

# Statistical Analysis Plan

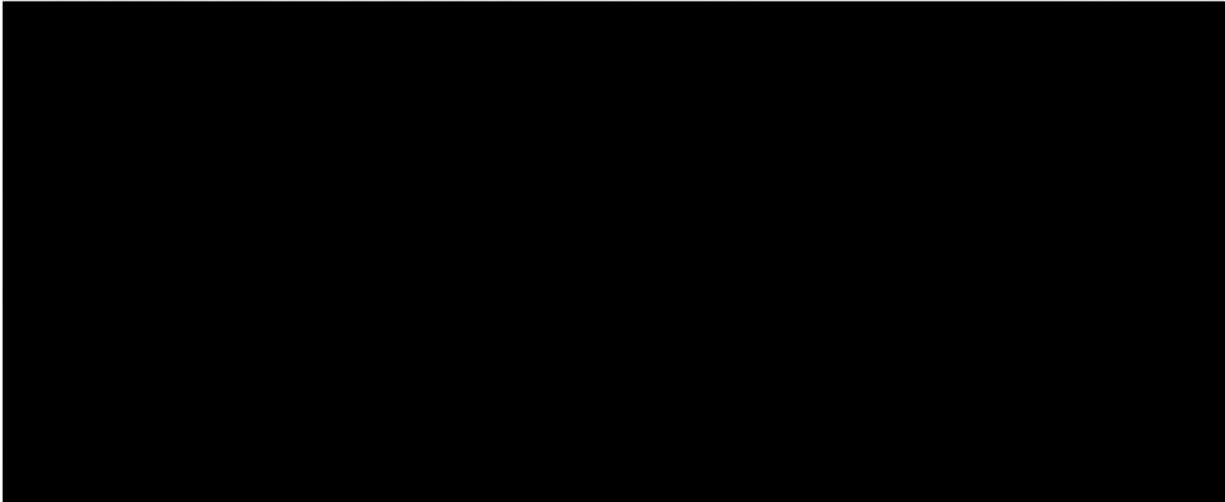
Protocol Title:	A Phase 3, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study to Evaluate the Efficacy and Safety of TNX-102 SL Taken Daily at Bedtime in Participants with Military-Related PTSD ('HONOR' Study)
Protocol Number:	Protocol No. TNX-CY-P301 (06 October 2016), Amendment 1, Dated 03 February 2017, Amendment 2, Dated 28 March 2017 and Amendment 3, Dated 18 May 2018
Investigational Product:	██████████ TNX-102 SL (cyclobenzaprine HCl sublingual tablets), 2.8 mg
Dose	5.6 mg taken daily at bedtime as two 2.8 mg sublingual tablets
Phase:	3
Sponsor:	Tonix Pharmaceuticals, Inc. 509 Madison Avenue, Suite 306 New York, NY 10022
SAP Author:	██████████ ██████████ ██████████
SAP Version:	Final 1.0
SAP Date:	16May2018

**CONFIDENTIAL**

### DOCUMENT HISTORY

Version	Date	Author	Description
1Draft	[REDACTED]	[REDACTED]	[REDACTED]
Final 1.0	16MAY2018	[REDACTED]	[REDACTED]

**SIGNATURE PAGE AND APPROVALS**



## TABLE OF CONTENTS

<b>DOCUMENT HISTORY .....</b>	<b>2</b>
<b>TABLE OF CONTENTS .....</b>	<b>4</b>
<b>ABBREVIATIONS.....</b>	<b>7</b>
<b>1. OVERVIEW .....</b>	<b>9</b>
<b>2. STUDY OBJECTIVES AND ENDPOINTS.....</b>	<b>10</b>
2.1 Study Objectives .....	10
2.1.1 Primary Objective .....	10
2.1.2 Secondary Objectives.....	10
2.2 Study Endpoints .....	10
2.2.1 Efficacy Endpoints.....	10
2.2.2 Exploratory Efficacy Endpoints.....	12
2.2.3 Safety Endpoints .....	12
<b>3. OVERALL STUDY DESIGN AND PLAN .....</b>	<b>13</b>
3.1 Selection of Study Population.....	13
3.2 Method of Treatment Assignment and Randomization .....	13
3.3 Treatment Blinding .....	14
3.4 Minimization of Missing Data / Dropout rate.....	14
<b>4. ANALYSIS AND REPORTING .....</b>	<b>16</b>
4.1 Interim Analysis.....	16
4.2 Final Analysis .....	16
<b>5. SAMPLE SIZE DETERMINATION.....</b>	<b>17</b>
<b>6. ANALYSIS POPULATIONS .....</b>	<b>18</b>
<b>7. GENERAL ISSUES FOR STATISTICAL ANALYSIS .....</b>	<b>19</b>
7.1 General Statistical Methodology.....	19
7.1.1 Adjustments for Multiplicity and Other Alpha Control.....	19
7.1.2 Data Handling for Participants Who Discontinue Study Drug or Withdraw from the Study.....	20
7.2 Efficacy Assessments.....	20
7.2.1 Primary Efficacy .....	21
7.2.2 Clinician-Administered PTSD Scale for DSM-5 (CAPS-5).....	21
7.2.3 Key Secondary Efficacy Assessments .....	22
7.2.4 Other Efficacy Outcomes.....	23

