STATISTICAL ANALYSIS PLAN

Protocol Number: AS140026-A5

Kappa Opioid Receptor Antagonism for the Treatment of Alcohol Use Disorder and Comorbid PTSD

SAP VERSION: V4.0

SAP DATE: August 2019

PI/SPONSOR: Lori Davis, MD
Ismene Petrakis, MD

PREPARED BY: RTI International.
3040 Cornwallis Rd
Research Triangle Park, NC 27709-2104

AUTHOR (S): Nathan Vandergrift, PhD
Shawn Hirsch, MPH

Version #	Implemented By	Revision Date	Approval Date	Reason
1.0	Shawn Hirsch	N/A	N/A	First version in Draft Protocol Amendment Version 7 is pending approval
2.0	Shawn Hirsch	02/23/2022	N/A	1. Changed the Fear Potentiated Startle (FPS) outcome data from secondary to exploratory outcome 2. Added further clarification for the TLFB derived outcome measures
3.0	Shawn Hirsch	03/29/2022	N/A	Removed evaluation of the FPS assessment from this primary SAP; those analyses will be detailed in a secondary supplemental SAP, as needed
4.0	Lori Davis and Shawn Hirsch	04/19/2022	N/A	1. For the primary outcome definition, changed a positive response for PTSD from a ≥ 8 point CAPS-5 decrease from baseline to ≥ 10 point decrease from baseline; 2. Clarified that the AUD outcome measure is the average number of drinks consumed over a 28 consecutive day look-back period (and not based on *drinking days*) 3. Added change in PeTH value from baseline to Weeks 8 and 12 to exploratory outcomes

The undersigned acknowledge they have reviewed the **Statistical Analysis Plan** and agree with the approach it presents. Changes to this **Statistical Analysis Plan** will be coordinated with and approved by the undersigned or their designated representatives.

Signature:		Date:	
Print Name:	Lori Davis		
Title:	Study Clinical Lead		
Role:	Principal Investigator		

Signature:		Date:	
Print Name:	Nathan Vandergrift		
Title:	Senior Statistician		
Role:	Senior Statistician/Alternate-PI		

Signature:		Date:	
Print Name:	Shawn Hirsch		
Title:	Statistician		
Role:	Lead Statistician		

Contents

LIST OF ABBREVIATIONS

Abbreviation:	Deciphered:
ADS	Alcohol Dependence Scale
AE	Adverse event
AUD	Alcohol Use Disorder
BrAC	Breathalyzed alcohol concentration
BIS	Barratt Impulsivity Scale
BUP	Buprenorphine
CAPS-5	Clinician Administered PTSD Scale for DSM-5
CGI-S/CGI-I	Clinical Global Impression-Severity & Clinical Global Impression-Improvement
CIWA-Ar	Clinical Institute Withdrawal Assessment of Alcohol Scale, Revised
COWS	Clinical Opiate Withdrawal Scale
CREEF	Connecticut Research & Education Foundation
CRC	Clinical Research Coordinator
C-SSRS	Columbia-Suicide Severity Rating Scale
DAQ	Desire of Alcohol Questionnaire
DoD	Department of Defense
DSM-5	Diagnostic and Statistical Manual-5th Edition
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
FDA	Food and Drug Administration
FPS	Fear-Potentiated Startle
GSC	Government Steering Committee
H&P	History and Physical
HR	Heart rate
IM	Intramuscular
IND	Investigational New Drug Application
IRB	Institutional Review Board
KOR	Kappa opioid receptor
LEC-5	Life Events Checklist for DSM-5
MINI-5	Mini-International Neuropsychiatric Interview for DSM-5
MMRM	Mixed-Effect Model of Repeated Measure
msec	Millisecond
N/A	Not Applicable
NTX	Naltrexone
OCDS	Obsessive Compulsive Drinking Scale
OSU-TBI	Ohio State University Traumatic Brain Injury Identification Method (Short Form)
PASA	Pharmacotherapies for Alcohol and Substance Abuse Consortium
PCL-5	PTSD Checklist for DSM-5
PEth	Phosphatidyl ethanol
PHQ-9	Patient Health Questionnaire
PI	Principal Investigator
PLC	Placebo
PTSD	Post-Traumatic Stress Disorder
RCT	Randomized Controlled Trial
SAE	Serious adverse event

SIP	Short Inventory of Problems
SL	Sublingual
SNRI	Serotonin and norepinephrine reuptake inhibitors
SSRI	Selective serotonin reuptake inhibitors
T	Telephone visit
TLFB	Timeline Follow Back
TRD	Treatment Resistant Depression
TREAC	Tuscaloosa Research & Education Advancement Corporation
TSH	Thyroid Stimulating Hormone
TVAMC	Tuscaloosa VA Medical Center
VA	Veteran's Administration
VACT	VA Connecticut Healthcare System
VR-12	Veterans RAND 12-item Health Survey
WHO	World Health Organization
XR	Extended release
XR-NTX	Extended-release injectable naltrexone

1 BACKGROUND AND PROTOCOL HISTORY

Alcohol use disorder (AUD) is highly prevalent in U.S. service members and among military Veterans and has a large detrimental impact on society. AUD affects 18 million Americans. One important comorbid condition for individuals with AUD is PTSD. PTSD has a past year and lifetime prevalence of 4.7% and 6.1%, respectively. The odds of having PTSD are 30% greater for those with a lifetime AUD than those with no lifetime AUD. Taken from the other direction, those with lifetime PTSD are at greater risk of having an AUD (20% increased risk) compared to those without PTSD. Comorbidity is associated with many detrimental outcomes including significant social instability, disability, and more severe symptoms, higher rates of relapse, suicidal behaviors, and medical complications. , Currently, there are only four FDA–approved medications (disulfiram, oral and long-acting injectable naltrexone, and acamprosate) to treat AUD and two (sertraline and paroxetine) to treat PTSD. Despite a growing body of research in this area, there is no medication with clear evidence of efficacy in AUD-PTSD comorbidity and there are no FDA-approved medications to treat these disorders when they co-occur. New treatment strategies are urgently needed. This study addresses the substantial deficit in pharmacologic treatments for comorbid PTSD and AUD.

The use of a medications that result in KOR antagonist effects represents a novel potential treatment for Veterans and Service Members with comorbid AUD and PTSD. Endogenous opioid systems in the brain are involved in regulation of mood, stress modulation, and cravings. Kappa opioid receptors (KOR) are densely localized in limbic and cortical areas comprising the brain reward system, which play a role in modulating stress and in promoting addictive behaviors. Endogenous dynorphins, called k-selective opioid peptides, are released during the stress response and contribute to the anxiogenic and dysphoric responses to the stressful experience. Further, the KOR dynorphin system impacts stress-induced drug and alcohol seeking behavior. In animal models of alcohol dependence, KORs are implicated in the excessive alcohol consumption, particularly during withdrawal. KOR antagonists block the actions of endogenous dynorphins and block the stress-induced reinstatement of extinguished cocaine- and ethanol-seeking behaviors in animals. Studies for animal models of cocaine dependence have shown promise for using a combination of buprenorphine and naltrexone, without producing opioid dependence.

KOR antagonists may be beneficial in the treatment of addictions, PTSD, and major depressive disorder. An example includes a short-acting potent KOR antagonist with selectivity over other

opioid receptors (CERC-501) which has demonstrated antidepressant-like effects in animal models of depression and been shown to reduce ethanol self-administration in alcohol-preferring rats. CERC-501 is currently unavailable for additional studies due to its being purchased by another pharmaceutical company that is pursuing its use in major depression as the first approved indication.

