Statistical Analysis Plan

Protocol Title:	A Phase 3, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study to Evaluate the Efficacy and Safety of TNX-102 SL in Participants with PTSD Taken Daily at Bedtime (Protocol No. TNX-CY-P302)
Protocol Number:	Protocol No. TNX-CY-P302 (21 November 2018); Amendment 1, Dated 02 January 2019; Amendment 2, Dated 19 March 2019; Amendment 3, Dated 23 December 2019
Investigational Product:	Tonmya®/TNX-102 SL (cyclobenzaprine HCl sublingual tablets), 2.8 mg
Dose	5.6 mg taken daily at bedtime (administered as two 2.8 mg sublingual tablets)
Phase:	3
Sponsor:	Tonix Pharmaceuticals, Inc. [redacted]
SAP Author:	[redacted]
SAP Version:	Final 2.0
SAP Date:	23 October 2020

CONFIDENTIAL

DOCUMENT HISTORY

[redacted]

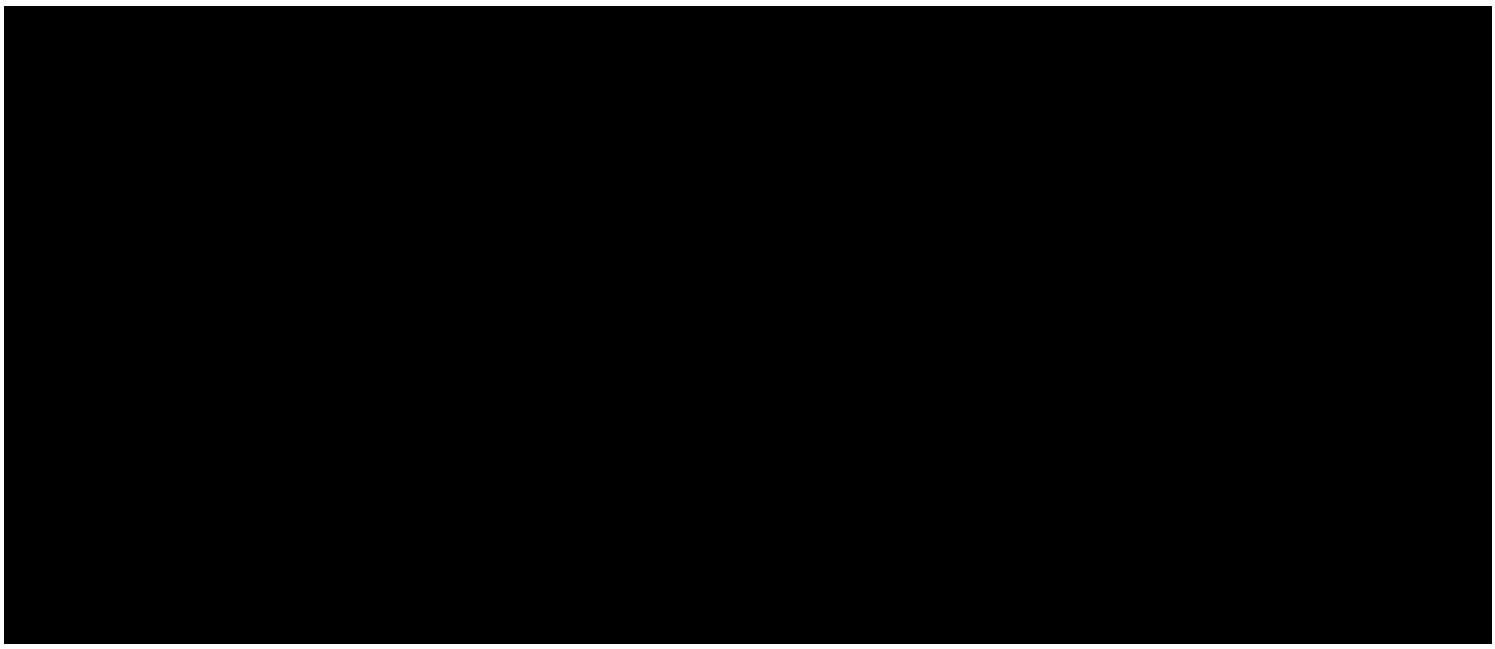

TABLE OF CONTENTS

ABBREVIATIONS

ABBREVIATION	DEFINITION OR DESCRIPTION
AE	Adverse Event
BMI	Body Mass Index
CA-AF	Criterion A – Assessment Form
CAPS-5	Clinician Administered PTSD Scale (for DSM-5)
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression- Severity
CI	Confidence Interval
CMH	Cochran Mantel Haenszel
CM	Clinical Manager-I
CRF	Case Report Form
CRO	Contract Research Organization
CSFQ-14	Changes in Sexual Functioning Questionnaire Short Form
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
ECG	Electrocardiogram
EDC	Electronic Data Capture
e.g.	*Exempli gratia* (for example)
EMA	European Medicines Agency
ET	Early Termination
FDA	Food and Drug Administration
HCl	Hydrochloride
ICH	International Council for Harmonization
IDMC	Independent Data Monitoring Committee
i.e.	*id est* (that is)
IP	Investigational Product
LEC-5	Life Events Checklist for DSM-5
LOE	Lack of Efficacy
LS	Least Squares
LTFU	Lost to follow-up
MAR	Missing at Random
MADRS	Montgomery-Åsberg Depression Rating Scale
MCMC	Markov Chain Monte Carlo
MDE	Major Depressive Episode
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MINI	Mini International Neuropsychiatric Interview
mITT	Modified Intent-to-Treat
MMRM	Mixed Model Repeated Measures
MNAR	Missing Not at Random
N, n	Number (of participants)

NIH	National Institutes of Health
RHNPCOT	Randomization Honoring Non-parametric Combination of Tests
PGIC	Patient Global Impression of Change Scale
PI	Principal Investigator
PROMIS	Patient-Reported Outcome Measurement Information System
PTSD	Posttraumatic Stress Disorder
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS®	Statistical Analysis System
SD	Standard deviation
SDS	Sheehan Disability Scale
SE	Standard Error
SL	Sublingual
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
THC	Tetrahydrocannabinol
TNX-102 SL	Cyclobenzaprine HCl sublingual tablets
WHO-DD	World Health Organization Drug Dictionary
WOC	Withdrawal of consent

1. OVERVIEW

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for protocol TNX-CY-P302: A Phase 3, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study to Evaluate the Efficacy and Safety of TNX-102 SL in Participants with PTSD Taken Daily at Bedtime, dated 21 November 2018; Amendment 1, Dated 02 January 2019; Amendment 2, Dated 19 March 2019; Amendment 3, Dated 23 December 2019. TNX-CY-P302 is intended to be a pivotal efficacy study. [redacted]

The structure and content of this SAP provides sufficient detail to meet the requirements identified by Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Council for Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials. All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association and the Royal Statistical Society, for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc, or unplanned exploratory analysis performed will be clearly identified as such in the final CSR.

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

The reader of this SAP is encouraged to also read the clinical protocol, and other identified documents, for details on the planned conduct of this study. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

2. STUDY OBJECTIVES AND ENDPOINTS

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective is to evaluate the efficacy of TNX-102 SL (Tonmya®) 5.6 mg (2 x 2.8 mg tablets) taken at bedtime in the treatment of posttraumatic stress disorder (PTSD).

2.1.2 Secondary Objectives

The secondary objective is to evaluate the safety of TNX-102 SL (Tonmya®) 5.6 mg (2 x 2.8 mg tablets) taken at bedtime in the treatment of PTSD.

2.2 Study Endpoints

2.2.1 Efficacy Endpoints

The primary efficacy endpoint is:

- The mean change from baseline (Visit 2) in the total CAPS-5 score after 12 weeks of treatment evaluated at Visit 6. The primary efficacy comparison will be the change from baseline in CAPS-5 total score for the 5.6 mg treatment arm compared to placebo.

Key secondary efficacy endpoints for labeling purposes include:

- Clinical Global Impression – Severity (CGI-S) score analyzed as a continuous variable (1-7) after 12 weeks of treatment comparing the 5.6 mg treatment arm to placebo.
- Change from baseline in Sheehan Disability Scale (SDS) total score after 12 weeks of treatment comparing the 5.6 mg treatment arm to placebo.
- Change from baseline in participants' quality of sleep using the Patient-Reported Outcome Measurement Information System (PROMIS) Sleep Disturbance (form 8a) scale after 12 weeks of treatment.

Other secondary efficacy endpoints include:

- Change from baseline in the disruption of work/school activities assessed using the SDS after 12 weeks of treatment.
- Change from baseline in disruption of social life/ leisure activities assessed using the SDS after 12 weeks of treatment.
- Change from baseline in the disruption of family life/home responsibilities assessed using the SDS after 12 weeks of treatment.
- Change from baseline in CAPS-5 Arousal and Reactivity (Criterion E) score after 12 weeks of treatment.
- Change from baseline in CAPS-5 Intrusion symptoms (Criterion B) score after 12 weeks of treatment.