7.3	Safety Assessments.....	24
7.3.1	Adverse Event and Prior/Concomitant Medication handling conventions .....	24
<b>8.</b>	<b>STUDY PARTICIPANTS AND DEMOGRAPHICS.....</b>	<b>26</b>
8.1	Disposition of Participants and Withdrawals.....	26
8.2	Protocol Deviations.....	26
8.3	Demographics and Other Baseline Characteristics.....	26
<b>9.</b>	<b>EFFICACY ANALYSES .....</b>	<b>27</b>
9.1	Primary Efficacy Analysis Approach .....	27
9.1.1	Primary Efficacy Analysis .....	27
9.1.2	Sensitivity and Supporting Analyses .....	29
9.1.3	Graphical Representations .....	29
9.2	Secondary Efficacy Analysis .....	30
9.2.1	Continuous outcomes.....	30
9.2.2	Categorical outcomes.....	31
<b>10.</b>	<b>SAFETY AND TOLERABILITY ANALYSIS.....</b>	<b>32</b>
10.1	Adverse Events .....	32
10.1.1	Adverse Events Leading to Discontinuation of Study Drug.....	33
10.1.2	Serious Adverse Events .....	33
10.1.3	Deaths .....	33
10.2	Clinical Laboratory Evaluations .....	33
10.3	Columbia Suicide Severity Rating Scale (C-SSRS).....	33
10.4	Vital Signs.....	34
10.5	Physical Examination and Visual Examination of Oral Cavity.....	34
10.6	Electrocardiogram (ECG).....	35
10.7	Assessment of Sedation Morning Treatment-Related Sedation Score (MTRSS).....	35
10.8	Assessment of Changes in Sexual Function: Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14) .....	35
<b>11.</b>	<b>MEDICATIONS .....</b>	<b>37</b>
11.1	Concomitant Medication.....	37
11.2	Exposure and Compliance .....	37
<b>12.</b>	<b>CHANGES FROM PLANNED ANALYSIS.....</b>	<b>38</b>
<b>13.</b>	<b>REFERENCES.....</b>	<b>39</b>

<b>14.</b>	<b>APPENDICES</b> .....	<b>40</b>
14.1	Appendix 1.....	40
	14.1.1 General Reporting Conventions.....	40
	14.1.2 Population Summary Conventions .....	41
14.2	Appendix 2: PROMIS T-score Conversions.....	42
14.3	Appendix 3: Seeds for Imputation:.....	43

## ABBREVIATIONS

<b>ABBREVIATION</b>	<b>DEFINITION OR DESCRIPTION</b>
AE	Adverse Event
ADaM	Analysis Data Model
BDI-II	Beck Depression Inventory–II
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 from Initiation of Treatment
CI	Confidence Interval
CMH	Cochran Mantel Haenszel
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 (Version 4)
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (Version 5)
ECG	Electrocardiogram
EDC	Electronic Data Capture
e.g.	<i>Exempli gratia</i> (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.	<i>id est</i> (that is)
ISAP	Interim Statistical Analysis Plan
IWRS	Interactive Web Response System
LEC-5	Life Events Checklist for DSM-5
LS	Least Squares
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MDE	Major Depressive Episode
MedDRA	Medical Dictionary for Regulatory Activities
Military-Related	Related to service in any branch of the armed services (active or veteran), or as a military contractor
MI	Multiple Imputation
MINI	Mini International Neuropsychiatric Interview
mITT	Modified Intent-to-Treat
MMRM	Mixed Model Repeated Measures
MNAR	Missing Not at Random
MTRSS	Morning Treatment-Related Sedation Scale