Another medication that acts as a net KOR antagonist is ALKS-5461, which is a combination drug formulation of buprenorphine and samidorphan. Buprenorphine acts as an antagonist at kappa and partial agonist of the mu receptors and samidorphan is a mu opioid receptor antagonist, resulting in a pharmacologically net effect of kappa-opioid receptor antagonism (i.e. the same mechanism of action as the combination approach proposed in this application). The combination drug binds with high affinity to opioid receptors with low net intrinsic signaling activity. ALKS-5461 is intended to support opioid tone in the brain regions with impaired endogenous activity and dampen opioid tone in upregulated regions. In a recent study, adjunctive ALKS-5461 significantly reduced depression scores compared to placebo in patients with treatment resistant depression. Following a positive phase II trial, ALKS-5461 was granted Fast Track Designation by the FDA for treatment resistant depression (TRD) in October 2013. In 2014, two of three core phase II trials showed that ALKS-5461 at daily doses of 0.5mg/0.5mg and 2mg/2mg were safe and well-tolerated but, disappointingly, the trials failed to meet their primary efficacy endpoints, due in part to an unusually strong placebo effect. However, some endpoints trended towards efficacy for 2mg/2mg/d and a statistically significant signal on the depression endpoint. In a sequential parallel comparison design study (stage 1 double-blind placebo-controlled parallel comparison with higher proportion of patients randomized to placebo than to active drug and those patients who meet placebo nonresponse are re-randomized in stage 2 to drug or placebo), the 2mg/2mg and 8mg/8mg doses of ALKS 5461 were tested in patients with TRD.20 The 2/2 dose was significantly better that placebo across the three depression outcomes. Although there was significant improvement in the 8mg/8mg dosage, it did not achieve statistical significance. The medication was well tolerated and there was no evidence of opioid withdrawal on treatment discontinuation. Three additional Phase 3 studies showed efficacy ALKS-5461 2mg/2mg in subjects with TRD maintained on their SSRI, SNRI or bupropion.

<u>Protocol History:</u>

- V(2):
    - Eliminated references to ‘sponsor’ since this is not an IND study
    - Added inclusion criteria: Must have a CIWA-Ar score of < 8 prior to randomization.
    - Defined that: ‘Current diagnosis of moderate or severe non-alcohol substance use disorder (except for caffeine and nicotine) during the preceding 1 month’, would be determined per PI discretion, based on participant screening interview.
    - Removed FIBSER assessment.
    - Updated eligibility and medication b/on decisions re: allowable concomitant medications.
    - Updated inclusion and exclusion wording
    - Updated based on Alkermes SAE reporting requirements
- V(3):
    - Updated exclusion criteria #2 and #4
    - Revised incentive payment wording
    - Clarified that research pharmacist will dispense buprenorphine
    - Revised which MINI modules are required for completion
    - Added military sexual trauma software to FPS
    - Clarified SAE reporting items; AE definition; and added study halting rules
- V(4):

  - Updated site information from Atlanta (AREEF) to Detroit (Wayne State University)
  - Added Birmingham VA as a satellite site for the Tuscaloosa VA
  - Simplified exclusion criteria to remove redundant wording about SUD
  - Clarified a visit will be cancelled if urine drug screen is positive for opiates or breathalyzer > 0.02 at any visit
  - Clarified C-SSRS can be done at screening or baseline
  - Removed CG-I and CG-S assessments at Weeks 1, 2, 6, and 10
  - Removed gabapentin (Neurontin), Abilify (aripiprazole), melatonin, and beta-blockers from 'unallowable' medication list
- V(5):
  - **Removed the high dose treatment arm (8 mg of SL-Buprenorphine)**
  - **Allow non-Veterans to be enrolled.**
  - Changed exclusion criteria to QTcF >500 msec on ECG
- V(6):
  - Revisions to address safety precautions to re-open study enrollment:
    - Added section 7.3.7 (COVID-19 Safety Procedures)
    - Psychophysiology assessment put on-hold
    - In-person visits required at screening/baseline, Week 4, Week 8, and Week 12 (due to required labs or med injections)
    - Telehealth visits allowed at Week 1, Week 2, Week 6, Week 10, and Week 14
      - Specific assessments not conducted over telehealth visits include vitals, BrAC, and CIWA and COWS (specific questions)
- V(7):
  - Expand recruitment area for Tuscaloosa to include Central Alabama VA Healthcare System (CAVAHS)
  - Updated to 37-month study period that includes 32-month enrollment period
  - Restarted the psychophysiological assessment, but is optional (were on-hold due to COVID)
  - Closed enrollment at the Wayne State/Detroit site
  - Updated prohibited medication list to exclude Lamotrigine, Quetiapine, and Atomoxetine
  - Revised I/E criteria (ALT and AST now allowed to be >5 ULN)

2 PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analysis to be performed to test the therapeutic actions of kappa-opioid receptor antagonism. We propose a proof-of-concept study to evaluate the effects of sublingual buprenorphine (BUP; Subutex) combined with extended-release injectable naltrexone (NTRX; Vivitrol) in the treatment of comorbid PTSD and AUD. No formal interim analyses will be conducted. As such, all analyses described will be performed at the end of the study excluding any efficacy and/or safety summaries provided for the DSMB, which are described in Section 7.4.

The deliverable will be data that advances the understanding of KOR antagonistic pharmacology in the treatment of comorbid AUD and PTSD with the aim to inform a decision about proceeding with full development of soon-to-be marketed KOR antagonists for the treatment of comorbid AUD and PTSD. Finding a novel pharmacologic treatment approach to improve the clinical outcomes for

Veterans and military Service-Members with comorbid PTSD and AUD is the focus of this project, which aligns exceptionally well with the goals of the PASA Consortium.

3 STUDY OBJECTIVES AND OUTCOMES

3.1 Study Objectives

The **specific aim** of this study is to evaluate the effects of sublingual buprenorphine (BUP; Subutex) combined with extended-release injectable naltrexone (NTRX; Vivitrol) in the treatment of comorbid PTSD and AUD.

The **objectives** of this aim are to compare safety and efficacy endpoints in the treatment group versus placebo after 8 weeks (with an additional follow up period extending to 12 weeks).

The study team hypothesizes that the efficacy of BUP/NTRX in the treatment of comorbid PTSD and moderate-to-severe AUD will be greatest in the treatment group followed by a very minimal effect for the placebo group.

SAP version 3.0 was updated to remove the evaluation of the Fear Potentiated Startle (FPS) assessment from this primary SAP. The FPS was changed from a secondary objective to an exploratory objective in SAP version 2.0 due to the pause in the FPS assessment as a result of the COVID pandemic. Exploratory analyses of the FPS assessment will be included and described in a secondary supplemental SAP, as needed.

3.1.1 Primary Safety Objectives

1. Determine if adverse event rates, severity, and relatedness differ between the BUP/NLTRX group compared to the placebo group over the 8-week study period.

3.1.2 Primary Efficacy Objectives

1. To evaluate the efficacy of BUP/NTRX in the treatment of comorbid PTSD and moderate-to-severe AUD based on a matrix outcome of response in terms of both PTSD and AUD outcomes over the 8-week study period.
2. Determine if percent of heavy drinking days (defined as >4 standard drinks/sessions for men and >3 standard drinks/sessions for women) as assessed by the Timeline Follow-Back (TLFB) are reduced for BUP/NTRX treatment over the 8-week study period.
3. Determine if PTSD symptoms, as measured by the change in the Clinician Administered PTSD Scale for DSM-5 (CAPS-5), are reduced for BUP/NTRX treatment over the 8-week study period.

3.1.3 Exploratory Efficacy Objectives

1. Examine if longer treatment is needed to show efficacy in the primary outcome measure between treatment groups at week 12.

3.2 Outcomes

A brief description of the safety and efficacy outcomes are described below. Additional detail regarding the definition of each of these outcomes is located in Section 9.2.

3.2.1 Primary Efficacy Outcomes

As mentioned above, the **primary outcome** is a matrix outcome of response in terms of both PTSD and AUD outcomes. The derivation of this measure will be a binary "yes/no" outcome. The PTSD outcome will be assessed using the CAPS-5 and be defined as "yes" if there is a decrease ≥ 10 points from baseline to week 8. The AUD outcome will be assessed using the TLFB and %HDD. The AUD outcome will be defined as "yes" if there is a reduction >1-shift in World Health Organization (WHO) risk levels of alcohol use from baseline to week 8 (note: the look-back period will consist of 28 consecutive days).