- Change from baseline in CAPS-5 Negative Cognition and Mood (Criterion D) score after 12 weeks of treatment.
- Change from baseline in CAPS-5 Persistent Avoidance (Criterion C) score after 12 weeks of treatment.
- Change from baseline in CAPS-5 Sleep Disturbance (item E-6) score after 12 weeks of treatment.
- Change from baseline in Clinical Global Impression – Improvement (CGI-I) score after 12 weeks of treatment.
- Patient Global Impression of Change Scale (PGIC) score after 12 weeks of treatment (as a continuous variable).
- Proportion of participants with a PGIC score of "much improved" or "very much improved" after 12 weeks of treatment.
- Proportion of CGI-I Responders, defined as a CGI-I score of 2 ("much improved") or 1 ("very much improved"), after 12 weeks of treatment.
- Proportion of participants with a CAPS-5 total score of 0 –10 (asymptomatic/few symptoms) after 12 weeks of treatment.
- Proportion of participants with a CAPS-5 total score of 0-22 (asymptomatic or mild PTSD/subthreshold) after 12 weeks of treatment.
- Proportion of participants with Response, defined as a ≥ 10-point improvement from baseline in CAPS-5 total score after 12 weeks of treatment.
- Proportion of participants with Loss of Diagnosis, defined as Response AND no longer meeting DSM-5 symptom criteria in any one or more of the four clusters (B, C, D, E) after 12 weeks of treatment.
- Proportion of participants in Remission, defined as Loss of Diagnosis AND Total CAPS-5 score ≤ 10 after 12 weeks of treatment.
- Proportion of participants achieving sustained Remission, defined as Loss of Diagnosis AND Total CAPS-5 score ≤ 10 after both 8 weeks AND 12 weeks of treatment.
- Proportion of participants with a ≥ 30% reduction from baseline in CAPS-5 total score after 12 weeks of treatment.
- Proportion of participants with a ≥ 50% reduction from baseline in CAPS-5 total score after 12 weeks of treatment.
- Change from baseline in in Montgomery-Åsberg Depression Rating Scale (MADRS) score after 12 weeks of treatment.
- Change from baseline in CAPS-5 Reckless or Self-Destructive Behavior (item E-2) score after 12 weeks of treatment.

- Changes from baseline in patient-rated Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14) in males and in females, analyzed separately

2.2.2 Exploratory Efficacy Endpoints

All time points of the above outcomes will be reported, but the primary time point of interest is Week 12.

Additionally, exploratory analyses of the CAPS-5 questions will be performed using non-parametric combination of test analyses, and, separately, a 14-item total CAPS-5 score.

2.2.3 Safety Endpoints

Safety will be assessed by:

- Adverse events (AE) and serious AEs (SAEs) throughout the entire duration of the study, including detailed assessment of AEs involving the oral cavity.
- Changes from baseline in clinical laboratory test results.
- Changes from baseline in vital signs and weight.
- Change from baseline in MADRS score.
- Changes from baseline in suicidal ideation or behavior as reported on the Columbia Suicide Severity Rating Scale (C-SSRS).
- Changes from baseline in participant-rated CSFQ-14.

3. OVERALL STUDY DESIGN AND PLAN

This is a Phase 3, 12-week, randomized, multicenter, double-blind, placebo-controlled, fixed dose study that will investigate the efficacy and safety of TNX-102 SL 5.6 mg taken daily at bedtime for the management of PTSD. This study is to be conducted at approximately 30 investigational sites in the United States.

The total duration of this study over 6 visits will be approximately 13-17 weeks, with 12 weeks of treatment duration and a variable length of screening. This study will consist of a Screening visit (Visit 1, Days −21 to −7), Randomization visit (Visit 2, Day 1), telephone visit (Visit 3, Week 2) and three monthly in-clinic visits (Visits 4, 5 and 6 at Weeks 4, 8, and 12, respectively).

Eligible participants who provide written informed consent to participate will have study assessments performed at Screening, including counseling regarding any required washout and instructions to refrain from use of excluded medications throughout the study.

The length of the pre-randomization screening period is variable (7 to 21 days) in order to accomplish wash-out of previous medications, where appropriate, and allow for return of clinical laboratory results necessary to confirm eligibility. If additional screening time is required, medical monitor approval must be granted for any extensions of the screening interval. At Visit 2, participants will return to the site for additional baseline assessments and randomization (Day 1), when, if they meet all study randomization criteria, they will be randomly assigned in a 1:1 ratio to receive placebo or TNX-102 SL 5.6 mg. A dynamic randomization procedure will be employed at Visit 2 to minimize trial-wide imbalances between the treatment groups for site and sex.

Participants will take the study drug sublingually daily at bedtime, starting on the day that they are randomized and continuing for 12 weeks. Participants will return to the study center for safety and efficacy assessments at Weeks 4, 8, and 12 (or early termination [ET]).

3.1 Selection of Study Population

For a complete list of inclusion and exclusion criteria please refer to the initial protocol issued 21 November 2018; Amendment 1, Dated 02 January 2019; Amendment 2, Dated 19 March 2019; and Amendment 3, Dated 23 December 2019.

3.2 Method of Treatment Assignment and Randomization

Once all pre-randomization assessments have been completed, only those participants who meet all of the following randomization criteria will be eligible to continue:

1. Participant continues to meet all inclusion and exclusion criteria, including urine and blood test results, and
2. Visit 2 CAPS-5 total score ≥ 33 (“Symptom Severity” version using 1-week recall), and
3. No active suicidal intent or plan, or suicidal behavior, based on Investigator’s judgment and Visit 2 C-SSRS responses (e.g., no C-SSRS Type 4-5 ideation or

suicidal behavior since Visit 1) and/or a score of > 4 on Item 10 of the Visit 2 MADRS.

If the participant meets all study randomization criteria, they will be randomly assigned in a 1:1 ratio to receive TNX-102 SL 5.6 mg or placebo for 12 weeks. A dynamic randomization procedure will be employed at Visit 2 to minimize trial-wide imbalances between the treatment groups for site and sex.

Treatment A: 2 x TNX-102 SL 2.8 mg tablets ("TNX-102 SL") to be taken sublingually once daily at bedtime.
Treatment B: 2 x placebo tablets ("placebo") to be taken sublingually once daily at bedtime.

3.3 Treatment Blinding

This is a double-blind study. Unless otherwise specified, all study personnel are to remain blinded to study drug. Treatment assignments will not be revealed until all participants have completed the study and the database has been finalized and closed. Completion of study may occur due to early stop based on interim analysis results (for futility); initial results may be released at that time, followed by a full final analysis for the clinical study report.

If AEs occur that are considered to be intolerable, the investigator must decide whether it is necessary for the participant to discontinue study drug; however, the investigator should not be unblinded unless it is imperative for the participant's overall safety to determine whether the participant received active study drug (e.g., in the event of overdose of investigational product).

3.4 Minimization of Missing Data / Dropout rate

It is important to avoid missing data from clinical trials. The following strategies are designed to minimize drop-outs and missing data in this study:

- Providing participants with greater background on the nature of placebo-controlled clinical trials and explaining that completing this study, regardless of the participant's level of treatment response, is essential to understanding whether TNX-102 SL may be helpful to others in the treatment of PTSD. Sites will explain that the study is not necessarily designed to benefit the individual participant but, rather, it will provide useful information for future therapeutics.
- Minimizing the burden on participants, with visits scheduled generally every 4 weeks (with reasonable visit window flexibility). Clinical site personnel will also receive guidance regarding the special considerations necessary when dealing with participants with PTSD.
- Training of site personnel on the importance of minimizing missing data.

- Providing payment for participants' time and effort at clinic visits, based on the duration of assessments and as approved by the Institutional Review Board. Reimbursement for travel expenses to and from the clinic sites may also be provided to further minimize the financial burden of participating in the study.
- Utilizing repeated assessments of outcome measures and analytical approaches that most appropriately compensate for missing data.

4. ANALYSIS AND REPORTING

4.1 Interim Analysis

An interim analysis will be performed when approximately 50% of the initially planned enrollment is evaluable for efficacy assessments. This interim analysis will be performed by an unblinded team separate from the team responsible for the conduct and analysis of the study. An Independent Data Monitoring Committee (IDMC) will review the data and recommend to the Sponsor to increase the sample size by a fixed amount, keep the current sample size and continue, or stop the study early for futility in the event the conditional power is <20%. If the IDMC recommends an increase in sample size, the maximum increase will be limited to 120 additional participants in blocks of 10 participants. If the conditional power is at or over 85% with the current sample size, then the IDMC will recommend continuing at the current sample size; if it is between 20% and <85%, the IDMC will recommend increasing the sample size. The increase will be chosen with the goal of increasing the sample size by the minimum number of subjects sufficient to raise the conditional power to at least 90%. If the conditional power at the current sample size is over 20% and the maximum increase of 120 subjects falls short of raising the conditional power to at least 90%, the committee will recommend an increase of 120 subjects. No further information will be provided to the sponsor beyond this recommendation.

The interim analysis will be conducted approximately 16 weeks after randomization of approximately 125 participants, i.e. when approximately 50% of the initially planned participant enrollment (250) is evaluable for efficacy assessments (approximately 12 weeks); and related data cleaning, database freeze, and administrative tasks have been completed on this cohort (approximately 4 weeks).

In case the study is stopped early for futility, participants will be managed according to a study termination plan.

In the case of an early stop for futility, initial results may be released prior to full database lock; however, clinical study report (CSR) results will be based on the final analysis described below.