N, n	Number (of participants)
NIH	National Institutes of Health
PGIC	Patient Global Impression of Change Scale
PROMIS	Patient-Reported Outcome Measurement Information System
PTSD	Posttraumatic Stress Disorder
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS <sup>®</sup>	Statistical Analysis System
SD	Standard deviation
SDS	Sheehan Disability Scale
SDTM	Study Data Tabulation Model
SE	Standard Error
SL	Sublingual
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TNX-102 SL	Cyclobenzaprine HCl sublingual tablets
WHO-DD	World Health Organization Drug Dictionary



## 1. OVERVIEW

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for protocol TNX-CY-P301: A Phase 3, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study to Evaluate the Efficacy and Safety of TNX-102 SL Taken Daily at Bedtime in Participants with Military-Related PTSD ('HONOR' Study), dated 06 October 2016, Amendment 1, dated 03 February 2017, Amendment 2, dated 28 March 2017 and Amendment 3, dated 18 May 2018. TNX-CY-P301 is intended to be a pivotal efficacy study.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by 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]
- [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 [REDACTED] 5.6 mg (2 x 2.8 mg tablets) taken at bedtime over 12 weeks of treatment using the total Clinician Administered PTSD Scale for DSM-5 (CAPS-5) score to assess military-related posttraumatic stress disorder (PTSD) symptom severity change from baseline to Week 12 (Visit 6).

#### **2.1.2 Secondary Objectives**

The secondary objective is to evaluate the safety of TNX-102 SL 5.6 mg taken at bedtime over 12 weeks of treatment.

### **2.2 Study Endpoints**

#### **2.2.1 Efficacy Endpoints**

The primary efficacy endpoint is:

- The mean change from baseline 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 total CAPS-5 score for the 5.6 mg treatment arm compared to placebo.

Key secondary efficacy endpoints for labeling purposes include:

- Clinical Global Impression – Improvement (CGI-I) 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.

Other secondary efficacy endpoints include:

- Change from baseline in participants' quality of sleep using the Patient-Reported Outcome Measurement Information System (PROMIS) Sleep Disturbance scale after 12 weeks of treatment.
- 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 Persistent Avoidance (Criterion C) 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 Sleep Disturbance (item E-6) score after 12 weeks of treatment.
- Change from baseline in CAPS-5 Exaggerated Startle (item E-4) score after 12 weeks of treatment.
- Patient Global Impression of Change Scale (PGIC) score after 12 weeks of treatment.
- 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 Total CAPS-5 score of 0 –10 (asymptomatic/few symptoms) after 12 weeks of treatment.
- Proportion of participants with a Total CAPS-5 score of 0-22 (asymptomatic or mild PTSD/subthreshold) after 12 weeks of treatment.
- Proportion of participants with Response, defined as a  $\geq 10$ -point improvement from baseline in Total CAPS-5 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  $\leq 10$  after 12 weeks of treatment.
- Proportion of participants achieving sustained Remission, defined as Loss of Diagnosis AND Total CAPS-5 score  $\leq 10$  after both 8 weeks AND 12 weeks of treatment.

- Proportion of participants with a  $\geq 50\%$  improvement from baseline in Total CAPS-5 score after 12 weeks of treatment.
- Change from baseline in Beck Depression Inventory–II (BDI-II) 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.

### **2.2.2 Exploratory Efficacy Endpoints**

Exploratory efficacy endpoints include:

- Change from baseline in the CAPS-5 item for depersonalization (CAPS-5 item 29) after 12 weeks of treatment
- Change from baseline in the CAPS-5 item for derealization (CAPS-5 item 30) after 12 weeks of treatment.

### **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 BDI-II.
- 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 Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14).
- Changes from baseline in participant-rated morning sedation as assessed by the Morning Treatment-Related Sedation Score (MTRSS).

### **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 military-related PTSD. This study is to be conducted at approximately 45 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 -35 to -7), Randomization visit (Visit 2, Day 0), telephone visit (Visit 3, Week 2) and three monthly in-clinic visits (Visits 4, 5 and 6 at Weeks 4, 8, and 12).

