Statistical Analysis Plan

TRIAL FULL TITLE	EN20-01: A 24-WEEK STUDY TO EVALUATE THE SAFETY AND EFFICACY OF CNTX-6970 IN SUBJECTS WITH MODERATE TO SEVERE KNEE OSTEOARTHRITIS PAIN
SAP VERSION	5.2
SAP VERSION DATE	July 19, 2024
STUDY STATISTICIAN	Thaddeus Tarpey, PhD
ASSOCIATED PROTOCOL VERSION	9.0
PROTOCOL PRINCIPAL INVESTIGATOR	Robert Edwards, PhD
SAP AUTHORS	Thaddeus Tarpey, PhD and Eva Petkova, PhD

History of changes

Version	Date	Modification(s)
1.1	February 24, 2021	First complete version
1.2	February 24, 2021	Minor modifications based on protocol clarifications
1.3	March 18, 2021	Remove original Aim 4 since that was based on screening measures
1.4	April 15, 2021	Fix numbering of aims and associated analysis plans, add separate Safety analysis section, add original aim 4 back in
1.5	April 29, 2021	Edit updates for revised protocol version 2.2
1.6	June 17, 2021	Update Inclusion/Exclusion and missing data due to discontinuation
1.7	July 21, 2021	Update to include NRS outcome
1.8	September 30, 2021	Modifications for updated protocol based on IND submission to the FDA
1.9	December 22, 2021	PPI changed to Robert Edwards, inclusion criteria 10 and 17 updated, criteria 18–21 added, exclusion criterion modified to QRS interval $\geq$120 ms, subsection on protocol violations added
1.10	April 15, 2022	Numerous changes to inclusion/exclusion to improve recruitment, including removing synovial fluid and EMA collection
2.0	May 4, 2022	Updated the details for stratifying on K-L grade for randomization
3.0	October 1, 2022	Updated for protocol version 7.0 changes. Remove inclusion/exclusion text. Add in synovial fluid collection.
4.0	April 8, 2023	Updated for protocol version 8.0 changes. Major change is to remove the Celecoxib and 100 mg CNTX arms from the study due to drug supply. Other changes are: to remove actigraphy and QST; PROMIS Sleep Disturbance Scale removed from V2, V4, V6, and V8; PGIC to be collected at V2 and V9; Removal of HADS from Baseline Visit, V2, V4, V6, and V8; Assessment of Stair-Evoked Pain at V2, V4, V6, V8. For WOMAC-A, no longer a criterion that must complete 4 out of 5 within specified amount of time; the new criteria states that subjects must complete 4 WOMAC-As at minimum during the screening period to determine the index knee. WOMAC-A can be administered throughout screening until 4 are completed.
5.0	January 8, 2024	Update for protocol versions 8.1, 8.2, and 9.0, increase sample size to 55 and provide additional details for the statistical analysis.
5.1	March 26, 2024	Added text for backup analysis plans for secondary aims if the data is insufficient.
5.2	July 19, 2024	Clarification on fitting models with insufficient data in Section 7, minor edits for clarity. Removed Aim 9 on enzyme CYP2CP since that is not in the protocol and added mediation analysis proposal for analysis of chemokines and cytokines for aim 5.

1 SAP Signatures

I give my approval of the attached SAP entitled ***EN20-01: A 24-WEEK STUDY TO EVALUATE THE SAFETY AND EFFICACY OF CNTX-6970 IN SUBJECTS WITH MODERATE TO SEVERE KNEE OSTEOARTHRITIS PAIN***

July 19, 2024

Protocol Principal Investigator

Name: _____

Signature: _____

Date: _____

CCC Principal Investigator

Name: _____

Signature: _____

Date: _____

DCC Principal Investigator

Names: _____

Signature: _____

Date: _____

Contents

2 Abbreviations and Definitions

AE	Adverse Event
BID	bis in die (twice a day)
CI	Confidence Interval
ECG	Electrocardiography
ITT	Intention-To-Treat
IXRS	Interactive [Web/Phone] Response System
MDS	Minimal Data Set
NRS	Numerical Rating Scale
NSAIDS	Nonsteroidal Anti-inflammatory Drugs
OA	Osteoarthritis
OUQ	Opioid use Questionnaire (OUQ)
PPI	Protocol Principal Investigator
PRN	As Needed
RCT	Randomized Clinical Trial
SAD	Single Ascending Dose
SAP	Statistical Analysis Plan
SAE	Serious Adverse Events
TEAE	Treatment Emergent Adverse Events

3 Introduction

3.1 Preface

The purpose of this trial is to investigate the study drug CNTX-6970 for the treatment of osteoarthritis (OA) of the knee. Chemokines are small chemotactic peptides that control the trafficking of leukocytes, and in particular monocytes, to their target tissue. Chemokine receptors (CCRs, CXCRs, or XCRs) form a growing family of receptors (CCR1 up to CCR10, CXCR1 up to CXCR7, CX3CR1, XCR1, etc.). CNTX-6970 is a highly selective CCR2 and CCR5 antagonist which makes it appealing as a targeted treatment for persistent pain conditions characterized by inflammatory processes. The pharmacological effect of CNTX-6970 is expected to come primarily from blockade of CCR2, although higher concentrations will also inhibit CCR5 as was demonstrated in a Phase 1 single ascending dose (SAD) study. Symptoms of pain are the major contributing factor to functional impairment in mild and moderate OA, which is both quite common and highly disabling. With an aging and increasingly obese population, this condition is becoming even more prevalent than in previous decades. The expectation is that CCR2 antagonism will not only reduce pain, but also lead to improvement of function, and thus be relevant for treating signs and symptoms of OA. Therefore, this study will investigate the effect of CNTX-6970 on both pain and function.

