCS if they meet the criteria listed in Table 1. The percentage will be calculated where the numerator is the number of subjects in the Safety Population with non-PCS values at baseline and PCS values at any postbaseline assessment; and the denominator is the number of subjects in the Safety Population with non-PCS values at baseline and at least one post-baseline assessment. A supportive listing will present all values of a parameter for subjects with at least one PCS value for that parameter.

The following graphical displays will be presented by treatment group:

- Mean (SE) change from baseline over time for selected chemistry and hematology

laboratory parameters.

Pregnancy and drug test data will be listed.

Positive opioid urine drug screen results will be summarized for each visit during the entire study by treatment group and overall for the Safety Population. Opioid urine drug screen tests include Methadone, Opiate, Buprenorphine and Oxycodone. Any positive testing results in these categories will be counted as "Positive". Baseline for the opioid urine drug screen is defined as last non-missing record before dose start date/time, which includes urine drug screen tests from either the screening visit, BUP Lead-in Visit or V5 Day 1 Visit.

Table 1: PCS Criteria for Laboratory Parameters

Category	Parameter	Criteria
Hematology	Eosinophils	$>1.0\times10^3$/uL
	Hematocrit	≤32% (Female) ≤37% (Male)
	Hemoglobin	≤9.5 g/dL (Female) ≤11.5 g/dL (Male)
	Leukocytes	$\leq2.8\times10^3$/uL $\geq16\times10^3$/uL
	Lymphocytes	$<1.0\times10^3$/uL
	Neutrophils, Absolute	$<1.5\times10^3$/uL
	Platelets	$<75.1\times10^3$/uL $\geq700\times10^3$/uL
Chemistry	Alanine Aminotransferase (U/L)	≥3 × ULN
	Albumin	<2.5 g/dL
	Alkaline Phosphatase (U/L)	≥3 × ULN
	Aspartate Aminotransferase (U/L)	≥3 × ULN
	Blood Urea Nitrogen (BUN)	>30 mg/dl
	Creatine Kinase (U/L)	≥3 × ULN
	Creatinine	≥2 mg/dL
	Gamma Glutamyltransferase	≥3 × ULN
	Glucose	<50 mg/dL >200 mg/dL
	Lactate Dehydrogenase (U/L)	>3 × ULN
	Chloride	≤90 mmol/L ≥118 mmol/L
	Potassium	<3 mmol/L >5.5 mmol/L

Category	Parameter	Criteria
	Sodium	<130 mmol/L >150 mmol/L
	Thyroid stimulating hormone (TSH)	>5.5 uIU/mL
	Total Bilirubin	≥2.0 mg/dL
Urinalysis	Glucose	at least 2+
	Protein	at least 2+

3.7.3. Vital Signs and Oxygen Saturation

Vital signs and oxygen saturation for each visit during the entire study will be summarized by treatment group and overall for the Safety Population. All vital signs and oxygen saturation data will be presented in the subject data listings.

Number and percentage of subjects with vital sign values considered PCS (see Table 2) occurring at any post-baseline visit will be summarized by treatment group and overall for the Safety Population. The numerator is the number of subjects in the Safety Population with non-PCS values at baseline and PCS values at any post-baseline assessment; and the denominator is the number of subjects in the Safety Population with non-PCS values at baseline and at least one post-baseline assessment. A supportive listing will present all values of a parameter for subjects with at least one PCS value for that parameter.

In addition, mean change from baseline along with the corresponding SE will be summarized by visit and treatment group for systolic blood pressure, diastolic blood pressure, and heart rate.

The following graphical displays will be presented by treatment group:

- Mean value and mean change from baseline along with the corresponding SE for vital sign and oxygen saturation parameters

Table 2: PCS Criteria for Vital Sign and Oxygen Saturation Parameters

Parameter	PCS Criteria
Systolic blood pressure	Low: ≤ 90mm Hg and decrease ≥ 20 mm Hg
	High: ≥180 mm Hg and increase ≥ 20 mm Hg
Diastolic blood pressure	Low: ≤ 50mm Hg and decrease ≥ 15 mm Hg
	High: ≥105 mm Hg and increase ≥ 15 mm Hg
Heart Rate	Low: ≤ 50 bpm and decrease ≥ 15 bpm
	High: ≥120 bpm and increase ≥ 15 bpm
Oxygen Saturation	< 93%

3.7.4. 12-Lead Electrocardiogram (ECG)

Electrocardiogram parameters (heart rate, RR interval, PR interval, QRS interval, QT interval, QTcB [QT interval corrected using Bazett's method], QTcF [QT interval corrected using

Fridericia's method]) will be summarized for each visit during the entire study by treatment group and overall for the Safety Population.

