TONIX PHARMACEUTICALS

TNX-CY-F304

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 PATIENTS WITH FIBROMYALGIA
"RELIEF STUDY"

Document Date: 16 OCTOBER 2020

NCT04172831

Statistical Analysis Plan

Protocol Title:	TNX-CY-F304 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 PATIENTS WITH FIBROMYALGIA (RELIEF STUDY)
Protocol Number:	Protocol No. TNX-CY-F304, Amendment 02 (27 March 2020)
Investigational Product:	TNX-102 SL (cyclobenzaprine HCl sublingual tablets)
Phase:	3
Sponsor:	Tonix Pharmaceuticals, Inc. 509 Madison Avenue, Suite 1608 New York, NY 10022
SAP Author:	[redacted]
SAP Version:	Final Version 2.0
SAP Date:	16 Oct 2020

CONFIDENTIAL

DOCUMENT HISTORY

Version	Date	Author	Description
1.0	[redacted]	[redacted]	[redacted]
1.0	[redacted]	[redacted]	[redacted]
2.0	16 Oct 2020	[redacted]	[redacted]

SIGNATURE PAGE AND APPROVALS

[redacted] Date

[redacted] Date

TABLE OF CONTENTS

ABBREVIATIONS

ABBREVIATION	DEFINITION OR DESCRIPTION
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical (WHO Drug Classification System)
BDI-II	Beck Depression Inventory II
BMI	Body Mass Index
CFB	Change from Baseline
CRF	Case Report Form
CSFQ-14	Changes in Sexual Functioning Questionnaire Short-Form
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
EMA	European Medicines Agency
ET	Early Termination
FDA	Food and Drug Administration
FIQR	Fibromyalgia Impact Questionnaire (Revised)
FM	Fibromyalgia
ICH	International Council for Harmonisation
IND	Investigational New Drug
ITT	Intent-to-Treat
IVRS	Interactive Voice Response System
LOE	Loss of efficacy
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model Repeated Measures
MNAR	Missing not at random
N	Number of subjects
NRS	Numeric Rating Scale
PGIC	Patient Global Impression of Change
PROMIS	Patient Reported Outcomes Measurement Information System
RHNP	Randomization Honoring Non-Parametric
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SF	Short Form
SOC	System Organ Class
TEAE	Treatment-emergent Adverse Event
TNX-102 SL tablets	Cyclobenzaprine HCl Sublingual Tablets
WHO	World Health Organization
WHO-DD	World Health Organization – Drug Dictionary

1. BACKGROUND

This Statistical Analysis Plan (SAP) for Protocol No. TNX-CY-F304 Study has been written [redacted] for the Phase 3 AFFIRM study (Protocol No.TNX-CY-F301/AFFIRM). [redacted]

[redacted]

2. OVERVIEW

This SAP describes the planned analysis and reporting for Protocol TNX-CY-F304 (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 Patients with Fibromyalgia).

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the FDA, European Medicines Agency (EMA), and International Council for Harmonisation (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 for statistical practice, as published by the American Statistical Association and the Royal Statistical Society.

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]

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 relating to collection and timing of planned clinical assessments are not repeated in this SAP unless they are relevant to the planned analysis.

3. STUDY OBJECTIVES AND ENDPOINTS

3.1 Study Objectives

3.1.1 Primary Objective

The primary objective is to evaluate the efficacy of TNX-102 SL taken at bedtime over 14 weeks of treatment using an 11-point (0-10) numerical rating scale (NRS).

3.1.2 Secondary Objectives

The secondary objective is to evaluate the safety of TNX-102 SL taken at bedtime over 14 weeks of treatment.

3.2 Study Endpoints

3.2.1 Efficacy Endpoints

The primary efficacy endpoint is:

- Change from Baseline to the Week 14 endpoint in the diary NRS weekly average of daily self-reported average pain severity scores.

Key secondary efficacy endpoints include:

- Proportion of subjects with a Patient's Global Impression of Change (PGIC) rating of "very much improved" or "much improved" at the Week 14 endpoint
- Change from baseline in the Fibromyalgia Impact Questionnaire – Revised (FIQR) symptoms domain score at the Week 14 endpoint
- Change from baseline in the FIQR function domain score at the Week 14 endpoint
- Change from baseline in the Patient Reported Outcomes Measurement Information System (PROMIS) score for sleep disturbance at the Week 14 endpoint
- Change from baseline in the PROMIS score for fatigue at the Week 14 endpoint
- Change from baseline in the weekly average of the daily diary assessment of sleep quality at the Week 14 endpoint

3.2.2 Exploratory Endpoints

Exploratory efficacy endpoints include:

- Proportion of subjects with a ≥30% improvement from baseline to Weeks 1-14 in the daily self-reported pain severity score
- Proportion of subjects with a ≥50% improvement from baseline to Weeks 1-14 in the daily self-reported average pain severity scores
- Proportion of subjects with a PGIC rating of "very much improved" or "much improved" at all post-randomization clinic visits
- Mean PGIC rating at all post-randomization clinic visits

- Change from baseline in the FIQR total score, overall impact domain score, and individual item scores at all post-randomization clinic visits
- Change from baseline in the FIQR symptoms domain score and function domain scores at all post-randomization clinic visits
- Change from baseline in the PROMIS score for sleep disturbance at all post-randomization clinic visits
- Change from baseline in the PROMIS score for fatigue at all post-randomization clinic visits
- Change from baseline in the weekly average of the daily diary assessment of sleep quality at Weeks 1-14
- Change from baseline in the weekly average of the daily self-reported average pain severity scores at Weeks 1-14
- Changes from baseline in patient-rated Changes in Sexual Functioning Questionnaire Short-Form (CSFQ-14) in females and in males, analyzed separately
- Change from screening in the in-clinic assessment of pain (24-hour & weekly recall)

3.2.3 Safety Endpoints

Safety is assessed by the monitoring and recording of Adverse Events (AEs), clinical laboratory tests, vital signs, and physical examinations including examinations of the oral cavity, as well as the monitoring of suicidality using the Columbia Suicide Severity Rating Scale (C-SSRS), and the monitoring of depression status via the Beck Depression Index (BDI-II).

4. OVERALL STUDY DESIGN AND PLAN

This is a Phase 3, randomized, parallel-group, double-blind, placebo-controlled, 14-week study designed to evaluate the efficacy and safety of TNX-102 SL 5.6 mg (2 x 2.8 mg tablets) taken daily at bedtime for the management of fibromyalgia. The study is being conducted at approximately 35-45 investigational sites in the United States (US).