Conditional power will be calculated using the following formula (Mehta, 2010):

$$CP(Z_1,\check{n}_2) = 1 - \Phi\left(\frac{Z_\alpha\sqrt{n_2} - Z_1\sqrt{n_1}}{\sqrt{\check{n}_2}} - \frac{Z_1\sqrt{\check{n}_2}}{\sqrt{n_1}}\right)$$

Where Z_1 is the value of the Z score at the interim, n_1 is the actual sample size at the interim, n2 is the total planned sample size (250 subjects), and ň2 is the total planned sample size minus the actual sample size at the interim.

4.2 Final Analysis

All final, planned analyses will be performed after the last participant has completed all study assessments, all relevant study data have been processed and integrated into the analysis database, and the database has been locked.

Any post-hoc, exploratory analyses completed to support planned study analysis, which were not identified in this SAP, will be documented and reported in appendices to the CSR. Any results from these unplanned analyses (post-hoc) will also be clearly identified in the text of the CSR.

5. SAMPLE SIZE DETERMINATION

The study is planned to enroll approximately 250 patients total in a 1:1 randomization, that is, 125 patients in each of the TNX-102 SL 5.6 mg and placebo arms. [redacted]

The interim analysis will re-evaluate these assumptions and the IDMC may recommend a sample size increase.

6. ANALYSIS POPULATIONS

The following analysis populations are planned for this study:

- **Safety Population (SAFETY):** All participants who receive at least 1 dose of study drug. Participants who are issued study drug, but return 100% of it (i.e., none consumed) will be excluded from the safety population; likewise, participants that have no follow up following receipt of study drug to indicate they took drug are excluded. All safety analyses and demographic/baseline characterization will be performed using this population, analyzed as treated; participants will be summarized under TNX-102 SL 5.6 mg if they were issued it at any visit.
- **Modified Intention-to-Treat Population (mITT)**: All randomized participants who have at least a baseline and one post-baseline CAPS-5 assessment. This is the primary population for efficacy analyses and participants will be analyzed based on their randomized treatment.

7. GENERAL ISSUES FOR STATISTICAL ANALYSIS

7.1 General Statistical Methodology

Descriptive summaries will be provided where appropriate for each of the primary and secondary variables. In general, tables will summarize data by treatment group and visit.

Baseline characteristic and safety tables will be completed for the Safety Population unless otherwise specified. Efficacy tables will be presented for the mITT Population.

Continuous, quantitative, variable summaries will include the number of participants (N) with non-missing values, mean, standard deviation (SD), median, minimum, and maximum.

Categorical, qualitative, variable summaries will include the frequency and percentage of participants who are in the particular category. In general, the denominator for the percentage calculation will be based upon the total number of participants in the study population for the treatment group unless otherwise specified.

Baseline values are defined as the last non-missing measurement prior to the first dose of study drug. Change from baseline will be defined as the post-baseline visit value minus the baseline value.

ET participants will only be followed if required to monitor an on-going adverse event or other condition.

Participants are not supposed to change study sites; however, should this be necessary to allow a participant to continue in the study, participants will be analyzed under the site where they were initially enrolled.

Study day after first dose is defined as assessment date – first dose date +1. Dates prior to first dose are defined as assessment date – first dose date.

All analyses will be performed using Statistical Analysis System (SAS®) Software version 9.4 or later.

Unless otherwise noted, 95% confidence intervals will be presented for statistical tests.

For the purposes of analyses that include a site covariate, sites that have two or fewer subjects will be pooled into a single site. If there are further problems with model convergence, other small sites causing issues will be included in this pooled site.

7.1.1 Adjustments for Multiplicity and Other Alpha Control

7.1.1.1 Multiplicity

To adjust for multiplicity and to control for overall type I error, a fixed sequence procedure will be applied to key secondary efficacy endpoints. If the primary analysis produces a result that is statistically significant at a given level, the key secondary endpoints will be tested in an ordered fashion at the same level. If the analysis for a secondary endpoint does not produce a statistically significant result, then the remaining secondary endpoint analyses will automatically be considered non-significant regardless of the p-value produced. In the case of an early stop or increase in sample size, the same

methodology/critical value used for the primary endpoint will be applied to the key secondary endpoints for the purposes of the sequential testing.

The order in which the key secondary endpoints, intended for labeling purposes, will be tested is as follows:

i. Change from baseline in CGI-S score after 12 weeks of treatment (as a continuous outcome)
ii. Change from baseline in the SDS total score after 12 weeks of treatment.
iii. Change from baseline in the PROMIS Sleep Disturbance instrument T-score after 12 weeks of treatment

No other adjustments for multiplicity will be made and other p-values displayed in the output will be considered for descriptive summary purposes only and will not be used for formal inference. Additional details regarding statistical analysis for the listed endpoints can be found in Section 9.2.

7.1.1.2 Sample Size Increase

In the case of a sample size increase, the same inverse-normal method described in the primary efficacy analysis will be applied to the key secondary analyses to adjust the reported p-value; these will be tested at two-sided [redacted] to account for the sample size increase.

7.1.2 Data Handling for Participants Who Discontinue Study Drug or Withdraw from the Study

Participants who withdraw/drop out from the study will have their ET visit data collected and included in the analysis based on the closest visit window (Week 4, 8 or 12). Visit windows will be assigned by splitting the periods between visits at the midpoint between the visits. If more than one record falls within the window, the one closest to the target date will be used in the analysis, with preference given to the scheduled visits in the case of equidistant visits.

7.1.3 Procedures to Identify Reasons for Early Termination/Discontinuation

Reason for withdrawal will strongly contribute to the primary imputation algorithm; as such, the sites and sponsor will use the following categories and procedures to ensure that the withdrawal reason selected for the subject on the discontinuation case report form (CRF) page accurately represents the reason for withdrawal.

7.1.3.1 Early Termination/Discontinuation Categories

The following categories for early termination/discontinuation are available to the sites:

Insufficient Therapeutic Response (aka Lack of Efficacy or 'LOE'): Patients who discontinue from the study due to the perception they are receiving little or no benefit from the Investigational Product (IP) and feel they must go on to some other option for therapy should be coded to Insufficient Therapeutic Response.

All patients are informed at the Screening Visit about the nature of double-blind, placebo-controlled clinical trials, including the fact that they may not receive any benefit from the

experimental treatment. It is emphasized that all information derived from their participation is important, whether or not they are experiencing improvement in their condition, and, via informed consent, they are agreeing to be compliant with the protocol for the 12 weeks regardless of symptom improvement. Therefore, in theory, LOE should rarely be utilized as a reason for discontinuing. However, occasionally real-world instances dictate that despite initially agreeing to complete all study visits and procedures, some patients feel they just cannot continue without some other therapeutic help and are therefore discontinued.

Occurrence of an Adverse Event ('AE'): Patients who discontinue from the study due to complaints of untoward medical events of a severity that interferes with their daily activities or causes much distress are coded to Occurrence of an AE. Patients who discontinue for this reason essentially are stating they would continue in the study if they were not experiencing the(se) adverse event(s).

Lost to follow-up ('LTFU'): Patients who miss study visits and are non-responsive to repeated phone calls, texts and/or emails requesting follow up should be coded to LTFU.

All patients are to receive at least 3 attempts at contact before the site sends a certified letter requesting the patient contact the site (and return study drug) before being deemed "Lost to Follow Up". Sites are to document in the source these 3 attempts made to contact the patient. Also, all sites are requested to ask patients at the Screening visit to provide the name and contact information of someone close to them whom the site may contact if they are having difficulty reaching the participant, to assist with successfully communicating with the participant. If the patient provided this information, this route is to be attempted before the certified letter. If the certified letter also fails to elicit a response from the participant, the site is expected to document the date the letter was sent, whether delivery and signature were confirmed, and that the patient is coded to LTFU.

Investigator Decision: Patients who are discontinued because the Principal Investigator (PI) thinks it is not in the best interest of the patient to continue study treatment should be coded to Investigator Decision. This category differs from AE in that there is not an identifiable treatment-emergent event that is driving this decision. An example would be if an investigator learns after randomization that the patient had not previously disclosed pre-existing medical condition that is contraindicated for inclusion, such as learning from medical records that do not arrive until after randomization that the patient has untreated or uncontrolled angle-closure glaucoma. This is a very rarely utilized reason for discontinuation.

Withdrawal of consent by patient ('WOC'): Patients who discontinue from the study due to not being able to continue with study visits because of life circumstances should be coded as discontinuing due to Withdrawal of Consent. The circumstances must be unrelated to perceived efficacy of the IP or to occurrence of AEs. Typical reasons in this category are the participant has to move out of range from the study site and return visits are no longer feasible, or the participant has had a life change, such as a new job, that does not allow time for study visits during available site hours. The burden is placed on the investigator to reasonably demonstrate that the discontinuation is not better accounted for by LOE or AE. Sponsor assessment of the appropriateness of WOC discontinuation

designation includes review of the participant's results efficacy measures, including CAPS-5 and PGIC, and recorded AE(s) occurring around the time of discontinuation. Any potential alternative explanation based on evidence of LOE or AE is presented to the investigator to challenge the categorization in WOC and encourage re-designation under LOE or AE if deemed more accurate and appropriate.

Patient Non-Compliance: Patients who are discontinued from the study due to not following study procedures, to a degree that they are impacting the integrity of the study, should be coded to Patient Non-Compliance. In most cases the patient has stopped taking their study drug.