Number and percentage of subjects with QTcB or QTcF parameter values considered PCS occurring at any post-baseline visit will be summarized by treatment group and overall for the Safety Population. Criteria for PCS values are listed in Table 3 and will be presented for each parameter. A subject will be counted only once in the highest category for a given parameter based on the largest post-baseline value. The percentages will be calculated relative to the number of subjects in the Safety Population with baseline values ≤450 msec and at least one post-baseline assessment.

A supportive listing will present all values of a parameter for subjects with at least one PCS value for that parameter.

Table 3: PCS Criteria for QTcB and QTcF

ECG Parameter	PCS Criteria
QTcB, QTcF	>500 msec, >480 - ≤500 msec >450 - ≤480 msec
QTcB, QTcF	Change from baseline of >60 msec Change from baseline >30 to ≤60 msec

3.7.5. Columbia-Suicide Severity Rating Scale

The C-SSRS is a questionnaire used to measure the presence and intensity of suicidal behavior and ideation.

Suicidal behavior and suicidal ideation will be summarized by treatment group and overall for the Safety Population. The proportion of subjects who meet the criterion for each of these categories will be summarized, as described in Table 4. A listing of results at all visit assessments will be provided.

Table 4: C-SSRS Categories for Analysis

Category	C-SSRS Item response is "YES"
Suicidal behavior	• Preparatory acts or behavior • Aborted attempt • Interrupted attempt • Actual attempt • Complete suicide

Suicidal ideation	• Wish to be dead • Non-specific active suicidal thoughts • Active suicidal ideation with any methods (not plan) without intent to act • Active suicidal ideation with some intent to act, without specific plan • Active suicidal ideation with specific plan and intent

3.7.6. Other Questionnaires

Questionnaire data collected during the study (ie, MMSE, Months of the Year Backward (MOTYB), SMAST, HAM-D will be descriptively summarized at each assessment visit by treatment group and overall for the Safety Population. A supportive listing for each questionnaire will be provided.

3.8. Pharmacokinetic/ Pharmacodynamics Data Analysis

No pharmacokinetic/pharmacodynamics data analysis is planned for this study.

4. INTERIM ANALYSIS

No interim analysis is planned for this study.

5. CHANGES IN CONDUCT OR PLANNED ANALYSES FROM PROTOCOL

The key secondary efficacy endpoint was changed to one of the secondary efficacy endpoints and was redefined as "number of days with COWS peak score ≤12 during the Treatment Period prior to the VIVITROL injection". A negative binomial regression model will be used to analyze this secondary endpoint.

The first secondary efficacy endpoint was changed to Mean score for "desire for opioids" VAS during the Treatment Period and VIVITROL Induction and Post-VIVITROL Observation Period (Days 1/1a-11).

Another secondary efficacy endpoint was changed to "number of post-VIVITROL days (Days 9-11) in which subjects in each group demonstrate mild opioid withdrawal (COWS ≤12)". A negative binomial regression model will be used to analyze this secondary endpoint.

6. DEFINITIONS AND CONVENTIONS FOR HANDLING OF DATA

6.1. Analysis Visit Windows

Dataset specifications will be provided in a separate document.

6.2. Handling of Partial Dates of Concomitant Medication

Partial start dates of prior and concomitant medications will be assumed to be the earliest possible date consistent with the partial date. Partial stop dates of prior and concomitant medications will be assumed to be the latest possible date consistent with the partial date. In the case of a completely missing stop date, medication will be assumed to be ongoing.

6.3. Safety Data Handling

All efforts should be made to obtain the missing information from the Investigator. For C-SSRS, vital signs, laboratory testing (chemistry, hematology, urinalysis), and 12-lead ECGs, only observed data will be used for analyses, and missing data will not be imputed.

7. GENERAL STATISTICAL METHODOLOGY

7.1. Reporting Precision

Summary statistics will be presented to the following degree of precision, unless otherwise specified:

Table 5: Degree of Precision

Statistics	Degree of Precision
Mean, Median, Confidence limit boundaries	One more than the raw data, up to three decimal places
Standard deviation, Standard error	One more than the mean, up to three decimal places
Minimum, Maximum	The same as the raw data, up to two decimal places
P-value	Rounded to three decimal places and therefore presented as 0.xxx; P-values smaller than 0.001 as "<0.001"; P-values greater than 0.999 as ">0.999"
Percentage	One decimal place. A percentage of 100% will be reported as 100.0%. Percentages of zero will be reported as 0

Fractional numeric values will be presented with a zero to the left of the decimal point (for example, 0.12 – 0.30).

For weight, height, and BMI, one decimal place will be used for summary statistics.

8. REFERENCE

1. Alkermes ALK6428-A302 Study Protocol Amendment 3 (dated 14 June 2017).