The study will consist of a Screening Visit (Visit 1, Days −35 to −8), a washout period and screening period of at least 7 days (for those subjects not requiring washout) and no more than 35 days, inclusive of a 7-day baseline data collection phase immediately preceding the Baseline visit. For extenuating circumstances, the duration of the screening period may be increased up to 49 days with Medical Monitor approval. Note, this trial was ongoing at the time the SARS-CoV-2 virus began to spread in the US, leading to the illness known as COVID-19 that is part of the COVID-19 pandemic. Due to the exceptional circumstances caused by the COVID-19 pandemic, an option for a telephone visit was made available for Weeks 2, 6, 10 (Visit 3, 4, 5), and, only with Medical Monitor approval, Week 14/Visit 6 (or Early Termination) for those unable to attend an in-clinic visit due to the COVID-19 pandemic. In cases in which it is feasible for a site's own research staff, with or without a research clinician, to make a home visit for Week 14/Visit 6 (or Early Termination), a home visit should be conducted rather than a telephone visit in order to collect greater safety data than the telephone visits allow. The Screening phase will be followed by the Baseline/Randomization Visit (Visit 2, Day 1), and 4 treatment visits at Weeks 2, 6, 10, and 14 (or early termination). The total duration of the study, including Screening, will be approximately 15-21 weeks. The maximum treatment duration will be 14 weeks.

Eligible subjects who provide written informed consent will have study assessments performed at Screening and will stop all excluded medications during the washout period through the Week 14 visit (Visit 6). Subjects will record their average pain intensity for the previous 24 hours in the evening every day from Visit 1 through Visit 6 on an electronic diary system. After recording Baseline Diary scores for at least 7 days, subjects will return to the investigative site for Baseline assessments and randomization (Day 1), where they will be randomly assigned to receive TNX-102 SL or matching placebo sublingual tablets in a 1:1 ratio. Subjects will take the study drug sublingually daily at bedtime, starting on the day that they are randomized (Day 1), for 14 weeks. For the first two weeks of treatment, subjects will start on TNX-102 SL 2.8 mg (1 tablet) or placebo. At the Week 2 visit (Visit 3), all subjects will have the dose increased to TNX-102 SL 5.6 mg (2 x 2.8 mg tablets) or 2 placebo tablets. Subjects will return to the site for assessment of efficacy and safety, as well as dose tolerability, at Weeks 6, 10, and 14 (or early termination), unless a telephone visit is conducted, in which case certain assessments that cannot be done over the phone (e.g. vitals and oral cavity exam) are allowed to be omitted, which is specifically stated in the protocol. In scenarios in which 5.6 mg dose is considered intolerable and would otherwise lead to study discontinuation, the Investigator may lower the dose back to TNX-102 SL 2.8 mg or one placebo per day. Re-challenge with TNX-102 SL 5.6 mg or 2 placebo tablets may be attempted at a later date if/when it is deemed clinically warranted by the Investigator, or the subject may remain on the lower 2.8 mg dose for the remainder of the study.

Subjects who wish to withdraw from the study may do so at any time.

Subjects will be trained on use of the diary system at the Screening visit. Each evening, when the subject utilizes the diary, the system will prompt the subject regarding daily average pain

intensity and sleep quality from the previous night, and, starting Day 2 post-randomization, regarding study drug dosing the previous night.

At Screening Visit 1 and after signing the written informed consent, any required washout should be discussed with the subject and plans made for an appropriate schedule for reducing/stopping any excluded medications. This down titration and medication withdrawal must be accomplished so that the subject is medication-free for at least 21 days prior to randomization. This will provide 14 days off the excluded medication before the subject starts the 7-day run-in phase during which critical baseline daily diary efficacy data are collected. Any additional time required for down-titration would be in addition to this 21-day washout requirement. For this reason, subjects can remain in screening for up to 35 days, which provides an additional 2 weeks for down-titration, if required.

4.1 Selection of Study Population

For a complete list of inclusion and exclusion criteria please refer to the study protocol referenced on page 1.

4.2 Method of Treatment Assignment and Randomization

The following randomization criteria must be satisfied at the Baseline visit (Visit 2) in order for the subject to continue in the study and be randomized:

1. The subject continues to meet all inclusion and exclusion criteria, including urine and blood test results and is successfully and consistently utilizing the diary system.
2. The subject's FM diary pain ratings satisfy the following criteria, as assessed by diary pain scores (24-hour recall):
    a. A mean pain intensity score ≥4 and ≤9 on the 11-point (0-10) NRS scale for the 7 days immediately preceding Visit 2, and
    b. No more than 2 individual days with a score <4 on the 7 days immediately preceding Visit 2, and
    c. No score of 10 on any of the 7 days immediately preceding Visit 2, and
    d. Pain scores must be recorded on at least 5 out of the 7 days immediately preceding Visit 2.

Before the start of the study, a computer-generated randomization schedule will be prepared. Based on the randomization schedule, eligible subjects will be randomly assigned in a 1:1 ratio to receive TNX-102 SL tablets or placebo tablets for 14 weeks.

Treatment A:	For Days 1-14, 1 tablet of TNX-102 SL 2.8 mg taken sublingually (under the tongue) each day at bedtime. For Days 15-98, 2 tablets of TNX-102 SL 2.8 mg (5.6 mg) taken simultaneously and sublingually each day at bedtime.
Treatment B:	For Days 1-14, 1 tablet of placebo taken sublingually (under the tongue) each day at bedtime. For Days 15-98, 2 tablets of placebo taken simultaneously and sublingually each day at bedtime.

4.3 Treatment Blinding

This is a double-blind study. Unless otherwise specified, all study personnel are to remain

blinded to study drug. With the exception of unblinded CRO staff performing the interim analysis and IDMC members, treatment assignments will not be revealed until all subjects have completed the study and the database has been finalized and closed.

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

4.4 Minimization of Missing Data

The Sponsor believes it has incorporated strong steps into the study design to ensure minimization of missing data during the treatment period. Multiple analysis techniques to examine the impact of missing data on the robustness of results will be carried out. Additionally, investigators have the option after the Week 2 visit when dose escalation has occurred, to reduce the treatment dose back to one tablet of TNX-102 SL 2.8 mg or one tablet of placebo if the higher dose is resulting in tolerability issues that would otherwise lead to discontinuation in the study. The importance of minimizing the amount of missing data will be discussed with all study investigators, and their awareness of the importance of subject compliance and minimal dropout rates is factored into their recruiting plans and daily subject management. In addition, investigators are advised to contact the medical monitor for guidance on available management options when needed to avoid subjects withdrawing from the study.