As a complement to the primary outcome, we will also examine each of the components individually (CAPS-5 Total Symptom Severity Score and %HDD)

3.2.2 Exploratory Efficacy Outcomes

3.2.2.1 PTSD score as measured by the PTSD Checklist for DSM-5 (PCL-5)

3.2.2.2 Depression score as measured by the Patient Health Questionnaire (PHQ-9)

3.2.2.3 Physical and Mental Health Summary scores as measured by the Veterans Rand 12-Item Health Survey (VR-12)

3.2.2.4 Change in phosphatidyl ethanol (Peth) from baseline to weeks 8 and 12

3.2.3 Safety Outcomes

Rates, severity and relatedness of adverse events including serious adverse events, study drug-related adverse events, and deaths will be evaluated. See section 8 of the protocol for specific details on safety parameters (i.e. definitions of AEs) and classifications (i.e. severity, relationship, and expectedness definitions).

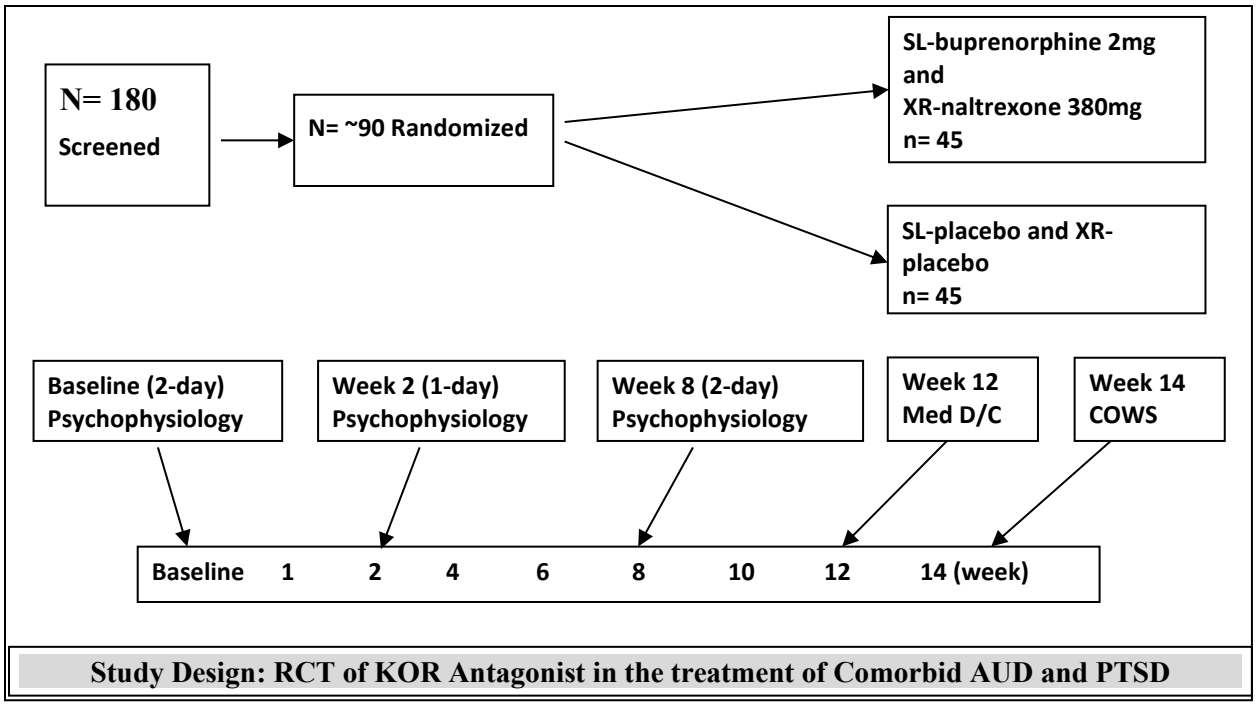
4 STUDY METHODS

4.1 Overall Study Design and Plan

This randomized, double-blind, placebo-controlled, multi-site study evaluates the efficacy of SL-BUP combined with XR-NTX in the treatment of comorbid AUD and PTSD. Recruitment is based on convenience sampling of treatment-seeking veterans and active duty service members. Following screening and baseline visits, eligible patients are randomization to receive one of two treatments in a double-blind fashion for 12 weeks:

- **Treatment A (SL-BUP 2mg and XR-NTX 380mg) or**
- **Treatment B (SL-PLC and XR-PLC)**.

The treatment allocation ratio for the treatment vs. placebo (PLC) regimens is **1:1** and is stratified by site, presence of concomitant antidepressants, and gender using a random permuted block scheme with variable block size (see below Figure). The study population will consist of ~90 Veterans with combat-related PTSD and co-morbid AUD. Detailed eligibility criteria for participant inclusion in this study are included in Section 4.2, below.



4.2 Study Population

4.2.1 Subject Characteristics

The study population is defined by the following eligibility criteria:

Inclusion Criteria

To be enrolled in this study, participants *must meet* the following criteria.

1. Male or female, 18 to 65 years of age, capable of reading and understanding English, and able to provide written informed consent (i.e. no surrogate).
2. Current moderate to severe AUD as determined by MINI International Neuropsychiatric Interview for DSM-5 (MINI-5).
3. At least two recent episodes of heavy drinking (>5 standard drinks/sessions for men and >4 standard drinks/sessions for women) over the past 30 days, and heavy drinking pattern defined as 14 drinks per week for women and 21 drinks per week for men for at least 2 of a 4-week interval within the 90 days prior to baseline; i.e. at least Moderate Risk level on WHO category.
4. PTSD diagnosis defined by MINI-5 at screening; symptoms must be present according to the DSM-5 criteria, i.e. minimum number in each cluster: at least 1 B symptom, 1 C symptoms, 2 D symptoms and 2 E symptoms.
5. Clinician Administered PTSD Scale for DSM-5 (CAPS-5) total score ≥26 for the past one week at baseline.
6. Females of child-bearing potential must be using medically acceptable birth control (e.g. oral, implantable, injectable, or transdermal contraceptives; intrauterine device; double-barrier method) AND not be pregnant OR have plans for pregnancy or breastfeeding during the study.

7. Must have a CIWA-Ar score of < 8 prior to randomization.
8. Willing and able to refrain from medications thought to influence alcohol consumption (other formulations of naltrexone, disulfiram, acamprosate, topiramate, ondansatron, baclofen, and gabapentin).
9. Willing and able to refrain from psychotropic medications: stimulants/ADHD treatment, Alzheimer's medications, antipsychotics, benzodiazepines, antianxiety medications, mood stabilizers, and other sedatives.

Exclusion Criteria:
To be enrolled in this study, participants *must not meet* the following criteria.

1. Current diagnosis of DSM-5 bipolar I, schizophrenia, schizoaffective, and/or major depressive disorder with psychotic features (defined by MINI-5 at screening).
2. Increased risk of suicide that necessitates inpatient treatment or warrants therapy excluded by the protocol, and/or current suicidal plan, per investigator clinical judgement, based on interview and defined on the Columbia Suicidality Severity Rating Scale (C-SSRS).
3. Treatment with trauma-focused therapy for PTSD (e.g. Cognitive Processing Therapy, Prolonged Exposure, or EMDR) within two weeks of baseline study visit. Note: Supportive psychotherapy in process for PTSD at time of Screening may be continued.
4. Current diagnosis of moderate or severe non-alcohol substance use disorder (except for caffeine and nicotine) during the preceding 1 month, based on participant screening interview. Participants must agree to abstain from illicit drugs during the study. Patients who utilize cannabis but do not meet substance use disorder criteria are permitted.
5. Use of opioids within 2 weeks of baseline or opioid use disorder in the previous 90 days.
6. History of severe traumatic brain injury (TBI) per Ohio State University TBI Identification Method. Note: history of mild or moderate TBI is allowed.
7. Any clinically significant, uncontrolled, or medical/surgical condition that would contraindicate use of SL-BUP + XR-NTX, or limit ability to complete study assessments, including seizures (other than childhood febrile seizures), severe renal insufficiency, significant arrhythmia or heart block, heart failure, or myocardial infarction within the past 2 years, severe thrombocytopenia or hemophilia, severe hepatic failure, complete hearing loss, and/or need for surgery that might interfere with ability to participate.
8. Clinically significant laboratory abnormalities, including a thyroid stimulating hormone (TSH) >1.5 times upper limit of normal, hyperthyroidism, and aspartate aminotransferase and/or alanine aminotransferase > 3 times upper limit of normal; cardiovascular findings QTcF >500 msec on electrocardiogram (ECG) or blood pressure >190/110.
9. History of allergic reaction, bronchospasm or hypersensitivity to a naltrexone or buprenorphine.
10. Unable or unwilling to refrain from medications thought to influence alcohol consumption (see inclusion criteria above.)
11. Unable or unwilling to refrain from psychotropic medications (see inclusion criteria above); with the exception of stable doses of antidepressants, prazosin, and non-

benzodiazepine hypnotics and non-benzodiazepine anxiolytics to treat PTSD or insomnia.