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 35 days) in order to accomplish wash-out of previous medications, where appropriate, and allow for return of clinical laboratory results necessary to confirm eligibility. At Visit 2, participants will return to the site for additional baseline assessments and randomization (Day 0), 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, sex, current tobacco/nicotine use (yes/no) and presence (yes/no) of current Major Depressive Episode (MDE) based on the Mini International Neuropsychiatric Interview, Version 7.0.2 (MINI 7.0.2) conducted at Visit 1.

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 06 October 2016, Amendment 1, dated 03 February 2017, Amendment 2, dated 28 March 2017 and Amendment 3, dated 18 May 2018.

#### **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 total CAPS-5 score  $\geq$  33 (“Symptom Severity” version using 1-week recall), and

3. At Visit 2, participant is not severely depressed (based on the investigator's judgment and Visit 2 C-SSRS, CAPS-5, and BDI-II, and considering the pre-randomization collaborative review by the investigator and medical monitor(s) of the participant's Visit 1 psychiatric history, presenting symptomatology, MINI 7.0.2, C-SSRS, CAPS-5, and BDI-II assessments).
4. No active suicidal intent or plan, 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).

If the participant meets all study randomization criteria, they will be randomly assigned in a 1:1 ratio to receive placebo or TNX-102 SL 5.6 mg for 12 weeks. A dynamic randomization procedure will be employed at Visit 2 to minimize trial-wide imbalances between the treatment groups for site, sex, current tobacco/nicotine use (yes/no) and presence (yes/no) of current MDE based on the MINI conducted at Visit 1.

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 (either for futility or efficacy); 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 designed to benefit the individual

participant but, rather, can only 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 military 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.
- Offering open-label continuation trial(s) for those successfully completing the double-blinded study and for randomized participants who have not yet completed the study if the study is stopped early for success.

## **4. ANALYSIS AND REPORTING**

### **4.1 Interim Analysis**

An interim analysis will be performed by a statistical group independent of both the sponsor and contract research organization (CRO) performing the final analysis when approximately 50% of the initially planned enrollment is evaluable for efficacy assessments. An 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, stop the study early for success, or stop the study early for futility. The ISAP, the IDMC charter and the Adaptive Study Simulation Report are being submitted together with this version of the SAP to the FDA.

The interim analysis will be conducted approximately 12 weeks after randomization of approximately 275 participants, i.e. when approximately 50% of the initially planned participant enrollment (550) is evaluable for efficacy assessments. These approximately 275 participants of the interim analysis will define the Early Stop Efficacy Population (see [Section 6](#)).

In case the study is stopped early for success, eligible participants who are not included in the early stop efficacy population will be immediately transitioned to an open label extension safety study.

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 or efficacy, initial results may be released prior to full database lock; however, CSR results will be based on the final analysis described below.

### **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. As described above, in the event of the interim analysis triggering an early stop for success, only those participants who are part of the early stop efficacy population will be analyzed, and eligible participants who are not included in the early stop efficacy population will be immediately transitioned to an open label extension safety study. Efficacy assessment will not be done prior to transition to TNX-CY-P306.

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 550 participants total in a 1:1 randomization, that is, 275 participants in each of the placebo and TNX-102 SL 5.6 mg groups.



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 (including participants that are rolled over to open-label TNX-102 SL 5.6 mg in the case of an early stop recommended by the IDMC during the planned interim analysis).
- **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.
- **Early Stop Safety Population (SAFETY):** This population is identical to the safety population but is limited to only the participants that are part of the interim analysis, (i.e. participants that are ongoing at the time of early stop are excluded). In the case of an early stop, safety results will be analyzed on both safety populations. Should the study continue to full enrollment (or full enrollment on an increased sample size), this population will not be utilized.
- **Early Stop Efficacy Population (mITT):** This population will be limited to those participants who are part of the interim analysis, i.e., among the approximately 275 patients defined as the population upon which the predefined interim analysis efficacy assessments will be conducted. Only those participants who have a baseline and at least one post-baseline CAPS-5 assessment will be evaluated for purposes of the interim analysis. Participants will be analyzed based on their randomized treatment. The dataset for this population will be defined by the completion or early termination of the last participant among this cohort.