4.4.1 Opioid Usage

In order to minimize confounding issues related to concomitant usage of opioids, investigators have been asked to identify candidates for this clinical trial that are not currently using chronic opioids. However, it is understood by the Sponsor that opioid usage is sometimes unavoidable for acute conditions. In the event that an opioid is required for the management of an acute pain condition, the subject will be instructed to contact the site immediately so that appropriate management decisions can be implemented and accurate medication records obtained. In addition, when feasible, study visits may be delayed to avoid the contamination of data by recent opioid usage. At a minimum, no opioid/narcotic should be utilized within 2 days of a study visit, and ideally there will be no usage during the 7 days prior to any visit.

A listing of concomitant medication CRF data identifying the use of opioids will be reviewed by the project team and approved by the Sponsor prior to database lock and unblinding, and will be used to flag records in the analysis database and in the by-subject concomitant medication listings. These flagged records will also be utilized to identify records to be censored in the sensitivity analysis described in Section 10.1.3.

4.4.2 Intermittent Missing Data

Intermittent missing daily data will not be imputed and weekly averages will be calculated using available values, even if only a single value is available for a week. Intermittent missing data (weekly averages and/or in-clinic assessments) occurring prior to discontinuing use of study drug is assumed to be missing at random, and analyses have been chosen to mitigate the impact of these missing data (Section 10.1).

5. ANALYSIS AND REPORTING

5.1 Interim Analysis

An optional interim analysis may 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 one of the following:

- Stop the study early for efficacy [redacted]
- Increase the sample size by a fixed amount
- Keep the current sample size and continue as planned
- Stop the study early for futility [redacted]

[redacted]

The interim analysis will be conducted approximately 18 weeks after randomization of approximately 236 participants, i.e. when approximately 50% of the initially planned participant enrollment (470) is evaluable for efficacy assessments (approximately 14 weeks); and related data cleaning, database freeze, and administrative tasks have been completed on this cohort (approximately 4 weeks).

In the case of an early stop for futility or efficacy, 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.

[redacted]

5.2 Final Analysis

All final, planned analyses will be performed after the last subject has completed the last study visit and end-of-study assessments and all relevant study data have been processed and integrated into the analysis data base. Any post-hoc, exploratory analyses completed to support planned study analyses, but 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 as such in the text of the CSR.

6. SAMPLE SIZE DETERMINATION

The study is planned to enroll approximately 470 subjects total in a 1:1 randomization, that is, 235 subjects in each of the TNX-102 SL and placebo arms. [redacted]

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

7. ANALYSIS POPULATIONS

The following analysis populations are planned for this study:

- **Safety Population (SAFETY):** all subjects who took investigational product. Subjects that were issued study drug, but do not return all of it will be included in the safety population. All safety analyses and demographic/baseline characterization will be performed using this population, analyzed as treated.
- **Intention-To-Treat Population (ITT)**: all subjects who were randomized. This is the primary population for efficacy analyses, and subjects will be analyzed based on their randomized treatment.
- **Interim Population:** approximately the first 235 subjects randomized; this will be utilized for the interim analysis, if performed. In the case of an early stop, this will serve as the population for efficacy analyses. The cutoff of the population will be the date that the 235th subject is randomized; thus it is possible to have a marginal overrun depending on the number of subjects randomized that day.
- **Early Stop Safety Population:** all subjects that are in the safety population and the interim population. This will serve as a population parallel to the Interim Population for the purpose of evaluating the risk/benefit on the same sample.

8. GENERAL ISSUES FOR STATISTICAL ANALYSIS

8.1 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. All subjects for a population in each treatment group will be grouped together; subjects that reduce dose from two tablets per day to one tablet per day will be not be reported separately for the analyses described in this SAP, unless otherwise noted. If there are more than 10 subjects in the active arm with a dose reduction at any point in the study, summary statistics on key outcomes may be reported for subjects with dose reductions.

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

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

Categorical, qualitative variable summaries will include the frequency and percentage of subjects who are in the particular category. In general, the denominator for the percentage calculation will be based upon the total number of subjects 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. For scores obtained via diary, baseline will be defined as the average of the scores from the 7 days prior to the baseline visit. Change from baseline will be defined as the post-baseline visit value minus the baseline value.

Study day is defined as assessment date – first dose date +1.

All analyses will be performed using SAS® Software version 9.3 or later.

8.1.1 Adjustments for Multiplicity and Other Alpha Control

8.1.1.1 Alpha Spend, Early Stops and Sample Size Increase

[redacted]

Under both cases (early stop for efficacy, study continues to full enrollment), the alpha adjustment being applied to the primary will also be utilized as the criteria for the key secondary outcomes.

Likewise, 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 with the adjusted alpha to account for the first stage alpha spend.

8.1.1.2 Multiplicity

To adjust for multiplicity and to control for overall type I error, a sequential test procedure will be applied to the primary and key secondary efficacy endpoints. If the primary analysis produces a result that is statistically significant at the two-sided [redacted] level, a significance level of [redacted] will be used for comparing the secondary endpoints in an ordered fashion. If the analysis for a secondary endpoint does not produce a statistically significant result ([redacted]), then the remaining secondary endpoint analyses will automatically be considered non-significant regardless of the p-value produced. [redacted]

The order of key secondary endpoints in which they are to be tested is as follows:

- Proportion of subjects with a PGIC rating of "very much improved" or "much improved" at Week 14
- Change from baseline in the FIQR symptoms domain score at Week 14
- Change from baseline in the FIQR function domain score at Week 14
- Change from baseline in the PROMIS score for sleep disturbance at Week 14
- Change from baseline in the PROMIS score for fatigue at Week 14
- Change from baseline to Week 14 in the weekly average of the daily diary assessment of sleep quality.

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 10.

8.1.2 Data Handling for Subjects Who Withdraw/Drop Out from the Study

Subjects who withdraw/drop out from the study will have the early termination (ET) data collected at their ET visit included in the analysis at the closest visit (Week 2, 6, 10, 14), using midpoints between visits to window the early termination. If this results in two records for a given visit, then the one closest to the targeted date will be used.

For example, a subject with valid Week 10 clinic assessment collected on Day 72 that has early termination data collected on Day 79, would have the early termination data mapped to Week 10; however, since Day 72 data is present and closer to the target date, it would be analyzed and the Day 79 data would be excluded. A subject without valid Week 10 data who early terminates on Day 79 would have the Day 79 data analyzed with the Week 10 data.

Subjects who provide 14 weeks of diary pain data will be analyzed as completers for purposes of the primary analysis, even if for some reason they are unable to attend a Week 14 final study visit. As a specific example, a subject who successfully provides 14 weeks of diary data as described in Section 8.2.1, but who is unable to attend a final study visit during the required time window because of travel or other circumstances, would be analyzed as a completer for purposes of the primary endpoint, but would be missing non-diary based secondary outcome measures at

the Week 14 endpoint. For safety data, the last observation available will be summarized in addition to the presentation above (grouping ET visits with the closest planned visit); this will combine the ET visit data with the completers' Week 14 data, regardless of when the ET occurred.