Before choosing this option, the Investigator must determine that the non-compliance is not due to LOE or AE. This category should seldom be selected and should only be applied to real world instances where the patient is not following study procedures despite attempts at re-education by site staff.

Poor compliance with IP due to patient forgetting doses or other problems/circumstances that interfered with dosing would not be a reason to discontinue a participant. Rather, generally, all efforts are made to have patients with poor compliance remain in the study after re-education regarding daily dosing of IP and importance of not missing doses. Whereas outright refusal to comply with dosing by the patient would be considered a reason for discontinuation due to patient non-compliance after LOE and AE have been ruled out as primary causes of the non-compliance with dosing.

Pregnancy: Patients who become pregnant while in the study must be discontinued, with the reason for discontinuation coded to Pregnancy.

Other: Patients who discontinue for a reason that cannot be categorized under any of the above-mentioned categories should be coded to Other. Examples are the sponsor discontinued the study or the site can no longer participate in the study.

All other discontinuation reasons must be ruled out before "Other" can be selected.

7.1.3.2 Procedures for Reviewing Early Termination/Discontinuation Categories

- In the weekly Tonix-CRO clinical study team meeting, all new ETs are reviewed, and the CRO Clinical Manager (CM) is instructed by the Sponsor when follow-up is needed on specific patients
- The sponsor instructs our CRO CM to contact each site for any ET due to Investigator Decision, WOC, Patient Non-Compliance, and Other. The CM has a templated email for this purpose, specifically to ensure that the PI has carefully evaluated the ET reason, including an evaluation of AE and efficacy data to ensure the ET reason is reasonable. Within the email the participant's CAPS-5 and PGIC efficacy data is included as well as a listing of all the participant's AEs to assist their review.
- The Sponsor conducts a periodic review of the discontinuation listings to ensure any necessary change of category has been made and that the specific reason supporting the categorization has been documented.

7.2 Efficacy Assessments

There is one primary efficacy endpoint and three key secondary endpoints intended for the labeling. Other secondary and exploratory efficacy endpoints are tested to better understand the potential clinical benefit of TNX-102 SL on PTSD participants.

7.2.1 Primary Efficacy

7.2.2 Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)

The primary efficacy endpoint is the change from baseline in total CAPS-5 score at Week 12 (Visit 6).

The CAPS-5 is an updated and validated version of a semi-structured interview that has been designed to assess the essential features of PTSD as defined by the DSM-5 (Weathers et al, 2013). The CAPS-5 affords the clinician flexibility to inquire about symptoms and diagnostic status over different time frames, such as past week, past month, and/or worst month for lifetime. For this study, a "Diagnostic" version of the CAPS-5 (past month recall) will be utilized at the Screening visit to confirm the diagnosis of PTSD and determine eligibility. A "Symptom Severity" version of the CAPS-5 (past week recall) will be completed by the clinician at all other time points (Baseline [Visit 2] and after 4, 8, and 12 weeks of treatment [Visits 4, 5, and 6, respectively]).

The CAPS-5 interview contains the following components

- Life Events Checklist for DSM-5 (LEC-5): Completed by the participant at the Screening Visit (Visit 1 only)
- Criterion A – Assessment Form (CA-AF): this semi-structured interview is based on the Criterion A discussion included in the CAPS-5, but modified (by Dr. Frank Weathers, CAPS-5 lead author, in conjunction with the Sponsor P302 study team) specifically for this study to ensure sufficient review of the traumas outlined on the participant-completed LEC-5 and to capture clear description of the participant's index/qualifying trauma (Visit 1 only)
- PTSD symptoms, onset/duration, distress/impairment, global rating, and dissociative symptoms:
    - o Criterion B: Items 1-5 (Intrusion symptoms)
    - o Criterion C: Items 6-7 (Avoidance symptoms)
    - o Criterion D: Items 8-14 (Negative alterations in cognitions & mood)
    - o Criterion E: Items 15-20 (Arousal and reactivity symptoms)
    - o Criterion F: Items 21-22 (Onset and duration of symptoms)
    - o Criterion G: Items 23-25 (Distress and impairment due to PTSD)
    - o Global ratings: Items 26-27 (Validity, severity)
    - o Dissociative symptoms: Items 29 (Depersonalization) and 30 (Derealization)

In addition to the total symptom score (obtained from the sum of Criteria B-E), the CAPS-5 affords the opportunity to examine clusters of symptoms, including Criterion B (intrusion symptoms), Criterion C (persistent avoidance), Criterion D (negative

cognitions and mood), and Criterion E (arousal and reactivity), all of which will be secondary efficacy endpoints.

In the case of CAPS-5 items that are not resolvable at the site level (via query, review of source), the following conventions will be employed for handling of missing CAPS-5 items:

- For ≤ 2 missing item scores in Criterion B, D, or E, the average of the other items within the criterion will be used for the missing item;
- For < 2 missing items scores in Criterion C, which only has 2 items, the value of the non-missing items will be used for the missing item;
- For > 2 missing items in Criterion B, D, or E, or > 1 missing item in the Criterion C, the total CAPS-5 score will be considered missing overall.

7.2.3 Key Secondary Efficacy Assessments

7.2.3.1 Clinician Global Impression – Severity (CGI-S)

The CGI-S will be completed by an Investigator to evaluate the patient's severity at the time of the assessment. The CGI-S status is a key secondary efficacy endpoint in this study. The CGI-S should be completed toward the end of each in-clinic study visit (except Screening), once all of the assessments are available for the investigator's review. An Investigator will complete the CGI-S assessment at Baseline, Week 4, 8 and 12 (Visits 2, 4, 5 and 6, respectively) in order to assess the overall patient severity status, and answer the following question:

Considering your total clinical experience with this particular population, how mentally ill is the patient at this time?

1 = Normal, not ill at all
2 = Borderline mentally ill
3 = Mildly ill
4 = Moderately ill
5 = Markedly ill
6 = Severely ill
7 = Among the most extremely ill patients

The key secondary analysis of CGI-S will treat it as a continuous variable (scored 1-7 as above).

7.2.3.2 Sheehan Disability Scale (SDS)

The SDS scale is a self-report questionnaire that was designed to assess the participant's view of the degree to which symptoms have disrupted work/school, social life/leisure activities, and family life/home responsibilities during the previous two weeks (Sheehan & Sheehan, 2008). In addition, the SDS asks the participant to provide the number of days of work/school lost as well as unproductive days in the past two weeks. The SDS scale will be completed by the participant at Baseline and after 4, 8, and 12 weeks of treatment (Visits 4, 5, and 6, respectively). A total score will be calculated summing the three individual 11-point (0-10) scales. For participants that do not respond to the work/school disruption because the participant checked the box indicating he/she has not

worked or attended school for reasons unrelated to the disorder (PTSD), the total score will sum the other two domain questions and multiply by 1.5 (rounding up to the nearest whole number) to maintain the scale of 0-30 for the total score. Participants with missing values on the other items will be missing for the total.

7.2.3.3 PROMIS Sleep Disturbance Instrument

PROMIS refers to the Patient-Reported Outcome Measurement Information System (www.nihpromis.org), a National Institutes of Health (NIH)-funded initiative to develop instruments to be used across chronic conditions.

The PROMIS sleep disturbance scale (short form 8a) will be assessed at Baseline and after 4, 8, and 12 weeks of treatment (Visits 4, 5, and 6, respectively).

The sleep disturbance scale will be calculated summing the individual item scores of the 8 items. These summed scores will be transformed to T-scores using the published conversions (see Section 14.2).

7.2.4 Other Efficacy Outcomes

7.2.4.1 Clinician Global Impression of Improvement (CGI-I)

The CGI-I will be completed by an Investigator to evaluate the participant's status since initiation of treatment. The CGI-I, status since initiation of treatment, is a secondary efficacy endpoint in this study. The CGI-I should be completed toward the end of each in-clinic post-Baseline study visit once all other assessments are available for the investigator's review. Once the participant has been randomized, an Investigator completes the CGI-I assessment after 4, 8, and 12 weeks of treatment in order to assess the overall change in the participant's status and answer the following question:

Since the initiation of treatment at Visit 2, the participant is:

1 = Very much improved
2 = Much improved
3 = Minimally improved
4 = No change
5 = Minimally worse
6 = Much worse
7 = Very much worse

The secondary analysis of CGI-I will treat it as a continuous variable (scored 1-7 as above). There will be an additional secondary responder analysis presented where a responder on the CGI-I is defined as a participant who is scored as 'Very much improved' or 'Much improved' and compared between the study arms.

7.2.4.2 Patient Global Impression of Change (PGIC)

The PGIC is a validated, self-report instrument to gauge the participant's assessment of change in condition. This form will be completed by the participant after 4, 8, and 12 weeks of treatment (Visits 4, 5, and 6, respectively). The participant will answer a single question:

Since the initiation of study medication, my PTSD symptoms are:

1 = Very much improved
2 = Much improved
3 = Minimally improved
4 = No change
5 = Minimally worse
6 = Much worse
7 = Very much worse

A responder on the PGIC is a participant who scores himself as 'Very much improved' or 'Much improved.'