## **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. Note that the protocol refers to “Day 0”; for the purposes of the datasets and analyses, this would appear as day 1.

All analyses will be performed using Statistical Analysis System (SAS<sup>®</sup>) Software version 9.3 or later.

Alpha levels are discussed in [Section 7.1.1.2](#) below; unless otherwise noted, 95% confidence intervals will be presented for statistical tests.

Presence of current tobacco/nicotine use and MDE is recorded first in the Interactive Web Response System (IWRS) during screening for purposes of randomization stratification and, after randomization, transcribed into the electronic data capture (EDC) system (as part of substance use history and the results for the MINI). The source for tobacco/nicotine use MDE in all analyses and statistical models will be the EDC data and not the IWRS.

#### **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. CGI-I score after 12 weeks of treatment (as a continuous outcome)
- ii. Change from baseline in the SDS total 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 Alpha Spend, Early Stops and Sample Size Increase

[REDACTED]

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 one-sided [REDACTED] to account for the first stage alpha spend.

#### **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 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.

#### **7.2 Efficacy Assessments**

There is one primary efficacy endpoint and two 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 military-related 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]).

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) 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  $\leq 2$  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 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 the first key 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 key secondary analysis of CGI will treat it as a continuous variable (scored 1-7 as above). There will be an additional outcome presented where a responder on the CGI-I is defined a participant who is scored as 'Very much improved' or 'Much improved' and compared between the study arms.

#### 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.

## 7.2.4 Other Efficacy Outcomes

### 7.2.4.1 PROMIS Sleep Disturbance Instrument

PROMIS refers to the Patient-Reported Outcome Measurement Information System ([www.nihpromis.org](http://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](#)).

### 7.2.4.2 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 BDI-II

The Beck Depression Inventory (BDI-II) is a 21-item measure of the severity of current depressive symptoms, extensively validated for use in both medical and mental health populations. While this instrument does not provide a psychiatric diagnosis of depression and has considerable overlap with PTSD associated symptoms, it does provide a continuous scale for measuring changes in the severity of symptomatology. The BDI-II will be completed by the participant at all visits.

### 7.2.4.4 CAPS-5 Additional Endpoint Definitions

#### 7.2.4.4.1 Response

Response is defined as  $\geq 10$ -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  $\leq 10$ . Remission that is sustained is defined by meeting remission criteria after both 8 AND 12 weeks of treatment.

A CAPS-5 total score of  $\leq 10$  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  $\leq 20$  (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  $((\text{remission score of } 20 / 17\text{-items CAPS for DSM-IV}) / 2 \text{ (scaling)}) * 20\text{-items CAPS-5} = 11.8$  or about 12. Although  $< 12$  or  $\leq 12$  were considered as more comparable with these definitions from studies using the prior version of CAPS, the definition of  $\leq 10$  was ultimately chosen as the more stringent approach.

### 7.3 Safety Assessments

Safety will be assessed by:

- AE 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 BDI-II (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.
- Changes from baseline in participant-rated morning sedation as assessed by the MTRSS.

#### 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.

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 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 19.0 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 total CAPS-5 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 and accounting for the two-stage design. The model will include all participants in the mITT population. The dependent variable will be the observed change from baseline in total CAPS-5 score at each post-randomization visit.

#### 9.1.1 Primary Efficacy Analysis

An adaptive group sequential design will be performed by an independent third party statistical group [REDACTED] to cover an early efficacy stop, sample size adjustment, or futility stop; the final CSR analyses in all cases will be performed by [REDACTED] (the primary biostatistics vendor). 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.

If the study is stopped early for efficacy, 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 cutoff will be offered an immediate transition into open label extension safety study (TNX-CY-P306) and will not be included in the efficacy analyses.

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:

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

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 550 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 total CAPS-5 score at each post-randomization visit. Covariates in the model will include the fixed categorical effects of treatment, site, sex, current tobacco use status, presence of current MDE, visit 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, sex, tobacco use, and presence of current MDE. 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, tobacco use, presence of current MDE, 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. The study may stop early for success in the case where the critical boundary is crossed (see Section 5.7 of the ISAP). Given an early efficacy stop, the remaining participants will be offered an immediate transition into an open label extension safety study (Protocol No. TNX-CY-P306) and will not be included in the efficacy analyses;

likewise ongoing subjects will not be included in final efficacy analyses in the case of a futility stop.