8.1.3 Imputation of Missing Data

For individual daily pain scores, since a mean is used in the calculation of the primary endpoint, it is not necessary to replace values missing on random intermittent days; weeks will have average values as long as a single value for that week is present.

Missing weekly pain scores for participants in the ITT population will be imputed via multiple imputation (MI). When the study medication was discontinued due to lack of efficacy (LOE) or due to an AE, missing values will be imputed drawing from the baseline values, conditioned on the non-missing weekly average post-baseline values, of all participants in the ITT population under the assumption that they are missing not at random (MNAR), with a covariate for site. If the study medication was discontinued for any other reason (including impacts from the COVID-19 pandemic), values will be imputed within treatment group using MI under the assumption that they are missing at random (MAR) with covariates for site and the average pain severity score recorded at each time point. The MAR approach will also be applied for sporadic missing values (prior to discontinuation). The MI process will use twenty repeats in the analysis datasets generated; seeds are given in Appendix D. For the purposes of the MI, all sites with less than 10 subjects will be pooled into a single large site; this will be done for both the interim and post interim samples.

The steps will be as follows:

1. Intermittent missing weekly average scores will first be imputed using the Markov Chain Monte Carlo (MCMC) method implemented with the SAS MI procedure, which is appropriate for non-monotonic missing data.
2. For subjects that do not discontinue for LOE or AEs, a MAR imputation will be applied: the monotone missing values will then be multiply imputed with the SAS MI procedure using the monotone regression method. Baseline average and the weekly scores will be included as covariates as well as site; imputation will be done within treatment group.
3. For subjects that discontinue for LOE or AEs, an iterative process will be applied moving through each week. The complete MAR dataset from step 2 above will be used and augmented with subjects missing for the week being imputed at that stage. The non-missing values for the week in question will temporarily be set to the subjects corresponding baseline value, so the data from LOE and AE subjects are imputed for that week drawing from the baseline distribution. This will proceed from Week 1 to Week 14 with the imputed complete dataset for a week serving as the basis for the subsequent week. Covariates at each stage will include the values for the prior weeks (observed, MAR imputed in steps 1& 2, or imputed previously in this step) and site.

The above process will be done independently for the subjects included in the interim analysis and those not in the interim.

Daily sleep diary data will be handled in an identical manner to the pain scores.

Continuous outcomes included in the list of key secondary outcomes that are collected in-clinic will be handled as above, minus the step of creating weekly averages of daily values and with steps being done by visit instead of week.

Subjects with missing PGIC at a given visit will be considered non-responders at that visit for the purposes of treatment comparisons and efficacy claims; additional tabulations will summarize the number and percentage in each category out of non-missing values as well as a count of the missing values.

No missing safety data will be imputed.

8.2 Efficacy Assessments

8.2.1 Average Daily Pain Score

The average daily pain score will be obtained using the daily, 24-hour recall, NRS pain data collected with the diary system. Since a mean is used in the calculation of the primary endpoint, it is not necessary to replace values missing on random intermittent days.

The baseline pain score will be defined as the average of the pain scores from the 7 days prior to the baseline visit. Pain scores must be recorded on at least 5 out of the 7 days immediately preceding the baseline visit in order for the subject to qualify for randomization.

The mean of the daily, subject self-reported, 24-hour recall, pain severity scores for the 7 days of each week will serve as the pain score for that study week. All available values will be used; if a subject has at least one value for the week, the week is non-missing. Because there is a -4 day and +7 day window around the scheduling of Visit 6 (Day 99), the study weeks will be anchored by the timing and completeness of their final study visit. Ideally, subjects appearing in-clinic on Day 99 (defining randomization as Day 1), would have last filled out their diary on Day 98; therefore, "Week 14" would include Days 92-98; the protocol-defined window allows visit 6 as early as Day 95, resulting in a nominal "Week 14" including Days 88-94. Likewise, subjects may appear in-clinic per protocol as late as Day 106, resulting in a "Week 14" interval running from Day 99-105.

Due to impacts from the COVID-19 pandemic, the criteria for calculating the Week 14 average will be marginally relaxed beyond the protocol-specified visit windows to allow for additional flexibility by permitting data collected during Week 13 to be used to calculate the Week 14 average: working backwards from Day 98, the first non-missing day (as early as Day 91) will serve as the LAST day of Week 14 for the purpose of anchoring and any non-missing values in the 6 days prior to that day will contribute to the average. If a subject does not have a non-missing value on/after Day 91, any available data from Day 85-91 will be averaged as Week 14 and Day 91 will serve as the anchor. Subjects without data on/after Day 85 will be missing for Week 14 and handled as described below. Specific examples of calculating the Week 14 pain score are described in Appendix C.

Each week prior to Week 14 will be based on 7-day intervals; depending on the interval selected for "Week 14", this will result in Week 1 having less than 7 days or some "extra" days prior to Week 1. Extra days will be dropped and not included in the analysis; in the case of less than 7 days, the available days will be averaged, as long as there are non-missing values. The earliest

day included in the Week 1 average will be the data entered the day after randomization (covering roughly the 24 hours following their first dose of study drug).

For subjects who withdraw from the study early, that do not have a final study visit or complete the study but do not have data in the 7 days preceding their last visit, the last day of diary data (at or prior to Day 106) will serve as the anchor point for dividing the available data into weeks. First, the nearest nominal end day for a weekly period is identified and this week will be the assigned the data from the 7 days immediately preceding the anchor date. For example, if a subject early terminates on study Day 24, the Week 3 (nominally ending on day 21) average will be based on the 7 days prior to Day 24 (Days 17-23). Lost to follow-up subjects that continue to fill out the diary into Week 14 and beyond will have non-missing values for the Week 14 timepoint for diary outcomes. Day 106 will be the last allowable day to be included in the average.

Week	Nominal Study Day Intervals
Baseline	Day -7 to -1
Day of Randomization	Day 1
Week 1	Day 2 to 7
Week 2	Day 8 to 14
Week 3	Day 15 to 21
Week 4	Day 22 to 28
Week 5	Day 29 to 35
Week 6	Day 36 to 42
Week 7	Day 43 to 49
Week 8	Day 50 to 56
Week 9	Day 57 to 63
Week 10	Day 64 to 70
Week 11	Day 71 to 77
Week 12	Day 78 to 84
Week 13	Day 85 to 91
Week 14	Day 92 to 98

Change from baseline will be defined as the pain score at each week minus the baseline pain score (with baseline and weekly pain scores derived as described above). Thus, negative changes will denote lesser pain and larger negative values will denote greater improvement.