12. Persons who are imprisoned, of minor age, diagnosed with dementia, diagnosed with a terminal illness, or otherwise require a surrogate to provide informed consent.

4.3 Study Arm Assignment and Randomization

At baseline, participants who meet eligibility criteria are randomized to treatment A (buprenorphine 2mg SL with naltrexone 380mg IM) or treatment B (SL placebo and IM placebo). The treatment allocation ratio for the treatment vs. placebo regimens is 1:1 and is stratified by site, presence of concomitant antidepressants, and gender using a random permuted block scheme with variable block size. The randomization schedule is created by the PASA statistician who coordinates with the local research pharmacies and develops the web-based randomization assignment system. At week 8, the participants will remain on the assigned treatment and dose from week 8 to 12.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

This is a double-blind study so participants and study staff are masked to treatment assignment. Any unmasking will be reported as a protocol deviation. In the event of an emergency, study investigators may unmask without advance permissions but must promptly report the unmasking within 24 hours by contacting PASA leadership.

4.4.2 Database Lock

The database will be locked after the completion of all study follow-up visits for all participants. Once all participants have completed all follow-up visits, a final query report will be generated and the study coordinators will address data queries within Medidata Rave. The database will be locked after all final queries have been resolved (estimated ~2 weeks after all follow-up visits completed). Randomization assignment will not be unmasked beyond the details of Section 4.4.1 until after database lock is finalized.

4.5 Study Flow Chart of Assessments and Evaluations

Assessments	Schedule of Assessments: Screening, Baseline, Week (#) or Telephone (T) Visit																	
	Screen	Baseline	Baseline (Day 2)	Week 1	Week 2	T	Week 4	T	Week 6	T	Week 8	Week 8 (Day 2)	T	Week 10	T	Week 12	T	Week 14
PAYMENTS	25		75	30	60		30		30			80		30		40		20
Assessments to Qualify for Study and Characterize Population																		
Demographics	X																	
Smoking Status	X										X					X		
MINI	X																	
H&P	X																	
VS, BMI, BrAC	X	X		X	X		X		X		X			X		X		X
Labs/ECG	X															X		
Pregnancy	X						X				X					X		
CIWA-Ar	X	X														X		X
COWS	X															X		X
ADS	X																	
OSU-TBI	X																	
BIS	X																	
Psychophysiology Study																		
Fear Conditioning		X									X							
Psychophysiological Stress Reactivity		X	X		X						X	X						
Alcohol Related Outcomes																		
TLFB	X	X		X	X		X		X		X			X		X		
OCDS		X		X	X		X		X		X			X		X		
Peth		X									X					X		
Serum BUP											X					X		
SIP		X									X					X		
DAQ		X			X						X							
Urine Drug Screen	X	X		X	X		X		X		X			X		X		
PTSD Related Outcomes																		
LEC-5 Checklist	X																	
PCL-5	X	X		X	X		X		X		X			X		X		
CAPS-5		X					X				X					X		
Psychological Symptoms and Quality of Life																		
C-SSRS		X									X							
PHQ-9		X					X				X					X		
VR-12		X									X					X		
CGI-S		X		X	X		X		X		X			X		X		
CGI-I				X	X		X		X		X			X		X		
Adverse Events and Concomitant Medications																		
Con-Meds	X	X		X	X		X		X		X			X		X		X
Adverse Events				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Study Drug* **																		
Dispense Drug			X	X	X		X		X			X		X				
Pill Counts				X	X		X		X		X			X		X		
Study Drug Knowledge																X		

5 ANALYSIS POPULATIONS

5.1 Intention-to-Treat

The primary and exploratory analyses will be based on the intention-to-treat (ITT) principle, with data from all participants analyzed according to the arm to which they were randomized irrespective of the amount intervention received. This includes participants initially randomized to the high-dose treatment arm (SL-BUP 8mg and XR-NTX 380mg) that was removed from the protocol (Version 5). Participants who receive intervention but do not complete the study will be used in all analyses for which data are available. This approach ignores noncompliance, protocol deviations, withdrawal and lost-to-follow-ups. Analysis of the ITT population avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers, accepting that protocol deviations occur in actual clinical practice (Heritier, Gebski, & Keech, 2003).

5.2 Per-Protocol

As in most clinical trials, some participants may not adhere to the intervention they were randomized to receive, reducing fidelity to the intervention as designed and potentially changing the effectiveness of the intervention. The most likely form of noncompliance will be lack of adherence to the randomized medication. The potential impact of this noncompliance is underestimating the magnitude of the true treatment effect. To assess this, in addition to the ITT analyses, we will also conduct all analyses at the primary time point (8-weeks) using a per protocol population; defined as completion of the NTX/Placebo injections at baseline and week 4, as well as completion of 80% of Buprenorphine pills dispensed from baseline through week 8. The per-protocol analysis provides an estimate of the true efficacy of an intervention (i.e., among those who completed the treatment as planned) (Ranganathan, Pramesh, & Aggarwal, 2016). No other analysis populations are defined for this protocol.

6 SAMPLE SIZE DETERMINATION

The sample size for a proof-of-concept study should be large enough to be able to detect a drug effect or surrogate signal, but at the same time expose a minimum number of subjects to an experimental drug combination in a reasonable timeframe within a restricted budget. Sample size was determined based on sample sizes used in similar double-blind, placebo-controlled studies of PTSD and AUD, and based on the estimated recruitment rate over the allotted enrollment timeline. With these factors in mind, we plan to enroll approximately 180 participants to allow screen failures/drop-out prior to randomization of approximately 90 participants across three sites (45 in each treatment arm) within a 24-month enrollment period. We estimate that attrition will be approximately 30%. However, all randomized participants will remain in analysis (intent-to-treat) regardless of how long they participate in the study.

For the primary outcome, we assumed the following success rates based on a similar previous study: Placebo arm 5/45 (11%) or less and 2mg BUP/NLTRX arm 17/45 (38%) or greater. We calculated power for two by two contingency tables for comparison of the treatment and Placebo arm using a Fisher's Exact test with the Walters approximation applying a nominal two-sided alpha level of 0.05. We observe power of 80% or greater for the above assumptions. In addition, a situation with more positive responses in the active treatment arm or fewer positive responses in the placebo arm would increase apparent power.

For secondary outcomes, we estimate that we will have >80% power to detect a medium effect size (Cohen's f=0.25, two-sided alpha level of 0.05) at 8 weeks. Under optimal conditions, where the correlation between consecutive observations on the same participant is high, and the variability observed between pairs of time points is fairly constant, fewer participants would be required (see table below). Under a less than ideal, worst-case situation where correlation between measurements is low and the variability observed between pairs of time points changes, the proposed sample size is still large enough to detect a moderate effect size, even after accounting for screen failures and drop-outs.

For the Psychophysiological and Fear Conditioning analyses, a total sample size of 90 (45 per group) will allow for a conservative dataset with over 80% power to detect significant differences between the treatment group against the placebo group at the $\alpha = 0.05$ level. Estimates were based on means and standard deviations from previous study samples.