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 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 10-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 10-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.

## 9.2 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-I score.
- Change from baseline in SDS total.
- Change from baseline in participants' quality of sleep using the PROMIS Sleep Disturbance scale.
- 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).
- Change from baseline in CAPS-5 Exaggerated Startle (item E-4).
- Change from baseline in CAPS-5 Reckless or Self-Destructive Behavior (item E-2).
- Change from baseline in CAPS-5 Depersonalization (item 29).
- Change from baseline in CAPS-5 Derealization (item 30).
- PGIC score.
- Change from baseline in BDI-II 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 Total CAPS-5 score of 0–10 (asymptomatic/few symptoms).
- Proportion of participants with a Total CAPS-5 score of 0-22 (asymptomatic or mild PTSD/subthreshold).
- Proportion of participants with Response, defined as a  $\geq 10$ -point improvement from baseline in Total CAPS-5 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  $\leq 10$ .
- Proportion of participants achieving sustained Remission, defined as Loss of Diagnosis AND Total CAPS-5 score  $\leq 10$  after both 8 weeks AND 12 weeks of treatment.
- Proportion of participants with a  $\geq 50\%$  improvement from baseline in Total CAPS-5 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)
- Assessment of Morning Sedation (via MTRSS)
- Changes from baseline in participant-rated CSFQ-14
- BDI-II

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 19.0.

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

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.

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 and 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 analyzed with the closest visit that does not have a valid assessment value.

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 analyzed with the closest visit that does not have a valid assessment value.

#### 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 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 Sedation Morning Treatment-Related Sedation Score (MTRSS)**

The MTRSS is designed to assess the participant's one-week recall of morning sedation as assessed at clinic visits. The MTRSS will be assessed at Baseline (Visit 2) and at Weeks 4, 8, and 12 (Visits 4, 5, and 6). The participant will complete the following questions:

(A) On most mornings during the past week, I have awakened feeling:

0. Alert with no grogginess
1. Mildly groggy
2. Moderately groggy
3. Very groggy

(B) If you reported any grogginess in the last week (responses 1-3), the grogginess lasted:

1. A few minutes
2. A short time (less than one hour)
3. A long time (more than one hour)

Note: if "Alert with no grogginess" is selected in Part A, Part B is not completed, and the score for part B is 0

Results will be summarized with counts and percentages at each visit.

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

The CSFQ-14 ([Keller et al. 2006](#)) is a 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 September 2016. 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  $\geq 60$  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 ( $\leq 4$  weeks (study day 0 to 28), 4 to  $\leq 8$  (study day 29 to 56), 8 to  $\leq 12$  (study day 57 to 84) and  $> 12$  weeks (study day  $\geq 85$ )) 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, 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. 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**

Study day as it appears in data listings and the Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets will differ from the protocol: The protocol describes the day of first dose as "Day 0"; the listings and datasets will conform to the standard of first dose day appearing as "Day 1".

### 13. REFERENCES

- Adaptive Study Simulation Report for TNX-CY-P301 Final v2.0, 07 May 2018.
- 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.
- IDMC Charter for TNX-CY-P301 Final v3.0, 10 May 2018.
- Interim Statistical Report for TNX-CY-P301 Final v3.0, 16 May 2018.
- January 2017 Assessment of Abuse Potential for Drugs – Guidance for Industry Protocol No. TNX-CY-P301 (06 October 2016), Amendment 1, Dated 03 February 2017, Amendment 2, Dated 28 March 2017 and Amendment 3, Dated 18 May 2018
- Carpenter, J. R. & Kenward, M. G. (2013), Multiple Imputation and Its Application. Chichester, UK: Wiley.
- 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.
- 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
- Sheehan KH & Sheehan DV. Assessing treatment effects in clinical trials with the Discontinuity 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](http://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.,  $\mu$ ,  $\alpha$ ,  $\beta$ ).
- 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) ITT, (c) Safety, and (d) PP.
- 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<sup>®</sup> 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

<b>Raw Score</b>	<b>T-score</b>	<b>SE*</b>
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.

9529179445  
4148794564  
6635287595  
3215648962  
5558675309  
4529575352  
3242564245  
8884574838  
5364925865  
5666852284