7.2.4.3 Montgomery-Åsberg Depression Rating Scale (MADRS)

The MADRS is a validated, 10-item, clinician-administered depression scale that has been used in clinical research since 1979 (Montgomery & Åsberg, 1979). The MADRS scale will be completed at Screening (Visit 1), Baseline (Visit 2) and after 4, 8 and 12 weeks of treatment (Visits 4, 5 and 6, respectively). The 10 items measure the core symptoms and cognitive features of clinical depression. This scale has historically been used in drug-treatment trials due to its particular sensitivity to detect treatment effects. Since there is a comparative lack of emphasis on somatic symptoms, the scale is useful for the assessment of depression in subjects with physical illness (as may be the case in subjects with PTSD).

7.2.4.4 CAPS-5 Additional Endpoint Definitions

7.2.4.4.1 Response

Response is defined as ≥ 10-point improvement from baseline in CAPS-5 total score.

7.2.4.4.2 Loss of Diagnosis

Loss of Diagnosis at any time point is defined as Response AND no longer meeting DSM-5 symptom criteria required for diagnosis on the CAPS-5 in any one or more of the four clusters (B, C, D, E).

7.2.4.4.3 Sustained Remission

Remission at any time point is defined as a Loss of Diagnosis and a CAPS-5 total score of ≤ 10. Remission that is sustained is defined by meeting remission criteria after both 8 AND 12 weeks of treatment.

A CAPS-5 total score of ≤ 10 was selected for remission based on Frank Weathers (National Center for PTSD) communication that empirical data suggested the CAPS-5 score range from 0 to 10 was "asymptomatic/few symptoms". In large multicenter pharmacotherapy trials of PTSD that utilized the prior 17-item CAPS for DSM-IV, remission was defined as achieving a score of < 20 (Krystal et al, 2011) or ≤ 20 (Davidson et al, 2006). Direct mathematical extrapolation of a score of 20 on prior CAPS version, considering the different scaling for intensity/frequency (0-4 on CAPS-5; 0-8 on CAPS for DSM-IV), would be ((remission score of 20/17-items CAPS for DSM-IV)/2

(scaling))*20-items CAPS-5) = 11.8 or about 12. Although <12 or ≤12 were considered as more comparable with these definitions from studies using the prior version of CAPS, the definition of ≤ 10 was ultimately chosen as the more stringent approach.

7.3 Safety Assessments

Safety will be assessed by:

- AEs and SAEs throughout the entire duration of the study, including detailed assessment of AEs involving the oral cavity.
- Changes from baseline in clinical laboratory test results.
- Changes from baseline in vital signs and weight.
- Change from baseline in MADRS (see Section 7.2.4).
- Changes from baseline indicative of increased suicidal ideation or behavior as assessed by the C-SSRS.
- Changes from baseline in participant-rated CSFQ-14.

7.3.1 Adverse Event and Prior/Concomitant Medication handling conventions

To handle missing or partial AE and prior/concomitant medication dates, the following rules will be applied.

For partial start dates:

1. If the year is unknown, then do not impute the date but assign a missing value.
2. If the month is unknown, then:
    a. If the year matches the year of the first dose date, then impute the month and day of the first dose date.
    b. Otherwise, assign “January.”
3. If the day is unknown, then:
    a. If the month and year match the month and year of the first dose date, then impute the day of the first dose date.
    b. Otherwise, assign “01.”

For partial end dates:

1. If the year is unknown, then do not impute the date but assign a missing value.
2. If the month is unknown, then assign “December.”
3. If the day is unknown, then assign the last day of the month.

Adverse events are categorized as either pre-treatment adverse events or treatment-

emergent adverse events based on the response to the CRF question "Did the AE start prior to the first dose?"

After implementing the rules above, to determine whether medications with missing start or stop dates are prior or concomitant medications, the following strategy will be used:

1. If the start date and stop date are both missing, then the most conservative approach is taken and the medication is considered to be a concomitant medication.
2. If the start date is missing but the stop date is not missing and is after the day of first study dose administration, then the most conservative approach is taken and the medication is considered to be concomitant.
3. If the start date is missing but the stop date is not missing and is on or before the day of first study dose and after the date of signed informed consent, then the medication is considered to be a prior medication.
4. If the start date is not missing but the stop date is missing, then the most conservative approach is taken and medication is considered to be concomitant (assuming that the start date is prior to ten days after last dose).

The missing severity of an AE will be imputed to "severe"; the missing relationship to study drug of an AE will be imputed to "related".

8. STUDY PARTICIPANTS AND DEMOGRAPHICS

8.1 Disposition of Participants and Withdrawals

The numbers and percentage of participants randomized, completing the study, and withdrawing from the study, along with reasons for withdrawal, will be tabulated overall and by treatment group. The number and percentage of participants in each analysis population will be reported. This summary will be based on all participants who have data entered into the database.

8.2 Protocol Deviations

Protocol deviations will be checked on complete data for all participants, determined during a (blinded) data review meeting before database lock, unblinding, and the final analysis.

Protocol deviations will be summarized by type and by treatment group for the Safety population. Protocol "violations" are not differentiated from deviations; instead, each deviation is identified either as "major" or "minor" depending upon its potential impact upon the integrity of the study data or the participant's well-being.

Individual participants with protocol deviations will be listed.

8.3 Demographics and Other Baseline Characteristics

Descriptive summaries of the demographic and other baseline characteristics will be completed for all participants in the safety population by treatment groups, unless otherwise specified.

Descriptive summaries of demographic and other baseline conditions will include:

- Demographics (age, gender, race/ethnicity, height, weight, body mass index (BMI), family status, education, presence of current major depressive episode (MDE), current nicotine, alcohol and THC usage, and employment status). In addition to the safety population, these will also be repeated for all randomized participants and the mITT population.
- Psychiatric history (MINI 7.0.2)
- Clinician administered scales
    - C-SSRS-Baseline/Screening Version
    - CA-AF
    - CAPS-5 Assessment (Diagnostic version)

Other assessment's baseline values will be reported with their respective follow-up measures.

Medical History will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 21.0 and summarized by System Organ Class (SOC) and Preferred Term using frequency counts by treatment group.

9. EFFICACY ANALYSES

9.1 Primary Efficacy Analysis Approach

The mean change from baseline in the CAPS-5 total score after 12 weeks of treatment in the TNX-102 SL 5.6 mg and placebo arms will be analyzed using a restricted maximum likelihood (REML) based repeated measures approach, including multiple imputation for missing values, accounting for the interim analysis. The model will include all participants in the mITT population. The dependent variable will be the observed change from baseline in CAPS-5 total score at each post-randomization visit.

9.1.1 Primary Efficacy Analysis

An adaptive group sequential design will be performed for sample size adjustment or futility stop. To assure independence of the stage wise test statistics, the first stage population is defined as all participants available for efficacy analysis in the interim analysis. The second stage population is defined as all participants randomized after the interim cut-off and participants randomized prior to the interim cut-off, who were not available for efficacy analysis in the interim analysis.

Participants that drop out or reach Week 12 by the time of the interim cut-off will be included into the interim analysis. If the study continues until full planned recruitment, participants who were randomized prior to the interim cut-off, but did not drop out or reach Week 12 by the time of the interim cut-off, will be included into the second stage test statistic. Participants that were randomized after the interim cut-off will be included in the calculation of the second stage test statistic. Note that in every case, only participants that meet the mITT population criteria will be included.

Data of different stages will be combined using the inverse normal method to test the null hypotheses that there is no difference in the change from baseline in CAPS-5 total score between TNX-102 SL and placebo treatment groups at Week 12 (Cui 1999):

$$Z_1 = \Phi^{-1}(1 - p_1)$$
$$and\ Z_2 = w_1 Z_1 + w_2 \Phi^{-1}(1 - p_2)$$

Where:

Z_1= the Z statistic for the first stage

Z_2= the combination test statistic at the end of the second stage

w_i = the weighting applied for each associated Z statistic

p_1= the first stage p-value

p_2 =the second stage p-value based on second stage participants

For maximum statistical efficiency, the weights are defined prospectively according to the square-root of the planned proportion of participants in the two stages, relative to the preplanned total enrollment of 250 participants, as $w_i = \sqrt{0.5}$. The calculation of these

weights is fixed and will not be changed due to unblinded data and is hence in line with the draft guidance on adaptive design clinical trials (CDER, CBER, February 2010). In order to control the type-I error, adaptive changes of the stage wise sample sizes will not lead to changes of the weights (Lehmacher & Wassmer 1999).

A Mixed Model Repeated Measures (MMRM) analysis will be performed for the change from baseline in CAPS-5 total score to estimate the treatment difference at Week 12. The model will include all participants in the mITT population, and the dependent variable will be the observed change from baseline in CAPS-5 total score at each post-randomization visit. Covariates in the model will include the fixed categorical effects of treatment, site, sex, and treatment by visit interaction, as well as the continuous fixed covariates of baseline score and baseline score by visit interaction. An unstructured covariance matrix will be used to model the within-subject variance-covariance errors. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Should the unstructured model fail to converge, an autoregressive AR(1) covariance structure will be attempted, then finally compound symmetric should AR(1) not converge.