Estimated Sample Sizes per Group for Detecting a Moderate Effect Size at 80% Power			
Modeling Conditions	AUD Outcome	PTSD Outcome	Higher of PTSD/AUD with Screen Failures and Drop-outs (+15)
High Correlation, Constant Variance	42	30	57
High Correlation, Non-constant Variance	42	63	78
Low Correlation, Constant Variance	81	57	96
Low-Correlation, Non-constant Variance	81	120	135

7 STATISTICAL / ANALYTICAL ISSUES

7.1 General Rules

All statistical computations will be performed and data summaries will be created using SAS 9.4 or higher. If additional statistical packages are required, these will be discussed in the study report. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of participants; continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by only presenting median and range. Model-based analyses described in Section 9 will be used to obtain point estimates and associated confidence intervals for various measures as well as p-values for comparisons of data between treatment groups. Note: initial models will include the high dose treatment arm; however, if the model cannot converge due to the small number of participants randomized to this arm prior to removal from the study, then the model will only compare outcomes between the low dose and placebo treatments. P-values presented will be based on two-sided tests unless otherwise specified and generally only adjusted for randomization factors (study site, gender, and presence of concomitant antidepressants). Note: in the event a model cannot converge due to small sample sizes for a specific randomization factor, it will be removed from the model. For continuous outcomes, checks of normality will be performed and if required, transformations or non-parametric tests will be employed.

7.2 Adjustments for Covariates

Since this is a randomized, double-blind, placebo-controlled study with equal balance in the treatment arms, no adjustment for demographic or baseline characteristics is planned. However, any assessment of the impact of covariates or adjustments for them will be done in an exploratory manner only and described in the study report.

7.3 Handling of Dropouts and Missing Data

Analysis of the primary and exploratory outcomes will be based on a mixed-effect model of repeated measures (MMRM) using data from all time points to account for missing data and thereby maximize information used for the analyses. These models treat missing data as ignorable missing, assuming any missing data are missing at random. No imputation processes will be used to replace missing data. However, every effort will be made to minimize missing data. Additionally, because missing observations have the potential to alter the results of analyses, we will examine whether the pattern of missing data is different among the groups. If there are no systematic differences between those with and without missing data, the data will be considered to be missing at random. If there are significant differences in dropout or missing data patterns between treatment arms, we will conduct sensitivity analyses to determine the impact of missing information on the treatment comparisons.

7.4 Interim Analyses and Data Monitoring

The PASA Consortium Leadership has established a Data Safety Monitoring Board (DSMB) to oversee this study. Members of the DSMB are independent of the study investigators and include representatives with substance abuse, pharmacology and psychology/psychiatry expertise, a biostatistician and an ethicist. None will have financial, scientific, or conflicts of interest which might interfere with their unbiased assessment of the progress of the trial. The DSMB will meet at least once every four months as specified in the DSMB charter to review the study, although may be convened between planned meetings to discuss study issues related to adverse events/safety. This protocol will be approved by the DSMB prior to initiation of recruitment. While there are study halting rules that will trigger a study review by the DSMB (details found in the protocol), there are no formal study stopping rules based on the safety review that would trigger permanent study termination without the review and recommendation of the DSMB.

This study may be suspended or prematurely terminated if there is reasonable cause. Written notification, documenting the reason for study suspension or termination, will be provided by the suspending or terminating party to the investigator, DoD Grants Officer Representative (GOR) for further distribution to the Programmatic Panel, PASA Consortium PI (or designee), DSMB, and regulatory authorities. If the study is prematurely terminated or suspended, the PI will promptly inform the IRB and will provide the reason(s) for the termination or suspension. Circumstances that may warrant termination or suspension include, but are not limited to:

- A pattern of AEs indicating that SL-BUP + XR-NTX is inappropriate or unsafe
- Determination of unexpected, significant, or unacceptable risk to participants
- New information emerges that suggests the risks are increased beyond an acceptable limit and outweigh benefits
- Significant and recurring protocol violations that threaten the integrity of the study data
- Data that are not sufficiently complete and/or evaluable
- Continuing and ongoing serious noncompliance with human research protections and regulations that warrants study closure
- Determination of futility. While there are no formal assessments of futility for efficacy, the study may be terminated if operationally futile, e.g. there is an inability to enroll or retain study participants. The study may resume once concerns about safety, protocol

compliance, data quality is addressed and satisfy the PASA Consortium, IRB and/or other regulatory agencies.

- Generally, any study suspension that could occur will be based upon review of the study and recommendations made by the DSMB. Details of safety oversight and associated study halting rules are provided in Sections 8.6 and 8.5, respectively.
- A local site that is not performing for one of the reasons above, may be selectively terminated and replaced with a new site.

A description of the report to be provided to the DSMB is listed in Appendix A.

7.5 Masked Data Review

A masked data review of outcome data for this study is not planned.

7.6 Multicenter Studies

This is a 3-center study including Tuscaloosa Research & Education Advancement Corporation (TREAC)/Tuscaloosa VA Medical Center (TVAMC), Connecticut Research & Education Foundation (CREEF)/VA Connecticut Healthcare System (VACT), and Wayne State University. Although we initially expected the total number of enrolled participants to be roughly equal across the three study sites, only one participant from the Wayne State site enrolled and completed the study. Since randomization was stratified based on study site, site will be included as a term in all model-based analyses. Since only one participant was enrolled at Wayne State, this participant will be removed during sensitivity analyses so comparisons can be made between study site.

7.7 Multiple Comparisons and Multiplicity

One single primary formal hypothesis test is planned for this study at the 0.05 level of significance, which is the comparison of the SL-BUP + XR-NTX treatment group vs. placebo for the primary outcome at the primary timepoint (Week 8). Additionally, regardless of significance of the primary outcome, statistical significance of the secondary outcomes will be assessed, with adjustment for multiple comparisons using the Benjamini-Hochberg false discovery rate correction to maintain the nominal alpha level of 0.05 (Benjamini & Hochberg, 1995) Primary and secondary hypothesis tests resulting in nonsignificant p-values will be interpreted as inconclusive. All other treatment group comparisons will be considered descriptive in nature with no adjustment for multiple comparisons, and all confidence intervals will be generated using 95% bounds.

7.8 Examination of Subgroups

No subgroup analyses are planned at this time.

7.9 Assessment Windows

All screening assessments will be performed prior to Day 0 (study baseline/admission). Screening will take place over 1 to 14 days and include the study instruments listed in the table above (section 4.5). Most screening assessments are repeated at Day 0 if occurring more than one-month post screening. The most recent data collected during the screening/baseline period will be used for summarizing and analyzing baseline data.

Participants will visit their respective site to complete study procedures at weeks 1, 2, 4, 6, 8, 10, and 12. The visit windows for these in-person visits is +/- 7-days. It is important to confirm

appointment with subjects at each visit, so that if adjustments need to be made so that the subject does not run out of medication between the scheduled in-person visits. The CRC will make a telephone contact with the participant between face-to-face visits to remind the participant of the next office visit and to check on adherence to the study medication and if there have been any adverse events. While every attempt is made to retain the participant in the study on medication for the full 12 weeks, the participant may discontinue medication at any point and remain in the study for assessments through week 12. Out-of-window visits or assessments must be reported as protocol deviations. If any exclusions are made for these reasons, these exclusions will be described in the study report.

8 STUDY SUBJECT CHARACTERIZATION

8.1 Subject Disposition

Participant eligibility status will be summarized and overall disposition of study participants will be described using a standard CONSORT diagram. The number of participants randomized and those completing or discontinuing from study therapy will be summarized.

8.2 Protocol Deviations

Protocol deviations will be listed with information such as type of deviation, time of occurrence, and reason. Incidence of protocol deviations will also be summarized overall and for each protocol deviation category. Incidence rate of protocol deviations will be calculated as the number of deviations divided by number of participant days.

8.3 Study Drug Exposure and Adherence

A dosing/pill count log will record the doses given each week (or since the last in-person visits) and whether the doses were distributed and taken per protocol. A descriptive table of overall dosing will be provided along with a listing of missed doses and reasons for the missed doses. To assess medication adherence, blood samples collected at study weeks 8 and 12 will be tested for buprenorphine and its metabolite, norbuprenorphine, using quantitative assessments. Results will be summarized graphically (box-plots) over time for the active treatment study group.