Based upon the underlying pathophysiology, two main types of pain can be distinguished: neuropathic pain and nociceptive/inflammatory pain. The most pronounced effects of CNTX-6970 are expected to pertain to nociceptive/inflammatory pain, which is mediated by pain receptors (nociceptors) that activate afferent somatic or visceral pain pathways via afferent nerves. Psychosocial comorbidities such as depression and anxiety commonly accompany chronic pain and can affect both the pain perception and tolerance thresholds. Given that the pathophysiology of OA is poorly understood, there are currently no curative or disease-

modifying treatments approved for patients who are not yet at a stage of their condition at which total joint replacement is being considered.

In terms of safety, CCR2 antagonism has not been associated with any of the known safety issues of nonsteroidal anti-inflammatory drugs (NSAIDs). One of the appealing features of CNTX-6970 is that it does not significantly cross the blood brain barrier and therefore is not expected to have any of the central nervous system side effects associated with opioids and many other analgesics. Importantly the liability for addiction is very low.

3.2 Scope of the Analyses

Using a multi-crossover trial, to study the safety and efficacy of 300 mg (BID) CNTX-6970 for the treatment of OA of the knee.

Figure 1 illustrates the crossover design for this study.

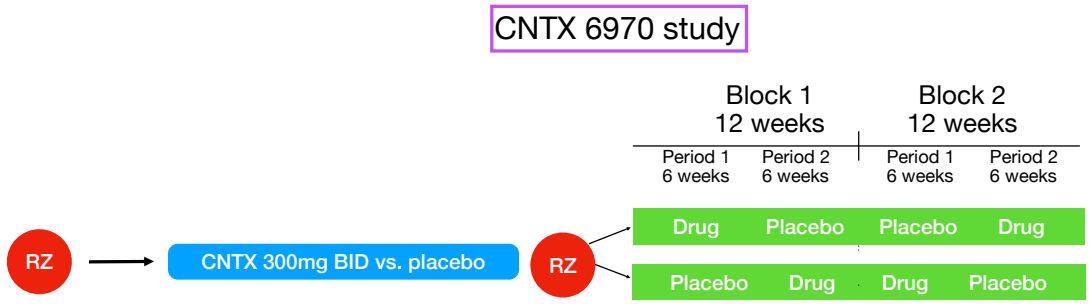

Figure 1: Schema for the Multi-Crossover Design

4 Study Objectives and Endpoints

The objectives of this study are categorized as Primary, Secondary and Exploratory.

4.1 Study Objectives

4.1.1 Primary Objectives

The primary objective of this study is to evaluate the safety and efficacy of CNTX-6970 300 mg BID for the treatment of pain related to OA of the knee compared to placebo. These objectives will be accomplished through the following specific aims listed below.

Aim 1: Assess the safety and tolerability of CNTX-6970 (300mg BID), and placebo.

There will be no hypothesis testing under Aim 1. To achieve Aim 1, all AEs and SAEs will be tabulated and classified them by severity and relatedness to treatment. Frequency of AEs and SAEs in the active and placebo groups will be tabulated with point estimates and 95% confidence intervals (CI). Similarly, we will summarize relevant laboratory measures with appropriate estimates.

Aim 2: Assess the efficacy of 300mg BID CNTX-6970 in comparison to placebo.

The primary outcome measure used to assess efficacy will be patient-reported knee pain using the WOMAC Part A, or WOMAC-A (Bellamy *et al.*, 1988).

HYPOTHESIS 1. CNTX-6970 300mg BID will be more effective than placebo with respect to pain as measured by WOMAC-A (primary outcome measure).

4.2 Secondary Objectives

The secondary objectives of this study are to evaluate the effect of CNTX-6970 (300mg BID) on general pain-related measures, including physical and psychosocial functioning, as well as biomarkers of pain and inflammation. This objective is addressed by the following specific aims:

Aim 3: Assess the efficacy of CNTX-6970 (300mg BID) in comparison to placebo with respect to secondary outcome measures related to OA on the knee: (a) WOMAC-C (function subscale) (Bellamy *et al.*, 1988); (b) Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983); (c) Patient Global Impression of Change (PGIC) (Kroenke *et al.*, 2019); (d) PROMIS Sleep Disturbance Scale – 6A (Yu *et al.*, 2011); (e) Sleep Duration Question; and (f) Daily Knee Pain Intensity on a 0-10 Numeric Rating Scale (NRS).