Missing data for participants in the mITT population will be imputed via multiple imputation (MI) (Carpenter, 2013; Mallinckrodt, 2013). Ten repeats of the imputation will be performed using Markov Chain Monte Carlo (MCMC) assuming non-monotone missing. Where the study medication was discontinued due to lack of efficacy, lost to follow-up, or to an AE, missing values will be imputed drawing from the baseline values, conditioned on the non-missing post-baseline values, of all participants in the mITT population under the assumption that they are missing not at random (MNAR); covariates will include site and sex. If the study medication was discontinued for any other reason, values will be imputed within treatment group using MI under the assumption that they are missing at random (MAR) with covariates for site, sex, and the CAPS-5 score recorded at each time point (including baseline). The MAR approach will also be applied for sporadic missing values (prior to discontinuation).

The test statistic based upon the least squares (LS) mean treatment difference and associated p-value [redacted] will be presented for each stage.

In addition, the least squares means and 95% confidence intervals (CIs) will be calculated for the treatment difference between TNX-102 SL and placebo at each time point.

9.1.2 Sensitivity and Supporting Analyses

The primary analysis will be followed by two sensitivity analyses to investigate the impact of missing data on the treatment estimates. Additionally, a MAR mixed model will be reported for reference.

9.1.2.1 'Other' Dropouts as MNAR

This sensitivity analysis will be a repeat of the primary efficacy analysis, but including participants that have 'other' marked as the reason for drop out included with the dropouts reasons to be imputed under the MNAR assumption (imputed values based on the baseline distribution).

9.1.2.2 All dropouts as MNAR

This sensitivity analysis will be a repeat of the primary efficacy analysis, but imputing all dropouts under the MNAR assumption with values based on the baseline distribution.

9.1.2.3 MAR Mixed Model Repeated Measures

This analysis will be identical to the primary analysis, but will not include the MI stage described above.

9.1.3 Graphical Representations

The time course of response of the CAPS-5 total score will be presented graphically with time/weeks on the x-axis and the CAPS-5 total score on the y-axis. The means for each treatment group will be plotted over time.

The percent of participants reaching varying levels of improvement in CAPS-5 will be displayed in several manners. For each, the percent of participants will be out of all participants in the mITT population; thus, missing participants will be considered to have no improvement. All graphs will be produced with separate bars/lines for each treatment.

First, a continuous responder graph will be produced with percent improvement to Week 12 on the X-axis and the percent of participants achieving that level of improvement on the Y-axis. The X-axis will start at 0, so the Y intercept represents participants with any improvement; because the X-axis starts at zero and percentages are reported out of the mITT population, those with worsening/no improvement and those with missing data will be grouped together and their frequency represented by the Y intercept.

Secondly, a similar graph will be produced, but substituting the absolute change from baseline for percent improvement, again with X-axis starting at 0, so the Y intercept represents subject with any improvement; those with worsening/no improvement and those with missing data will be grouped together.

Additionally, a bar chart will be presented showing the percent of participants in 5-point groupings of Week 12 CAPS-5 scores. Percent of participants with missing data will be reported.

Finally, a bar chart will be presented showing the percent of participants in 5-point groupings of Week 12 change from baseline CAPS-5 scores. Those with worsening/no improvement and those with missing data will be presented as separate groups and the percent reported.

9.1.4 Exploratory Analyses of CAPS-5 scores

9.1.4.1 Randomization honoring non-parametric combination of tests

[redacted]

[redacted]

9.1.4.2 14-item total CAPS-5 score

[redacted]

[redacted]

9.2 Secondary Efficacy Analysis

9.2.1 Continuous outcomes

The analyses of continuous secondary endpoints will be analyzed using an MMRM model identical to the primary endpoint, however, there will be no multiple imputation applied. These endpoints will include:

- CGI-S score.
- Change from baseline in SDS total.
- Change from baseline in participants' quality of sleep using the PROMIS Sleep Disturbance scale T-scores.
- CGI-I score
- Change from baseline in the disruption of work/school activities assessed using the SDS.
- Change from baseline in disruption of social life/ leisure activities assessed using the SDS.
- Change from baseline in the disruption of family life/home responsibilities assessed using the SDS.
- Change from baseline in CAPS-5 Arousal and Reactivity (Criterion E).
- Change from baseline in CAPS-5 Intrusion symptoms (Criterion B).
- Change from baseline in CAPS-5 Negative Cognition and Mood (Criterion D).
- Change from baseline in CAPS-5 Persistent Avoidance (Criterion C).
- Change from baseline in CAPS-5 Sleep Disturbance (item E-6).
- Mean PGIC score

- Change from baseline in CAPS-5 Reckless or Self-Destructive Behavior (item E-2).
- Changes from baseline in patient-rated Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14) in males and in females, analyzed separately
- Change from baseline in MADRS score.

The model will be changed to include the corresponding baseline value of the secondary endpoints instead of baseline CAPS-5. Randomized participants who have a baseline and at least one post-baseline assessment for the endpoint being analyzed will be included in the model. To estimate the difference between the treatment arms in CGI-I and PGIC scores (which do not have baseline assessments), the same method will be used, however, with the observed baseline response and associated interactions removed from the list of covariates. All visits will be included in the analysis. P-values from these comparisons will be considered nominal (conclusions from the sequential testing algorithm will be reported separately).

For each outcome, descriptive statistics (mean, SD, median, minimum and maximum) will be reported as well as LS means, standard error (SE) and p-values comparing the treatment and placebo arms at each visit. The time point of primary interest for all the above outcomes will be 12 weeks. In addition, for the SDS, the frequency of participants not working or attending school due to reasons unrelated to PTSD will be reported overall and within each visit.

9.2.2 Categorical outcomes

Binary data over time will be analyzed using a Cochran Mantel Haenszel (CMH) test stratified by presence of current major depressive episode. Examples include the CGI-I and the PGIC, in which responders are defined as those who score '2' (much improved) or '1' (very much improved) on these 7-point scales. Participants with missing data will be analyzed as though they are non-responders. This will be reported for each visit with the primary time point of interest at Week 12 (with remission spanning data at weeks 8 and 12). Odds ratios for treatment contrasts along with two-sided 95% CIs will also be reported. Two-sided p-values for the test of no difference between the TNX-102 SL and placebo will be provided.

These endpoints will include:

- Proportion of participants with a PGIC score of "much improved" or "very much improved".
- Proportion of participants with a CGI-I score of "much improved" or "very much improved".
- Proportion of participants with a CAPS-5 total score of 0 –10 (asymptomatic/few symptoms).

- Proportion of participants with a CAPS-5 total score of 0-22 (asymptomatic or mild PTSD/subthreshold).
- Proportion of participants with Response, defined as a ≥ 10-point improvement from baseline in CAPS-5 total score.
- Proportion of participants with Loss of Diagnosis, defined as Response AND no longer meeting DSM-5 symptom criteria in any one or more of the four clusters (B, C, D, E).
- Proportion of participants in Remission, defined as Loss of Diagnosis AND Total CAPS-5 score ≤ 10.
- Proportion of participants achieving sustained Remission, defined as Loss of Diagnosis AND Total CAPS-5 score ≤ 10 after both 8 weeks AND 12 weeks of treatment.
- Proportion of participants with a ≥50% improvement from baseline in CAPS-5 total score.

10. SAFETY AND TOLERABILITY ANALYSIS

The safety analysis will be run on the Safety population. The analysis of safety assessments in this study will include summaries of the following safety and tolerability data collected for each participant:

- Adverse Events
- Clinical Laboratory Investigations
- C-SSRS
- Vital Signs
- Visual Examinations of Oral Cavity
- Electrocardiogram (ECG) Parameters (Obtained at Screening only)
- Changes from baseline in participant-rated CSFQ-14
- MADRS

Summaries of continuous parameters will include raw values and change from baseline, as appropriate. Listings of safety data will also be presented.

10.1 Adverse Events

All AEs, treatment-emergent adverse events (TEAEs), and SAEs will be coded using MedDRA, version 21.0.

TEAEs are defined as either new onset AEs with an onset at the time of or following the start of treatment through ten days after the last dose of study medication, or a recurrence of an AE (or medical history) present prior to randomization but increasing in severity, frequency or relationship at the time of or following the start of treatment

An AE summary table will be presented for the following:

- TEAEs by severity

- TEAEs leading to study drug discontinuation
- TEAEs by relationship
- SAEs
- Oral Cavity AEs

Summaries of incidence rates (frequencies and percentages) of individual AEs by MedDRA SOC and preferred term will be prepared. Such summaries will be displayed for all TEAEs, TEAEs by maximum severity, and TEAEs by strongest relationship to study drug.

Each participant will be counted only once within each summation level (SOC; preferred term). If a participant experiences more than one TEAE within each summation level only, the TEAE with the strongest relationship or the maximum severity, as appropriate, will be included in the summaries of relationship and severity.

Additionally, a summary of the Oral Cavity AE characteristics will be presented; this includes information such as temporal proximity to the dose, duration and whether the AE was present the following morning.

Adverse events of special interest pertaining to abuse potential (as defined in the January 2017 Assessment of Abuse Potential for Drugs – Guidance for Industry) will be analyzed separately by treatment group, SOC and preferred term, and displayed in tabular format.

In the AE data listings, all AEs will be displayed. AEs that are treatment-emergent will be flagged.

10.1.1 Adverse Events Leading to Discontinuation of Study Drug

A summary of incidence rates (frequencies and percentages) of TEAEs leading to discontinuation of study drug by treatment group, SOC, and preferred term will be prepared for the Safety population.