8.4 Demographic and Baseline Characteristics

All data collected at baseline will be summarized descriptively for the entire efficacy (intent-to-treat) population. Demographic variables of interest include, but are not limited to, age (years), gender, race and ethnicity, veteran status, service-connected disability, and education level. Baseline efficacy variables include, but are not limited to CAPS-5, PCL-5, %HDD, PHQ-9, and VR-12.

9 EFFICACY AND SAFETY ANALYSES

9.1 Overview of Analysis Methods

The statistical goal for the study is to evaluate the effects of sublingual buprenorphine (BUP; Subutex) combined with extended-release injectable naltrexone (NTRX; Vivitrol) in the treatment of comorbid PTSD and AUD. Although the analyses will use data collected from baseline through week 12, the primary hypothesis that the primary outcome differs between the treatment arm vs. placebo will be tested at Week 8. The analysis also will formally test treatment differences for the PTSD and AUD outcomes, independently. Primary and exploratory analyses will use model-based approaches that take advantage of the longitudinal structure of the outcome data to address missing data caused by patient loss to follow-up or nonresponse and take into

consideration correlated data collected across time. Additional analytic details are in the sections below.

9.2 Efficacy and Safety Outcome Variables

<table>
<tr><th>Variable</th><th>Type</th><th>Definition</th></tr>
<tr><td colspan="3">Primary Outcomes</td></tr>
<tr><td>Composite AUD + PTSD outcome</td><td>Binary (Yes/No)</td><td>A positive response is defined as a decrease from baseline of at least 10 points on the CAPS-5 score (Schnurr et al., 2022) and a reduction of at least 1 risk level of alcohol use, as defined by WHO, at week 8. AUD data is collected on the TLFB, a semi-structured interview that uses a calendar prompt and other memory aids to facilitate accurate recall of the numbers of drinks consumed each day during a given target period (Sobell et al., 1996). The definition of WHO risk drinking levels are defined by Knox et al. (2018), and displayed in the below table
<table>
<tr><th>Wave 1 WHO risk drinking level</th><th>Definition of each level, in grams (US standard drinks)</th></tr>
<tr><td rowspan="2">Very high</td><td>>100 g (>7.1 drinks) for men;</td></tr>
<tr><td>>60 g (>4.3 drinks) for women</td></tr>
<tr><td rowspan="2">High</td><td>60–100 g (4.3–7.1 drinks) for men;</td></tr>
<tr><td>40–60 g (2.9–4.3 drinks) for women</td></tr>
<tr><td rowspan="2">Moderate</td><td>40–60 g (2.9–4.3 drinks) for men;</td></tr>
<tr><td>20–40 g (1.4–2.9 drinks) for women</td></tr>
<tr><td rowspan="2">Low</td><td>1–40 g (<2.9 drinks) for men;</td></tr>
<tr><td>1–20 g (<1.4 drinks) for women</td></tr>
</table>
</td></tr>
</table>

Variable	Type	Definition
		The TLFB assessment is given at screening, baseline, and weeks 1, 2, 4, 6, 8, 10, and 12. The screening measurement uses a 90-day recall period whereas the subsequent (post treatment follow up) measurements use a recall period of "since the last measurement." The baseline WHO risk level will be derived from TLFB data collected at both the screening and baseline visit. The look-back period to define the WHO risk level will be conducted over 28 consecutive days and based on the average number of reported standard drinks during those 28 consecutive days. *Each derived outcome measure for a specific timepoint will be based on the baseline visit date*. *In the event 28 consecutive days are not captured for the full look-back period, participant data will be excluded from analyses in the event the number of total days included in look-back period is less than one standard deviation of the entire population.*
Percent Heavy Drinking Days (%HDD)	Continuous (range between 0 and 100)	Percent Heavy Drinking Days (%HDD) is measured via the Timeline Follow Back instrument (Sobell et al., 1996). The screening (baseline) measurement uses a 90-day recall period whereas the subsequent (post treatment follow up) measurements use a recall period of "since the last measurement." Baseline %HDD will be derived from TLFB data collected at both the screening and baseline visit. The look-back period to define %HDD will be conducted over 28 consecutive days and based on the number of heavy drinking days reported during this time period. *Each derived outcome measure for a specific timepoint will be based on the baseline visit date. In the event 28 consecutive days are not captured for the full look-back period, participant data will be excluded from analyses in the event the number of total days included in look-back period is less than one standard deviation of the entire population.* A heavy drinking day is defined as follows: *>4 standard drinks/sessions for men and >3 standard drinks/sessions for women*

Variable	Type	Definition
CAPS-5 Total Symptom Severity Score	Continuous (range between 0 and 80)	CAPS-5 total symptom severity score is measured via the Clinician-Administered PTSD Scale For DSM-5 (CAPS-5). The score is calculated by summing severity scores for the 20 *DSM-5* PTSD symptoms (Weathers et al., 2013), with a higher score indicating worse PTSD symptoms.
Exploratory Outcomes		
PCL-5 Score	Continuous (range between 0 and 80)	The PCL-5 is a 20-item self-report measure that assesses the 20 DSM-5 symptoms of PTSD. Its purposes include screening for PTSD and/or provisional diagnosis, and monitoring symptom change before, during, and after treatment. A total symptom severity score ranging from 0 to 80 is possible (Weathers et al., 2013). The score is calculated by summing each of the 20 *DSM-5* PTSD questions, with a higher score indicating worse PTSD symptoms.
PHQ-9 Score	Continuous (range between 0 and 27)	Depression symptoms will be assessed using the validated PHQ-9 (Kroenke, Spitzer, & Williams, 2001) in support of exploratory analyses. The PHQ-9 was developed as a short form of the full Patient Health Questionnaire (which was a self-administered version of the PRIME-MD instrument). Severity scores range from 0 to 27, with a score of 5 to 9 representing mild depression, a score of 10 to 14 representing moderate depression, and a score of 15 or greater representing a severe level of depression (Kroenke et al., 2001).
VR-12 Scores (physical and mental health summaries)	Continuous (T-score with a mean of 50 and standard deviation of 10)	The VR-12© was developed from the Veterans RAND 36 Item Health Survey (VR-36©; formerly called the Veterans SF-36), which was adapted from the RAND 36-Item Health Survey (RAND 36 version 1.0) at the RAND Corporation and the Medical Outcomes Study. The twelve items provide an estimate for eight domains of functional health and well-being: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health. Together, the first four domains constitute a Physical Health

Variable	Type	Definition
		summary (PCS) measure, and the second 4 constitute a Mental Health summary (MCS) measure (Ware, Kosinski, & Keller, 1996). PCS and MCS scores are derived using an algorithm that is referenced to a metric centered at 50.0 using the 2000–2002 US Medical Expenditure Panel Survey population.
Phosphatidyl ethanol (PEth)	Continuous (range between 0 and >210 ng/mL)	Phosphatidyl ethanols (PEth) are a series of abnormal phospholipids located in the cell membrane and formed in the presence of ethanol and phospholipase D. PEth serves as a biomarker of alcohol ingestion in the past 2-3 weeks using liquid chromatography tandem mass spectrometric assessment of human blood (whole or dried spots). PEth in blood if often used as a marker of chronic alcohol use and was collected at baseline and weeks 8 and 12. At the present time, the international scientific community has not yet established a cut-off value for PEth concentration in blood to be used for differentiating an acceptable social ethanol intake (<40 g for males and <20 g for females, according to the WHO parameters), from an at-risk-alcohol-use (40–60 g/die) and chronic excessive drinking behavior (>60 g/die). However, per LabCorp, PEth levels in excess of 20 ng/mL are considered evidence of moderate to heavy ethanol consumption.
Safety Outcomes		
Adverse Events	Counts and proportions	Total number and proportion of treatment emergent adverse events. Comparison of AE rates by treatment group will also be stratified by type and severity.
Withdrawals due to Adverse Events	Proportion	Proportions of individuals by treatment group who withdrew from the study due to adverse events
Adherence and Retention Outcomes		

Variable	Type	Definition
Serum Buprenorphine and Norbuprenorphine	Continuous	Buprenorphine and its metabolite norbuprenorphine are assayed using liquid chromatography tandem mass spectrometry by a commercial laboratory. Sample results are available via the commercial laboratory portal using a study specific username/password combination and the unique participant code/requisition number. To assess medication adherence, blood samples collected at study weeks 8 and 12 will be tested for buprenorphine and its metabolite, norbuprenorphine, using quantitative assessments.
Pill Count	Count and proportion	A dosing log case report form (CRF) is completed weekly (or since the last in-person visit) for each participant. The number of completed NTX injections received (at baseline, week 4, and week 8) is recorded, as well as the number of missed SL-BUP pills per week. For each medication (NTX and SL-BUP), the proportion of subjects who missed any treatment will be derived, as well as the total number of missed injections or pills, per participant.