8.2.2 Average Daily Sleep Quality Score

The average daily sleep quality score will be obtained in the same manner as the average daily pain score, via electronic Interactive Response Technology (IRT) system. Derivation of the weekly average sleep quality scores and censoring of data will be treated the same as the weekly average pain scores described in Section 8.2.1, using the same windows based on the date of collection. It should be noted that the sleep prompt asks about the prior night's sleep; thus, subjects report their sleep starting the first night that they take randomized drug through the night

between day 97-98, if their final visit is day 99. As with pain, negative changes will denote improvement in sleep quality and larger negative values will denote greater improvement.

8.2.3 Patient Global Impression of Change (PGIC)

PGIC is a question completed by the subject at Weeks 2, 6, 10, and 14. The subject will rate the change in their overall fibromyalgia symptoms on a 1-7 Likert scale, where 1 is "Very Much Improved" and 2 is "Much Improved". Scores of 1 and 2 will be considered PGIC responders, and all other scores will be considered non-responders for that visit. The proportion of responders in each treatment arm will be analyzed. Any missing PGIC score will be considered a non-responder for that visit.

8.2.4 Revised Fibromyalgia Impact Questionnaire (FIQR)

The FIQR is made up of 3 domains: (1) functional (9 questions); (2) overall impact (2 questions) and (3) symptoms (10 questions). If 2 or more items are missing from the functional domain, the domain will be considered invalid for that visit. Likewise, if *any* item is missing from the overall impact domain or the symptom domain, those domains will be considered invalid.

To account for missing items in the functional domain, the score will be recalculated as:

$$\text{New score} = (\text{Raw score} / \#\text{ of items answered}) \times \#\text{ of items in domain}$$

The FIQR total score can be determined after all domain scores have been calculated using the following steps. First, divide the function domain score by 3, divide the overall impact domain score by 1, and divide the symptom domain score by 2. Next, add the three resulting domain scores to obtain the total score of FIQR. If any domain score is missing, then the total is missing.

8.2.5 Patient Reported Outcomes Measurement System (PROMIS) Instruments

8.2.5.1 PROMIS Short-Form (SF) Fatigue and Sleep Disturbance Instruments

The PROMIS Short-Form Fatigue and Sleep Disturbance Instruments each consist of 8 items in which responses are scored 1 to 5 for each item. Scores for all items are totaled to create a raw score. Either instrument will be considered invalid if 50% of the items are missing. If more than 50% of the items are answered, the raw total score will use the following formula to determine a new calculated score to account for missing items.

$$\text{New raw score} = (\text{Raw score} / \#\text{ of items answered}) \times \#\text{ of total items}$$

New raw scores that are fractions are rounded up to the nearest whole number. Once a new raw score has been calculated, a T-score will be determined using Appendix A for Fatigue Inventory and Appendix B for Sleep Disturbance. The T-scores will be analyzed.

Note that items 1, 2, and 8 of the PROMIS Sleep Disturbance Instrument will need to have their directionality reversed for calculating the totals.

8.2.6 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 patient to self-evaluate his or her sexual behaviors or problems in a number of areas. The CSFQ-14 will be administered at Baseline (Visit 2) and Week 14/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; 5=never). Items 10 and 14 are included in the total score but not in any subscale scores.

8.3 Safety Endpoints

To handle missing or partial AE and 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 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?”

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

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

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 treatment.
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 treatment.
4. If the start date is not missing but the stop date is missing, then the most conservative approach is taken and the medication is considered to be concomitant.

9. STUDY SUBJECTS AND DEMOGRAPHICS

9.1 Disposition of Subjects and Withdrawals

The numbers and percentage of subjects screened, 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 subjects in each analysis population will be reported. Additionally, the number and percentage of subjects that have an investigator-directed dose reduction from two tablets per day to one tablet per day will be summarized. The disposition and withdrawal summaries will be based on all subjects who have data entered into the database.

9.2 Protocol Violations and Deviations

Protocol deviations will be checked on complete data for all subjects. Protocol deviations will be summarized by type, status as major vs minor, and by treatment group for the Safety population.

Individual subjects with protocol deviations or violations will be listed.

9.3 Demographics and Other Baseline Characteristics

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

Descriptive summaries of demographic and other baseline conditions will include:

- Demographics (age, gender, race/ethnicity, height, weight, BMI, family status, education and employment status)
- Tobacco/nicotine, alcohol, and THC/cannabis use history
- ACR Criteria

Medical History will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.1 and summarized by SOC and Preferred Term using frequency counts by treatment group. Physical examination data, as well as findings from dedicated oral cavity examinations, will be presented in listings.

10. EFFICACY ANALYSES

10.1 Primary Efficacy Analysis

10.1.1 Estimand

The primary ITT analysis will provide an estimate of the following causal estimand: the difference in the weekly mean change from Baseline of the daily subject self-reported 24-hour recall average pain severity rating using an 11-point (0-10) NRS evaluated at the Week 14 endpoint in all randomized subjects attributable to the initially randomized treatment assignment, under the assumption that those withdrawing due to AEs and lack of efficacy revert to the baseline pain levels (on average).

10.1.2 Primary 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 the first half of the subjects randomized, planned as 236. The second stage population is defined as all subjects randomized after the interim cut-off.

The interim analysis will be executed once all subjects in the Interim Population either drop out or reach Week 14.

If the study is stopped early for efficacy, subjects who were randomized after the interim cut off will not be included in the efficacy analyses. For early stop, safety analyses will be reported both on all subjects exposed (the Safety Population) and all subjects exposed in the Interim Population (Early Stop Safety Population).

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 weekly average of daily pain between TNX-102 SL and placebo treatment groups at Week 14: (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 470 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).

The mean change from baseline in the weekly average of daily pain scores from baseline to each week in the TNX-102 SL and placebo arms will be analyzed using a restricted maximum likelihood (REML) based repeated measures approach with data imputed using multiple imputation (see Section 8.1.3). Week 14 will serve as the primary time point of interest. Models will be run on each of the twenty imputation sets individually and will include the fixed, categorical effects of treatment, site, study week, and treatment by study week interaction, as well as the fixed covariates of baseline value and baseline value score-by-study week interaction. An unstructured covariance structure will be used to model the within-subject errors. If this analysis fails to converge, the following structures will be tested in this order: first order ante dependence, heterogeneous first order autoregressive, heterogeneous compound symmetry and compound symmetry with the first to give convergence used. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. LS means and differences across the twenty MI reps will be combined using SAS procedure MIANALYZE (Rubin, 1976). Significance testing will be based on least-squares means and two-sided 95% confidence intervals will be presented. [redacted]

10.1.3 Sensitivity Analyses

Four sensitivity analyses will be performed:

For the first, pain scores will be censored on the day(s) on which an opioid is utilized. If the subject utilizes an opioid for any indication, then the pain score on each impacted day will be replaced by the score obtained on the day immediately prior to the first day of opioid use (and until the opioid is no longer used). This will be done prior to the weekly averaging, and all other calculations will be performed as described in the primary. This will investigate whether use of concomitant opioids influences the interpretation of study results.