A data listing of AEs leading to discontinuation of study drug will also be provided, displaying details of the event(s) captured on the CRF.

10.1.2 Serious Adverse Events

A summary of incidence rates (frequencies and percentages) of SAEs by treatment group, SOC, and preferred term will be prepared for the Safety population. A data listing of SAEs will also be provided, displaying details of the event(s) captured on the CRF.

10.1.3 Deaths

A listing of deaths will also be provided for the Safety Population.

10.2 Clinical Laboratory Evaluations

Laboratory data include analyses for Chemistry and Hematology will be summarized by treatment and visit for the Safety Population. Descriptive summaries of actual values and changes from baseline will be presented by study visit and each treatment group. 95% confidence intervals will be presented for change from baseline for each visit and last on-treatment assessment. ET data will be mapped to the closest visit window as described in

Section 7.1.2, but will typically only be used if a valid assessment was not collected then (scheduled visits will usually be closer to the target visit day and will take precedence if the ET and scheduled visit are equidistant to the target visit day).

Laboratory values will be displayed in the data listings and those that are outside the normal range ("H" or "L") will be flagged, along with corresponding normal ranges. Values pre-defined as potentially clinically significant ("HH" or "LL") will also be flagged. For each laboratory analysis, shifts in assessments of abnormality from baseline to each scheduled time point will be presented in shift tables.

A by-participant listing of all clinical laboratory data will also be provided.

10.3 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is an instrument that measures suicidal ideation and behavior. Items measuring suicidal ideation and suicidal behavior are displayed in the table below. Frequency counts of yes/no responses to items below and whether any suicidal ideation or behavior is present will be summarized as described below.

The overall number of participants with lifetime and/or recent (past 6 months) suicidal ideation (by item, e.g., Ideation Types 1-5), and lifetime and/or recent (past 12 months) suicidal behavior (by item; e.g., Preparatory Act, Aborted Attempt, Interrupted Attempt, Actual Attempt and Completed Suicide), or non-suicidal self-injurious behavior at the screening and baseline visit will be summarized by visit and treatment group. Additionally, the overall number of participants with any suicidal ideation or behavior (by type and in total) or self-injurious behavior while on-treatment will be provided by treatment group. For ideation, participants will only be counted once at each visit and/or time frame at the worst case response for ideation type (1-5), where 1 is the least severe and 5 is the most severe type of ideation.

Category	Items
A) Suicidal Ideation	(1) Type 1: Wish to be dead (2) Type 2: Non-specific active suicidal thoughts (3) Type 3: Active suicidal ideation with any methods (not plan) without intent to act (4) Type 4: Active suicidal ideation with any some intent to act, without specific plan (5) Type 5: Active suicidal ideation with specific plan and intent
B) Suicidal Behavior	(1) Preparatory acts or behavior (2) Aborted attempt (3) Interrupted attempt (4) Actual attempt (5) Completed suicide Suicidal Behavior present (composite of items 1-5) Non-Suicidal Self-Injurious Behavior

Suicidal intensity of ideation will be calculated by tallying up the five intensity items to

create a total score ranging 0-25. If a participant does not have any suicidal ideation, a score of 0 will be given. Separate tables will be created for the entire Safety Population as well as the population comprising only participants exhibiting any suicidal ideation. Suicidal intensity of ideation total score will be summarized using descriptive statistics by treatment group.

A data listing of C-SSRS results will include only participants with suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent. For participants with suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent at any time, data from all visits are displayed.

10.4 Vital Signs

Descriptive summaries (mean, SD, median, minimum, and maximum) of actual values and changes from baseline at each assessment time point and last on-treatment assessment will be calculated for vital signs including weight, BMI, body temperature, pulse rate, systolic blood pressure and diastolic blood pressure. 95% confidence intervals will be presented for change from baseline.

These summaries will be presented by treatment and assessment time for the Safety population. ET data will be mapped to the closest visit window as described in Section 7.1.2, but will typically only be used if a valid assessment was not collected then (scheduled visits will usually be closer to the target visit day and will take precedence if the ET and scheduled visit are equidistant to the target visit day).

10.5 Physical Examination and Visual Examination of Oral Cavity

A full physical examination will be performed at Screening Visit 1 only. A data listing of the physical examination results will be performed.

A visual examination of the oral cavity will be assessed at Screening, Baseline, and Week 12. A visual examination is to be done any time a participant spontaneously reports an oral adverse event (aside from AEs known to be sensory-only, such as numbness, tingling or bitter taste) to confirm presence or absence of any signs of mucosal irritation or other visible abnormalities. A data listing of the visual examination results performed at each scheduled visit and at the time of an adverse event will be presented.

10.6 Electrocardiogram (ECG)

A 12-lead ECG will be performed only at the Screening Visit. The standard ECG parameters including rhythm, heart rate, and intervals for PR, QRS, QT and QTcF (Fridericia's) correction for heart rate will be recorded. A data listing of ECG results will be provided.

10.7 Assessment of Changes in Sexual Function: Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14)

The CSFQ-14 (Keller et al. 2006) is validated scale with internal reliability designed to allow a participant to self-evaluate his or her sexual behaviors or problems in several areas. The CSFQ-14 will be administered at Baseline (Visit 2) and Week 12/ET (Visit 6). It yields a total score, three subscales corresponding to phases of the sexual response cycle (i.e. desire, arousal, orgasm), and five subscales corresponding to important

dimensions of sexual functioning. It is considered a useful scale for assessing sexual side effects of medications. For all items, higher scores reflect higher sexual functioning. For 12 of the 14 items, higher sexual functioning corresponds to greater frequency or enjoyment/pleasure (e.g. 1=never to 5 = every day). For two items (item 10, assessing loss of interest after arousal for women and priapism for men, and item 14, assessing painful orgasm), higher sexual functioning corresponds to lower frequency (e.g. 1=every day to 5=never). Items 10 and 14 are included in the total score but not in any scale score.

Descriptive summaries (mean, SD, median, minimum, and maximum) of actual values and changes from baseline at each assessment time point will be reported.

11. MEDICATIONS

11.1 Concomitant Medication

All medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) Version B2 enhanced March 2018. Prior and concomitant medications will be summarized by treatment group and by the number and percentage of participants taking each medication, classified by using WHO-DD Anatomical Therapeutic Chemical Level 3 and preferred term.

Prior medications are defined as medications or therapies initiated prior to the start of the study drug and terminating prior to the start of study drug. Hence, these medications or therapies will have end dates prior to the first dose date of study drug. Concomitant medications are defined as any medications other than the study drug that a participant receives concurrently with the study drug. These medications will have end dates on or after the first dose date of the study drug until ten days after last dose date.

Prior and Concomitant medications will be summarized. All medications will be presented in a listing.

Participants' Lifetime Prior Psychiatric Medications (i.e., discontinued ≥ 60 days prior to the Screening visit) will also be summarized by treatment group and by the number and percentage of participants taking each medication, classified by using WHO-DD Anatomical Therapeutic Chemical Level 3 and preferred term.

Please refer to Section 7.3 for the conventions used to impute partial start dates and end dates of concomitant medications.

11.2 Exposure and Compliance

The treatment duration will be calculated as (number of days=last dose date - first dose date+1) and summarized. Participants that are lost to follow up or have a missing last dose date will be assigned a last dose date of the day before the last attended clinic visit for analysis purposes.

All pill count shortage (negative pill count) of greater than 4 doses (8 tablets) per 4-week assessment period and/or any between-visit compliance <70% will be considered protocol deviations, and the reason for the pill count discrepancy will be discussed with the participant and documented in the CRF to ensure that any cases of potential abuse or misuse are identified. Participants with more than one significant incidence of negative pill count discrepancy during the double-blind phase of the study will be closely monitored by the investigator.

Exposure will be measured using the last date of treatment and first date of treatment. Total exposure will be defined as the last date of treatment minus the first date of treatment plus one. The number of participants with total exposure by visit weeks (≤ 4 weeks (study Day 0 to 28), 4 to ≤ 8 (study Day 29 to 56), 8 to ≤ 12 (study Day 57 to 84) and > 12 weeks (study Day ≥ 85)) will be presented. Additionally, exposure days, defined as the total number of tablets taken divided by 2 (and rounded up to the nearest whole number), will be summarized.

Compliance will similarly be summarized across all study visits for each treatment arm. Study drug compliance as a percentage will be defined as the number of pills taken by the participant divided by the total number of pills that the participant was assigned to take multiplied by 100.

For calculating the compliance and exposure days, the following convention will be applied: for periods where drug is issued, but the participant does not return any bottles or tablets, or the participant misses the visit entirely, that period will not contribute to the compliance or exposure calculation unless the bottles are returned at a later visit.

Compliance will be summarized with descriptive statistics by treatment arm. The number and percentages of participants within certain categories of compliance e.g. < 50%, 50% to < 70%, 70% to ≤ 100%, greater than 100% will be presented. Compliance between 70% to 114% will not be considered a protocol deviation. Tablet counts, calculation of compliance overall and by visit, which bottles of study drug were returned at each visit, and participant-reported reasons for all tablet discrepancies at all visits will be presented in a listing.