9.3 Primary Efficacy Analysis Methods

The primary outcome for this study is a binary measure assessed at weeks 4, 8, and 12. A positive response is defined as a decrease from baseline of at least 10 points on the CAPS-5 score (Schnurr et al., 2022) and a reduction of at least 1 risk level of alcohol use, as defined by WHO (Knox et al. 2018). The treatment effect on comorbid PTSD + AUD over time will be assessed using a mixed logistic regression model. The model will be used to generate point and interval estimates of the risk differences and to test differences in proportions of PTSD+AUD resolution between treatment groups at the 8-week visit (primary timepoint), as well as at weeks 4 and 12 (secondary timepoints). The model will include fixed effects for the treatment group, visit (as a categorical variable), gender, site, presence/absence of anti-depressants, treatment-by-visit interaction, and random effects for participant (see model statement below). This approach will provide consistent estimates and valid inferences under missing at random (MAR) data assumptions while accounting for correlation among multiple measures on the same participant. This mixed model will improve the power of the study and the precision of all estimates by allowing all available measures for an individual to be incorporated in the analysis, even if other timepoints are missing. Sensitivity analysis maybe considered by excluding participants that have a low or moderate WHO risk level at baseline. This model is further described below:

$$g(E(Y_{ij})) = \beta_0 + \beta_1 Visit_{ij} + \beta_2 Site_{ij} + \beta_3 Treatment_{ij} + \beta_4 Gender_{ij} + \beta_5 Anti\text{-}depressant_{ij} + \beta_6 (Treatment*Visit)_i + s_i + \varepsilon_{ij}$$

Where g(E(.)) indicates the identity link function on the expected value of Y, Y_{ij} is the jth measure of PTSD+AUD measurement in subject i, β_0 is the intercept, β_1 through β_6 are coefficients, s_i is a random subject effect and ε_{ij} is the residual error term.

Statistical analysis modeling will be carried out utilizing SAS/STAT PROC GLIMMIX (SAS Institute Inc. 2017) with the general structure of the SAS code for this model is listed below. Generalized linear mixed models (GLMMs) are a flexible generalization of the linear mixed effect model that allows response outcomes to have an error distribution other than the normal distribution (McCullagh, & Nelder, 1989). The GLMM will employ random intercepts for the between participants' part of the covariance model, and the within participants' part of the covariance model will be initially assumed as autoregressive. The study's primary hypothesis will be tested and associated odds ratio obtained using the 'oddsratio' statement which estimates the risk difference in PSTD+AUD resolution between the treatment arms.

```
proc glimmix data=anlydata;
   class patient treat gender visit site antidepress;
   model primoutcome = treat gender visit site antidepress
   treat*visit/ solution link=logit dist=binary oddsratio;
   random int / type=ar(1) subject=patient;
   ods select parameterestimates Oddsratios;
run;
```

9.3.1 CAPS

As mentioned above, the CAPS-5 is a continuous measure that is assessed at baseline and weeks 4, 8, and 12. The CAPS-5 Total Symptom Severity Score ranges from 0 to 80, with higher scores indicating greater PTSD symptoms (Weathers et al., 2018). The treatment effect on the clinical criteria of PTSD as measured by the CAPS-5 over time will be

assessed using a generalized-linear mixed model to account for temporal correlation between measurements and the clustering of the data. This model will include fixed effects for treatment group, visit (as a categorical variable), gender, site, presence/absence of anti-depressants, and a treatment-by-visit interaction term, and a random effect for participant (see model statement below).

$$Y_{ij} = \beta_0 + \beta_1 Visit_{ij} + \beta_2 Site_{ij} + \beta_3 Treatment_{ij} + \beta_4 Gender_{ij} + \beta_5 Anti\text{-}depressant_{ij} + \beta_6 (Treatment*Visit)_i + s_i + \varepsilon_{ij}$$

Where Y_{ij} is the jth measure of CAPS-5 in subject i, β_0 is the intercept, β_1 through β_6 are coefficients, s_i is a random subject effect and ε_{ij} is the residual error term.

Statistical analysis modeling will be carried out utilizing SAS/STAT PROC MIXED (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006) with the general structure of the SAS code for this model shown below. Adjusted marginal means (also called LS means) will be used to report and test for differences in mean change in CAPS-5 score at 8 weeks for the placebo and BUP/NTRX groups, providing 95% confidence intervals and a p-value for the difference between them.

```
proc mixed data = anlydata;
     class patient treat gender visit site antidepress;
     model CAPS5_Score = treat gender visit site antidepress
     treat*visit /s;
     random int / subject= patient;
     repeated visit / type=cs subject= patient;
     lsmeans treat*visit;
run;
```

9.3.2 %HDD

As mentioned above, %HDD is assessed by the TLFB and is collected at baseline and weeks 2, 4, 6, 8, 10, and 12. The treatment effect on the clinical criteria of AUD as measured by %HDD over time will be assessed using a generalized-linear mixed model to account for temporal correlation between measurements and the clustering of the data. A similar model to that listed in Section 9.3.1 will be run to produce adjusted marginal means to report and test for differences in mean %HDD for the placebo and BUP/NTRX groups, providing 95% confidence intervals and a p-value for the difference between them. NOTE: the lookback period to derive %HDD will be performed at 28-days, each derived measure at a specific timepoint based on the baseline visit date. Sensitivity analysis maybe also performed using a 14-day look-back period.

9.4 Exploratory Efficacy Analysis Methods

9.4.1 PCL-5

While the CAPS-5 is considered the gold standard for diagnosing PTSD, the PCL-5 is often used to screen for PTSD and monitor symptom changes before, during, and after treatment. A total symptom severity score ranges from 0 to 80 (Weathers et al., 2013). The PCL-5 is collected at baseline and weeks 1, 2, 4, 6, 8, 10, and 12. The treatment effect on PTSD as measured by the PCL-5 over time will be assessed using a generalized-linear mixed model

to account for temporal correlation between measurements and the clustering of the data. A similar model to that listed in Section 9.3.1 will be run to produce adjusted marginal means to report and test for differences in mean PCL-5 for the placebo and BUP/NTRX group, providing 95% confidence intervals and a p-value for the difference between them.

9.4.2 PHQ-9

The PHQ-9 is the major depressive disorder (MDD) module of the full PHQ. This instrument is used to provisionally screen, diagnose, and monitor depression and grade severity of symptoms in general medical and mental health settings (Kroenke K., Spitzer R., Williams, J., 2001). A total symptom severity score ranges from 0 to 27. The PHQ-9 is collected at baseline and weeks 4, 8, and 12. The treatment effect on depression as measured by the PHQ-9 over time will be assessed using a generalized-linear mixed model to account for temporal correlation between measurements and the clustering of the data. A similar model to that listed in Section 9.3.1 will be run to produce adjusted marginal means to report and test for differences in mean PHQ-9 for the placebo and BUP/NTRX group, providing 95% confidence intervals and a p-value for the difference between them.