For the second, subjects that discontinue due to “withdrawal of consent” or “investigator decision” will be grouped with the LOE and AE dropouts when performing the multiple imputation procedure. This analysis will investigate whether the categorization of these reasons for dropout influences the interpretation of study results.

For the third, ALL subjects that discontinue will be grouped with the LOE and AE dropouts when performing the multiple imputation procedure. This analysis will sample from the baseline distribution of all dropouts and will investigate whether treating dropouts as treatment failures (on average) influences the interpretation of study results.

Finally, a tipping point analysis will be performed on the values for the primary analysis. A shift parameter will be applied to the active treatment arm starting at .25 and increasing in increments of .25 through either 5 or when the outcome is no longer significant. This will only be performed if the primary analysis is significant. This will test the overall robustness of the

primary analysis.

10.2 Secondary Efficacy Analyses

Secondary efficacy analyses will be based on the ITT population only.

Continuous Endpoints

For the purposes of possible label claims and reported p-values entered into the multiplicity algorithm, an approach identical to the primary analysis will be used for all continuous outcomes. See Section 8.1.3 for the imputation algorithm and Section 10.1.2 for the analysis approach.

Outcomes using these analyses will include:

- Change from baseline in the FIQR symptoms domain score
- Change from baseline in the FIQR function domain score
- Change from baseline in the PROMIS score for sleep disturbance
- Change from baseline in the PROMIS score for fatigue
- Change from baseline in the weekly average of the daily diary assessment of sleep quality.

For each, all time points will be summarized, but the primary time point of interest is Week 14. Time points other than Week 14 are considered exploratory. The estimand for each is identical to the primary, substituting for the endpoint in question.

For the sleep quality assessment collected via a daily subject diary, the baseline score and the weekly scores for each subject will be calculated as the weekly average score based on the mean of the scores recorded for that study week (Section 8.2.2).

This will be repeated on the outcomes above, but using observed data only without imputation.

Categorical Endpoints

A categorical analysis of PGIC will be performed using a logistic regression model for each week with effects for treatment and investigative site. If the model does not converge, site will be removed from the model. Subjects with results of "very much improved" or "much improved" (defined as responders) will be compared to all other categories (defined as non-responders). Subjects with missing data will be considered non-responders. In addition, odds ratios and corresponding 95% CI intervals for being "very much improved" or "much improved" versus all other categories will be presented. A summary of frequency counts for PGIC responses for each time point will be presented; Week 14 is of primary interest and all other time points are considered exploratory.

10.3 Exploratory Analyses

All efficacy analyses at time points other than Week 14 are considered exploratory as described above.

Additional exploratory efficacy endpoints include:

- FIQR total score, overall impact domain score, and individual item scores at all post-randomization clinic visits
- Proportion of subjects with a ≥30% improvement from baseline to Weeks 1-14 in the daily self-reported pain severity score
- Proportion of subjects with a ≥50% improvement from baseline to Weeks 1-14 in the daily self-reported average pain severity scores
- PGIC as a continuous variable with 1-7 scoring
- CSFQ-14 at Week 14
- In-clinic assessment of pain (24-hour & weekly recall)
- Permutation test of change from baseline in the weekly average of the daily pain scores at week 14.

The proportion of improvement endpoints will be analyzed in using a logistic regression model for each week with effects for treatment and investigative site. If the model does not converge, site will be removed from the model. Subjects with missing data will be considered non-responders. In addition, odds ratios and corresponding 95% CI intervals for being a pain responder will be presented.

In addition to the responder analysis in the secondary outcomes, PGIC will be analyzed as a continuous variable scoring the responses 1-7. Results will be reported by treatment with summary statistics and an MMRM Model will be used to compare treatments. The model will include the fixed, categorical effects of treatment, site, study week, and treatment by study week interaction. An unstructured covariance structure will be used to model the within-subject errors. If this analysis fails to converge, the following structures will be tested in this order: first order ante dependence, heterogeneous first order autoregressive, heterogeneous compound symmetry and compound symmetry with the first to give convergence used. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Significance testing will be based on least-squares means using a two-sided alpha=0.05 test and two-sided 95% confidence intervals will be presented.

The CSFQ-14 is completed at baseline and Week 14/ET. It has a male and a female version, which will be analyzed separately. The CSFQ-14 will be analyzed for its total score, three subscales on phases of sexual response, and five subscales for dimensions of sexual functioning by using an Analysis of Covariance (ANCOVA) model. The model will include the fixed, categorical effects of treatment and site, plus a baseline covariate corresponding to the subscale (or total) being analyzed. Significance testing will be based on least-squares means using a two-sided alpha=0.05 test and two-sided 95% confidence intervals will be presented. This outcome is considered to be supportive of the product's safety; other exploratory analyses are for further support of product efficacy.

The FIQR exploratory endpoints and the in-clinic pain assessments will be analyzed using an MMRM including the fixed, categorical effects of treatment, site, study week, and treatment by study week interaction, as well as the fixed covariates of baseline value and baseline value score-

by-study week interaction. An unstructured covariance structure will be used to model the within-subject errors. If this analysis fails to converge, the following structures will be tested in this order: first order ante dependence, heterogeneous first order autoregressive, heterogeneous compound symmetry and compound symmetry with the first to give convergence used. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Note that the screening values will be used as baseline for the in-clinic pain outcomes.

10.3.1 Exploratory Analyses of Daily Pain Scores

Randomization Honoring Non-Parametric (RHNP) hypothesis test.

In this exploratory analysis, the primary endpoint of Week 14 change from baseline in daily pain scores will be analyzed by a RHNP method for computing p-values for the hypothesis that treatment has no effect.

Subjects will be randomly reassigned to treatment 10,000 times using the original randomization algorithm, in the order subjects actually enrolled, using the original block sizes and randomization strata. The apparent effect of the treatment under each of these hypothetical assignments will be compared to the apparent effect for the actual assignment, approximating the distribution of the apparent treatment effect under the null hypothesis that treatment makes no difference whatsoever.

The apparent effect size for each random allocation will be quantified by the median difference between the responses of subjects assigned to treatment by that allocation and those assigned to placebo by that allocation.

The test statistic will be calculated for 10,000 random allocations and for the original observed data with actual treatment codes. The percentile rank of the statistic for the true treatment assignment among the distribution of the statistics for the permutations is one-sided p-value for an exact randomized test of the null hypothesis that treatment makes no difference whatsoever. If treatment assignment had no impact, the test statistic for the original observed data would be unlikely to be in the tail of the distribution of values of the test statistic for the random assignments. The percentile is the p-value of the null hypothesis.