In addition, participants with significant over-usage or otherwise unaccounted-for missing tablets resulting in >114% compliance for any visit (e.g., failure to return >8 tablets per 4 week interval) will be summarized by visit and overall. These listings will include both the participant-reported reason for the discrepancy plus the investigator's assessment of potential abuse, misuse or diversion. Tables will be created for higher level abuse-related MedDRA terms as listed in Appendix 4. Additional terms may be identified and included; note that somnolence is an anticipated drug effect and would only be considered indicative of potential abuse in the presence of other indicators of substance abuse. Clinically significant drug accountability discrepancies associated with missing medication, loss of drug, or cases in which the PI confirms concern over drug abuse, misuse or diversion, will be examined and discussed in the CSR.

12. CHANGES FROM PLANNED ANALYSIS

No changes from the protocol are planned at this time.

13. REFERENCES

Arboretti, R et al. Test statistics in medical research: traditional methods vs multivariate NPC permutation tests. Urology 2015; 85 (2): 130-136.

Carpenter, J. R. and Kenward, M. G. (2013), Multiple Imputation and Its Application, New York: John Wiley & Sons.

Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics. 1999;55(3):853-7

Davidson J, Rothbaum BO, Tucker P, Asnis G, Benattia I, Musgnung JJ. Venlafaxine extended release in posttraumatic stress disorder: a sertraline- and placebo-controlled study. J Clin Psychopharmacol. 2006;26(3):259-67.

Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), February 2010.

Guidance for Industry Assessment of Abuse Potential for Drugs, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), January 2017.

Keller A, McGarvey EL, Clayton AH. Reliability and construct validity of the changes in sexual functioning questionnaire Short form (CSFQ-14). J of Sex and Marital Ther. 2006; 32:43-52.

Krystal JH, Rosenheck RA, Cramer JA, Vessicchio JC, Jones KM, Vertrees JE, Horney RA, Huang GD, Stock C; Veterans Affairs Cooperative Study No. 504 Group. Adjunctive risperidone treatment for antidepressant-resistant symptoms of chronic military service-related PTSD: a randomized trial. JAMA. 2011;306(5):493-502.

Lehmacher, W., Wassmer, G. Adaptive sample size calculations in group sequential trials. Biometrics 1999;55: 1286-1290.

Mallinckrodt C, Roger J, Chuang-Stein C, Molenberghs G, Lane PW, O'Kelly M, Ratitch B, Xu L, Gilbert S, Mehrotra DV, Wolfinger R, Thijs H. (2013), Missing Data: Turning Guidance Into Action, Statistics in Biopharmaceutical Research, 5:4, 369-382

Montgomery SA and Asberg, M. A new depression scale designed to be sensitive to change. The British Journal of Psychiatry 1979; 134: 382-389.

Mehta C and Pocock S. Adaptive increase in sample size when interim results are promising: A practical guide with examples. Statistics in Medicine. 2010; 30: 3267–3284.

Pesarin F, Salmaso L. Permutation tests for complex data. theory, applications and software. Chichester: John Wiley & Sons, Ltd. ISBN: 978-0-470-51641-6; 2010: Chapter 4, "The Nonparametric Combination Methodology" 117-175.

Protocol TNX-CY-P302: A Phase 3, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study to Evaluate the Efficacy and Safety of TNX-102 SL in Participants with

PTSD Taken Daily at Bedtime, dated 21 November 2018; Amendment 1, Dated 02 January 2019; Amendment 2, Dated 19 March 2019; and Amendment 3, Dated 23 December 2019.

Sheehan KH & Sheehan DV. Assessing treatment effects in clinical trials with the Discan metric of the Sheehan Disability Scale. Int Clin Psychopharmacol 2008:23:70-83.

Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx, et al. The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). 2013: Instrument available from the National Center for PTSD at www.ptsd.va.gov.

14. APPENDICES

14.1 Appendix 1

The following reporting conventions will be adopted for the presentation of study data. These conventions will enhance the review process and help to standardize presentation with common notations.

14.1.1 General Reporting Conventions

- All tables and data listings will be developed in Landscape Orientation, unless presented as part of the text in a clinical study report (CSR).
- Figures will be presented in Landscape Orientation, unless presented as part of the text in a CSR.
- Legends will be used for all figures with more than one variable or item displayed.
- Figures will be in black and white, unless color figures have been identified as useful for discriminating presentation in the figure. Lines in figures should be wide enough to view the line after being photocopied.
- Specialized text styles, such as bolding, italics, borders, shading, superscripted and subscripted text will not be used in tables, figures, and data listings unless they add significant value to the table, figure, or data listing.
- Only standard keyboard characters should be used in tables and data listings. Special characters, such as non-printable control characters, printer specific, or font specific characters, will not be used on a table, figure, or data listing. Hexadecimal character representations are allowed (e.g., μ, α, β).
- All titles will be centered on a page. The ICH numbering convention is to be used for all tables, listings, and graphs (TLGs).
- All footnotes will be left justified at the bottom of a page. Footnotes must be present on the page where they are first referenced. Footnotes should be used sparingly and must add value to the TLG. If more than four footnote lines are planned then a cover page may be used to display footnotes.
- Missing values for both numeric and character variables will be presented as blanks in a table or data listing. A zero (0) may be used if appropriate to identify when the frequency of a variable is not observed.
- All date values will be presented as DDMMMYYYY (e.g., 29AUG2001) format. A four-digit year is preferred for all dates.
- All observed time values will be presented using a 24-hour clock HH:MM:SS format (e.g., 01:35:45 or 11:26). Seconds should only be reported if they were measured as part of the study.
- Time durations will be reported in mixed HHhr MMm SSs notation (e.g., 5h 32m, or 27h 52m 31s). The use of decimal notation to present (display) time

durations should be avoided (e.g. 0.083h = 5m) unless it is necessary to show the computation of time differences in a table, figure, or data listing, in which case both notations may be used to display the time duration.

- All TLGs will have the name of the program, location, programmer, and a date stamp on the bottom of each output.
- All analysis programs developed for a TLG display will be self-contained to facilitate transfer of programs to multiple computing environments and transfer to a regulatory agency (if requested).

14.1.2 Population Summary Conventions

- Population(s) represented on the tables or data listings will be clearly identified in the last title of the Table as "<name of population>" and will be identical in name to that identified in the protocol or SAP.
- Consistent terminology will be used to define and identify a population. Common nomenclature may include (a) All Participants, (b) mITT, and (c) Safety.
- Sub-population(s) or special population(s) descriptions will provide sufficient detail to ensure comprehension of the population (e.g., MITT >60 years of age) used for analysis in a table or figure.
- Population sizes may be presented for each treatment or dosing category as totals in the column header as (N=xxxx), where appropriate.
- Population sizes shown with summary statistics are the samples sizes (n) of Participants with non-missing values.
- All population summaries for categorical variables will include all categories that were planned and for which the Participants may have had a response. Percentages corresponding to null categories (cells) will be suppressed.
- All population summaries for continuous variables will include: n, mean, SD, minimum, and maximum. Other summaries (e.g. number missing, median, quartiles, 95% confidence intervals, CV or %CV) may be used as appropriate.
- All percentages are rounded and reported to a single decimal point (xx.x%). A percentage of 100% will be reported as 100%. A percentage of zero will be reported as 0.
- Population summaries that include *P* values will report the *P* value to three decimal places with a leading zero (0.001). All *P* values reported on default output from statistical software (i.e., SAS® Software) may be reported at the default level of precision. *P* values <0.001 should be reported as <0.001 not 0.000.

14.2 Appendix 2: PROMIS T-score Conversions

Sleep-Related Impairment 8a *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	30.0	5.4
9	35.1	4.6
10	38.7	4.2
11	41.4	3.8
12	43.6	3.6
13	45.5	3.4
14	47.3	3.1
15	48.9	2.9
16	50.3	2.7
17	51.6	2.6
18	52.9	2.6
19	54.0	2.5
20	55.1	2.5
21	56.1	2.5
22	57.2	2.5
23	58.2	2.4
24	59.3	2.4
25	60.3	2.4
26	61.3	2.4
27	62.3	2.3
28	63.3	2.3
29	64.3	2.3
30	65.3	2.3
31	66.3	2.3
32	67.3	2.3
33	68.4	2.3
34	69.5	2.4
35	70.7	2.4
36	71.9	2.5
37	73.3	2.6
38	75.0	2.8
39	76.9	3.1
40	80.0	3.9

*SE = Standard Error on T-score metric

14.3 Appendix 3: Seeds for Imputation:

The following is the list of seeds to be used for any imputations; they will be used in the order that they appear in the production dataset programming first for primary analyses, then secondaries in the order given in this SAP. In the instance that a sample size increase is needed, the post-interim imputations will use a seed that is 10 greater than the one used on the corresponding pre-interim code. If the list below is insufficient for the number of seeds required, the list will repeat with values 1 greater than those given.

549179445

374794564

446287595

823648962

455675309

884575352

293564245

608574838

795925865

935852284

14.4 Appendix 4: Abuse Terms

The following preferred terms will be used to identify AEs potentially related to abuse.

- Euphoric mood
- Elevated mood
- Feeling abnormal
- Feeling drunk
- Feeling of relaxation
- Dizziness
- Thinking abnormal
- Hallucination
- Inappropriate affect
- Psychosis
- Aggression
- Confusion
- disorientation
- Drug tolerance
- Habituation
- Drug withdrawal syndrome

- Overdose
- Misuse
- Abuse
- Diversion
- Seizure
- Dizziness and Somnolence