9.4.3 VR-12

The VR-12 is a patient-reported global health measure that is used to assess a patient's overall perspective of their health (Jones, Kazis, Lee, et al., 2001). Answers on the 12-item questionnaire are summarized into two scores, a physical component score (PCS) and a mental component score (MCS). The PCS and MCS scores are derived using an algorithm that is referenced to a metric centered at 50.0 (with a standard deviation of 10) using the 2000–2002 US Medical Expenditure Panel Survey population. The VR-12 is collected at baseline and weeks 8 and 12. The treatment effect on these global health measures over time will be assessed using a generalized-linear mixed model to account for temporal correlation between measurements and the clustering of the data. A similar model to that listed in Section 9.3.1 will be run to produce adjusted marginal means to report and test for differences in mean PCS and MCS for the placebo and BUP/NTRX group, providing 95% confidence intervals and a p-value for the difference between them.

9.4.4 Phosphatidyl Ethanol (PEth)

Testing for phosphatidyl ethanol (PEth) is a relatively new tool for detecting and grossly quantifying a person's chronic alcohol consumption. PEth values were collected at baseline and weeks 8 and 12. At each collection timepoint, results were reported as either "positive" or "negative", and if "positive", the PEth value in the blood was reported in ng/mL. For analysis purposes, "negative" results will have a PEth value inputted as 5 ng/mL, the midpoint of the 'not detectable range' (which is reported as less than 10 ng/mL). The treatment effect on PEth measurement over time will be assessed using a generalized-linear mixed model to account for temporal correlation between measurements and the clustering of the data. A similar model to that listed in Section 9.3.1 will be run to produce adjusted marginal means to report and test for differences in mean PEth for the placebo and BUP/NTRX group, providing 95% confidence intervals and a p-value for the difference between them.

9.5 Safety Analysis Methods

9.5.1 Adverse Events

Any AEs will be listed and summarized by system organ class and preferred event term. Summaries will include the number of individuals experiencing events over the duration of the study. These summaries will be created for all AEs and AEs by severity (mild or moderate vs. severe or greater). Summaries will be provided by treatment group. If models can be run, point and confidence interval estimates of the risk difference will be obtained by modeling the probability of experiencing at least one adverse event. The model is specified with the equation:

$$g(E(Y_{ij})) = \beta_0 + \beta_1 Treatment_{ij} + s_i + \varepsilon_{ij}$$

Where g(E(.)) indicates the identity link function on the expected value of Y, Yij is the occurrence of an AE outcome (any AE, any SAE, and any related AE) in subject i, β_0 is the intercept, β_1 is the treatment group coefficients, si is a random subject effect, and εij is the residual error term.

Point and confidence intervals for the difference in experienced AE proportion between treatment groups will be estimated with a generalized linear mixed model (GLMM), assuming a binomial distribution and identity link. GLMMs are a flexible generalization of the linear mixed effect model that allows response outcomes to have an error distribution other than the normal distribution (McCullagh, & Nelder, 1989). Statistical analysis modeling will be carried out utilizing SAS/STAT PROC GLIMMIX (SAS Institute Inc. 2017) with the general structure of the SAS code for this model is listed below:

```
proc glimmix data=anlyaedata;
   class patient treatment;
   model any_ae = treatment / link=identity dist=binomial;
   random int / subject=patient;
run;
```

A similar model will be run for the difference in proportion of withdrawn subjects due to adverse events between treatment groups. If the above specified models cannot be fit due to limitations of the study data, frequencies will be analyzed using chi-square test.

9.6 Adherence and Retention Analysis Methods

9.6.1 Buprenorphine and Norbuprenorphine Assays

Buprenorphine and its metabolite norbuprenorphine are assayed using liquid chromatography tandem mass spectrometry by a commercial laboratory at weeks 8 and 12. Processed blood samples to quantify and compare the amount of buprenorphine and norbuprenorphine at weeks 8 and 12 by treatment group will be evaluated using an independent group T-test.

9.6.2 Pill Count and Received Injections

A self-report medication adherence case report form (CRF) is administered at each visit to document the completion of NTX injection and the number of missed and taken SL-BUP pills. For each medication (NTX and SL-BUP), the proportion of subjects who missed any treatment will be derived, as well as the total number of missed injections or pills per visit by treatment group. Differences in self-reported medication adherence by treatment group will be presented in the form of line bars (numbers of missed SL-BUP pills) or bar charts (proportion of missed NTX injections).

10 REFERENCES

Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B, 57, 289-300.

Jones D., Kazis L., Lee A., Rogers W., Skinner K., Cassar L., Wilson N., Hendricks A. Health status assessments using the Veterans SF-36 and SF-12. Methods for evaluating outcomes in the Veterans Health Administration. Journal of Ambulatory Care Management 2001; 24(3):1-19.

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x

Knox J, Wall M, Witkiewitz K, Kranzler HR, Falk D, Litten R, Mann K, O'Malley SS, Scodes J, Anton R, Hasin DS; Alcohol Clinical Trials (ACTIVE) Workgroup. Reduction in Nonabstinent WHO Drinking Risk Levels and Change in Risk for Liver Disease and Positive AUDIT-C Scores: Prospective 3-Year Follow-Up Results in the U.S. General Population. Alcohol Clin Exp Res. 2018 Nov;42(11):2256-2265. doi: 10.1111/acer.13884. Epub 2018 Oct 3. PMID: 30204248; PMCID: PMC6263142.Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS® for Mixed Models, Second Edition*, Cary, NC: SAS Institute Inc.

McCullagh, P, & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC Press.

SAS Institute Inc. 2017. *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

Sobell LC, Sobell MB, Buchan G, Cleland PA, Fedoroff I, Leo GI (1996). The reliability of the Timeline Followback method applied to drug, cigarette, and cannabis use. Paper presented at the 30th Annual Meeting of the Association for Advancement of Behavior Therapy, New York, NY, November 1996.

Schnurr PP, Chard KM, Ruzek JI, et al. Comparison of Prolonged Exposure vs Cognitive Processing Therapy for Treatment of Posttraumatic Stress Disorder Among US Veterans: A Randomized Clinical Trial. JAMA Netw Open. 2022;5(1):e2136921. Published 2022 Jan 4. doi:10.1001/jamanetworkopen.2021.36921

Ware, J., Kosinski, M., Keller, SD. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care, 34, 220-33. Doi:10.1097/00005650-199603000-00003.

Weathers F., Litz B., Keane T., Palmieri T., Marx B., Schnurr P. (2013). The PTSD Checklist for DSM-5 (PCL-5). Scales available from the National Center for PTSD at www.ptsd.va.gov

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military Veterans. Psychological Assessment, 30, 383-395. doi:10.1037/pas0000486.

Appendix A DSMB Report

The following will be included in a report provided for periodic review of the study by the DSMB committee:

a) A summary table of demographic information on all participants including age, sex, race, ethnicity, and whether the participant has served in the US military.

b) A summary table of disposition. Specifically: # of individuals consented, randomized, admitted, and completing each study visit, as well as final disposition based on final study status.
c) A summary table of overall study therapy receipt and completion. Specifically: Proportion of BUP doses taken and missed and NTX injections received.
d) A listing of all adverse events (including SAEs) and summary of proportion of subjects by day with at least one AE.
e) A listing of all protocol deviations and summary of proportion of subjects with at least one deviation.
f) Tabulation of primary and exploratory endpoints by study site.

Appendix B List of Potential Tables and Figures

Tables
a) Demographics and Baseline Characteristics by Treatment Group
b) Comparison of Primary Composite Outcome Measure by Treatment Group
c) Comparison of Change in %HDD by Treatment Group
d) Comparison of Change in CAPS-5 by Treatment Group
e) Comparison of Change in PCL-5 by Treatment Group
f) Comparison of Change in PHQ-9 by Treatment Group
g) Comparison of Change in VR-12 Measures by Treatment Group
h) Comparison of AEs by Treatment Group
Figures
a) Study CONSORT
b) Line plots with error bars for mean change from baseline by treatment group for the following outcome measures: CAPS-5, %HDD, PCL-5, PHQ-9, and VR-12
c) Box and whisker plots for vital measures
d) Bar charts for missed SL-BUP and NTX injections by treatment group