The primary test-statistic of interest for the first stage will be the difference in the medians between the treatment and control groups at Week 14. This statistic makes sense for ordinal variables such as the Likert scales that comprise each response. Other test statistics may also be investigated including:

- The difference in treatment means at Week 14
- The LS means from an ANCOVA at Week 14 with covariates for baseline value, site and sex
- The LS means at Week 14 from an MMRM identical to the primary efficacy analysis

In addition to performing the above analyses on the observed data, the randomization test will be repeated using multiple imputation implemented in a manner identical to that described for the primary analysis.

In a third analysis, the RHNP method will handle the missing data by assigning each subject with missing Week 14 values a +10 value for their change; since medians will be utilized, the particular value has no impact as long as it is larger than any possible change and less than 50% of the data are missing. This amounts to assuming that missing responses are "bad" and should be scored as more extreme; the more data are missing, the more it will pull the median in the direction of a finding that treatment is ineffective.

The performance characteristics of the RHNP method will be compared against the parametric approaches described for the primary analyses of the change from baseline to the Week 14 endpoint in the diary NRS weekly average of daily self-reported average pain severity scores.

10.4 Analyses by Dose

Given that TNX-102 SL is generally well tolerated, it is unlikely that a large number of subjects will have a dose reduction to one tablet and hence it is nearly certain that an insufficient number will have such a reduction so as to allow the multiple imputation algorithm return stable results taking dose into account. If more than 10 subjects in the active arm have a dose reduction at any point in the study, summary statistics on observed values for the primary and key secondary outcomes will be reported for placebo and each dose, with active treatment subjects grouped by whether they had a PI-directed dose reduction at any time point, regardless of the dose they were on at a particular time point. Placebo subjects that had a "dose reduction" will not be grouped separately.

11. SAFETY AND TOLERABILITY ANALYSES

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

- Adverse Events
- Clinical Laboratory Investigations
- C-SSRS
- BDI-II
- Vital Signs
- Physical Examinations and examinations of the oral cavity

11.1 Adverse Events

All AEs, TEAEs, and SAEs will be coded using the MedDRA Dictionary Version 22.1.

The collection of adverse events will begin at the time the subject is consented and screened for the study. Treatment-emergent adverse events (TEAEs) are defined as either new onset AEs with an onset at the time of or following the start of treatment, as indicated by a "no" answer to "Did AE start prior to the first dose?", 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:

- All TEAEs
- TEAEs by severity
- TEAEs leading to study drug discontinuation
- TEAEs by relationship
- SAEs
- Oral cavity TEAEs
- Oral cavity TEAEs by severity

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, oral cavity TEAEs, TEAEs by maximum severity, TEAEs by strongest relationship to study drug and post-treatment AEs.

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

In the AE data listings, all AEs will be displayed. AEs that occur prior to randomization will be considered pre-treatment AEs, and will be determined by a "yes" response to the CRF question, "Did the AE start prior to the first dose?" AEs that start on the date of randomization but have a "yes" response to this question will be categorized as pre-treatment AEs. TEAEs will be defined

from the date of randomization and a "no" response to the CRF question, "Did the AE start prior to the first dose?"

11.1.1 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.

11.1.2 Serious Adverse Events

A summary of incidence rates (frequencies and percentages) of serious adverse events (SAE) 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.

11.1.3 Oral TEAEs

In light of the study drug's sublingual route of delivery, subjects will undergo a detailed examination of the oral cavity at screening and brief examinations at all other visits.

AEs involving the oral cavity may be spontaneously reported by the subject, observed during an oral cavity examination, or both. Oral cavity AEs will be identified by a "yes" response to the CRF question, "Is the AE in the oral cavity?" Additional information will be collected on oral cavity AEs (based on a "yes" response to the above question), including whether the AE occurs immediately or very soon after dosing, an approximation of the duration of the AE (less than or greater than 60 minutes), and whether the AE is still present the next morning upon awakening.

A separate by-subject listing of oral cavity AEs (including pre-treatment oral cavity AEs) will be provided.

11.1.4 Deaths

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

11.2 Clinical Laboratory Evaluations

Laboratory data (analytes 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 last available assessment for each clinical laboratory analyte and each treatment group. 95% confidence intervals will be presented for change from baseline. 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 with their corresponding normal ranges, and those values that are outside the normal range will be flagged. For each laboratory analyte, shifts in assessments of abnormality from baseline to each scheduled time point will be presented in shift tables.

A by-subject listing of all clinical laboratory (Chemistry, Hematology and Urinalysis) data will

also be provided.

11.3 Columbia Suicide Severity Rating Scale (C-SSRS)

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

The overall number of subjects with lifetime and/or current suicidal ideation (by item and category), suicidal behavior (by item and category), or self-injurious non-suicidal behavior at the screening and baseline visit will be summarized by visit and treatment group.

Additionally, the number of subjects with increase over baseline in suicidal ideation at any time point will be reported. The maximum ideation across all visits for a subject will also be summarized with counts and percentages both for all subjects and among subjects with an increase in ideation. Likewise, the count and percentage of subjects with any suicidal behavior will be reported along with a summary of the most extreme behavior each subject reported.

Category	Items
A) Suicidal Ideation	(1) Wish to be dead
	(2) Non-specific active suicidal thoughts
	(3) Active suicidal ideation with any methods (not plan) without intent to act
	(4) Active suicidal ideation with any some intent to act, without specific plan
	(5) Active suicidal ideation with specific plan and intent
B) Suicidal Behavior	(6) Preparatory acts or behavior
	(7) Aborted attempt
	(8) Interrupted attempt
	(9) Actual attempt
	10) Completed suicide
	Suicidal Behavior present (composite of items 6-10)
	Non-Suicidal Self-Injurious Behavior

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

11.4 Vital Signs

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

11.5 Physical Examination and Oral Cavity Exam

A standard physical examination will be performed at Screening, Baseline, and Week 14. In addition, a separate examination of the oral cavity will be performed at each visit (Screening through Week 14). Oral cavity examination findings will be documented separately from other physical examination findings.

A data listing of the results at each scheduled visit will be presented for both the standard physical examination and the examination of the oral cavity.

12. MEDICATIONS

12.1 Concomitant Medication

All medications will be coded using the World Health Organization Drug Dictionary (WHO-DD Version 2019-Sep). Prior and concomitant medications will be summarized by treatment group and by the number and percentage of subjects taking each medication, classified by using WHO-DD Anatomical Therapeutic Chemical Levels 1 and 3 and preferred term.

Prior medications are defined as medications or therapies initiated prior to the start of the study drug and terminated 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 subject receives concurrently with the study drug. These medications will have end dates on or after the first dose date of the study drug.

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

Please refer to Section 8.3 to impute the partial start date and end date of concomitant medication.

12.2 Opioid Medications

A list of subjects' concomitant medication records from the CRF data identifying opioid usage will be reviewed by the project team and approved by the sponsor prior to database lock. This listing will be used to flag opioid records in the analysis database and in the by- subject concomitant medication listings.

12.3 Exposure and Compliance

The treatment duration will be calculated and summarized based primarily on CRF data for first and last dose dates (number of days=last dose date - first dose date+1). If these values are missing or subjects' diary data indicate dates that exceed this period, the diary dates will be utilized.

Days of exposure will be based on subjects' responses to the daily questions regarding medication usage. The number of days on study drug is the total number of days a subject respond that study drug was taken. If the CRF data for the last dose date is a date not included in the diary, it will be added to the count (this should be true for most completing subjects since they do not fill out a diary the day of their last visit). The number of subjects with total exposure by visit weeks (≤2 weeks, 2 to ≤4, 4 to ≤8, 8 to ≤12, 12 to ≤14 and >14 weeks) will be presented. Missing days where the subject did not complete the diary will be treated as though study drug was not taken. Days of exposure will also be calculated for the days on each dose of study drug.

Additionally, the number and percentage of subjects that drop back to one tablet per day will be summarized and the week in which they reduced their dose will be tabulated (among those that had a dose reduction).

Compliance will be similarly summarized across all study visits for each treatment arm. Study drug compliance as a percentage will be defined as the exposure days defined above divided by the total number of expected days on treatment multiplied by 100. The expected number of days will be the date of last diary recorded dose-randomization date +1. This compliance reporting

will take into account whether or not they took any medication, but will not account for the number of tablets they took on a given day.

Compliance will be summarized with descriptive statistics by treatment arm. The number and percentages of subjects within certain categories of compliance (e.g. < 60%, 60% to < 80%, 80% to <= 100%, greater than 100%) will also be presented.

A listing of drug accountability data based on CRF data will be provided.

13. CHANGES FROM PLANNED ANALYSIS

The protocol inadvertently omitted the in-clinic pain assessments from the list of exploratory endpoints; it was added to this SAP.

14. REFERENCES (AVAILABLE UPON REQUEST)

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.

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

Columbia-Suicide Severity Rating Scale (C-SSRS) Scoring Guide. Available at http://cssrs.columbia.edu/scoring_cssrs.html

Keller A, McGarvey EL, Clayton AH. Reliability and construct validity of the Changes in Sexual Functioning Questionnaire short-form (CSFQ-14). J Sex Marital Ther. 2006 ;32(1):43-52.

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

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

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.

PROMIS Fatigue. (2013, May 14). *A brief guide to the PROMIS fatigue instruments*. Retrieved from http://www.assessmentcenter.net/documents/PROMIS Fatigue Scoring Manual.pdf

PROMIS Global Short Form. (2010, October 10). *Scoring PROMIS Global short form*. Retrieved from https://assessmentcenter.net/documents/Scoring%20PROMIS%20Global%20short%20form.pdf

PROMIS Sleep Disturbance. (2013, May 14). *A brief guide to the PROMIS sleep disturbance instruments*. Retrieved from https://www.assessmentcenter.net/documents/PROMIS%20Sleep%20Disturbance%20Scoring%20Manual.pdf

Rubin, D. B. (1976), “Inference and Missing Data,”Biometrika, 63, 581–592.

TNX-CY-F304 IDMC Charter Final 13 Aug 2020

TNX-CY-F304 ISAP Final 13 Aug 2020

15. APPENDICES

15.1 Appendix A

Fatigue 8a *Short Form Conversion Table*		
Raw Score	T-score	SE*
8	33.1	4.8
9	38.5	2.7
10	41.0	2.2
11	42.8	2.0
12	44.3	1.9
13	45.6	1.8
14	46.9	1.8
15	48.1	1.8
16	49.2	1.8
17	50.4	1.8
18	51.5	1.7
19	52.5	1.7
20	53.6	1.7
21	54.6	1.7
22	55.6	1.7
23	56.6	1.7
24	57.5	1.7
25	58.5	1.7
26	59.4	1.7
27	60.4	1.7
28	61.3	1.7
29	62.3	1.7
30	63.3	1.7
31	64.3	1.7
32	65.3	1.7
33	66.4	1.7
34	67.5	1.7
35	68.6	1.7
36	69.8	1.8
37	71.0	1.8
38	72.4	2.0
39	74.2	2.4
40	77.8	3.7

*SE = Standard Error

15.2 Appendix B

Sleep Disturbance 8a *Short Form Conversion Table*		
Raw Score	**T-Score**	**SE***
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.8	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE= Standard Error on T-score metric

Adult version

15.3 Appendix C

The following conventions will be used for calculation of a subject's nominal Week 14 score. All other week scores are calculated relative to the nominal Week 14 interval.

Landmark Endpoint Visit Day	Last Available Diary Score Recorded	Days Used to Calculate "Week 14"	Comment
99	98	92-98	Ideal condition with final study visit on Day 99, last diary data will be day 98.
100	99	93-99	
101	100	94-100	
102	101	95-101	
103	102	96-102	
104	103	97-103	
105	104	98-104	
106	105	99-105	
Out of window (late)	106+	99-105	Since the Week 14 endpoint is Day 99 +/- 4 days, days after 106 are outside the Week 14 nominal interval. Any scores recorded after day 106 will be ignored.
98	97	91-97	
97	96	90-96	
96	95	89-95	
95	94	88-94	
94	93	87-93	Out of window, but allowed
93	92	85-92	Out of window, but allowed
92	91	85-91	Out of window, but allowed

<92	<91	85+	Must have a valid score occurring on/after day 85 or will be missing for week 14

15.4 Appendix D

The following list of numbers will be used for random seeds where required for MI processes:

9179445

2389546

9954263

5612645

2599732

7189411

6986208

6225777

5483414

8849556

These will generally be used in order for the primary analysis, then secondary endpoints and sensitivity analyses. For cases where identical code may be applied to more than one outcome, the second to last digit will be incremented by 1 to produce new seeds for the subsequent outcome. If a single dataset requires more than the 10 seeds above, additional seeds will be generated by incrementing the last digit in the last by 1. In no case will the same seed be used in more than one procedure; all seeds used will be documented in the programming specifications and the programs themselves.

15.5 Appendix E

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.

15.5.1 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 YYYY-MM-DD (e.g., 2001-08-01) 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 and a date stamp on the bottom of each output.

15.5.2 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 Subjects, (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., ITT >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 Subjects with non-missing values.
- All population summaries for categorical variables will include all categories that were planned and for which the subjects 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.