HYPOTHESIS 2: CNTX-6970 300mg BID will be superior to placebo with respect to symptoms measured by (a) WOMAC-C (function subscale); (b) Hospital Anxiety and Depression Scale (HADS); (c) Patient Global Impression of Change (PGIC); (d) PROMIS Sleep Disturbance Scale– 6A; (e) Sleep Duration Question; and (f) Daily Knee Pain Intensity on a 0-10 NRS.

Aim 4: Assess the efficacy of CNTX-6970 (300mg BID) in comparison to placebo with respect to general outcomes of pain: (a) Pain Catastrophizing Scale –Short Form 6 (Sullivan *et al.*, 1995); (b) PROMIS Physical Functioning Short-Form 6b (Schalet *et al.*, 2016); (c) Patient Health Questionnaire – 2 item scale (PHQ-2) – Depression (Arroll *et al.*, 2010); (d) Generalized Anxiety Disorder – 2 item scale (GAD-2) (Kroenke *et al.*, 2007; Plummer *et al.*, 2016); (e) Tobacco, Alcohol, Prescription medication, and other Substance use Tool (TAPS-1) (Gryczynski *et al.*, 2017); and (f) Opioid Use Questionnaire (OUQ).

HYPOTHESIS 3: CNTX-6970 300mg BID will be superior to placebo with respect to outcomes measured by (a) Pain Catastrophizing Scale –Short Form 6; (b) PROMIS Physical Functioning Short-Form 6b; (c) Patient Health Questionnaire (PHQ) – Depression; (d) Generalized Anxiety Disorder – 2 item scale (GAD-2) and (e) Tobacco, Alcohol, Prescription medication, and other Substance use Tool (TAPS-1); and (f) Opioid Use Questionnaire.

Aim 5: Assess the effect of CNTX-6970 (300mg BID) in comparison to placebo with respect to biomarkers of pain and inflammation: (a) Staircase-Evoked Pain Assessment; (b) serum and synovial fluid levels of chemokines and cytokines; and (c) synovial monocyte chemoattractant protein-1/CCR-2 receptor binding inhibition in blood and synovial fluid.

HYPOTHESIS 4: Compared to placebo, CNTX-6970 300mg BID will result in greater improvement with respect to (a) Staircase-Evoked Pain Assessment; (b) serum and synovial fluid levels of chemokines and cytokines; and (c) synovial monocyte chemoattractant protein-1/CCR-2 receptor binding inhibition in blood and synovial fluid.

4.3 Exploratory (Tertiary) Objectives

The tertiary objectives of this study are to obtain deep phenotyping of the target population both prior to study enrollment and during treatment, and to identify biomarkers for treatment response. This objective will be accomplished through the following specific aims:

Aim 6. If 300mg BID CNTX-6970 is more effective than placebo, evaluate the following characteristics of its effect: (a) onset of action; (b) carryover effect after treatment discontinuation.

Aim 7. Identify biomarkers related to pain from OA of the knee and to response to treatment with CNTX-6970 at 300mg BID.

Aim 8: Explore sociodemographic and clinical predictors of response to CNTX-6970.

4.4 Endpoints and Biomarkers

4.4.1 Primary Safety Endpoint

The primary safety endpoint is the incidence of treatment emergent adverse events (TEAEs), reported between the administration of study drug on Day 1 and the completion of the study at Week 24 or earlier if a patient discontinues treatment.

4.4.2 Primary Efficacy Endpoint

The primary efficacy endpoint is pain in the index knee, measured weekly using the WOMAC-A (Pain subscale) (Bellamy *et al.*, 1988). The index knee will be determined using the averaged screening period WOMAC-A scores by knee – the knee with the greater score at the end of screening is defined as the index knee going forward. Subjects must complete 4 WOMAC-A's at minimum during the screening period to determine the index knee. We will use the numerical rating scale (NRS) version of the WOMAC-A with the subject assessing each of 5 questions using an 11-point (0 to 10) scale; the total score is the sum of the individual item scores (range 0-50). A higher WOMAC score represents worse symptom severity. All WOMAC subscales show strong evidence of validity and reliability (Salaffi *et al.*, 2003), and are the most frequently used primary endpoints for OA trials because of their strong assay sensitivity (Jung *et al.*, 2018). Assessments during the last 2 weeks of each treatment period will be used to test the primary hypotheses in order to avoid possible carryover effects of unknown duration. All weekly measurements of WOMAC-A will be used to study the course of OA pain symptoms during treatment, including onset of action and carryover effects.

4.4.3 Secondary Endpoints

The following measures are considered secondary outcomes related to pain due to OA of the knee. They will be assessed at Baseline and at study visits through Week 24 specified in the protocol, except Daily Knee Pain Intensity on a 0-10 NRS which will be recorded daily the week prior to each study visit.

- Daily Knee Pain Intensity on a 0-10 Numeric Rating Scale (NRS). Pain intensity is reported by patients with chronic pain as one of the most important targets of treatment, and daily pain intensity ratings are a recommended core outcome measure for clinical trials of treatments for chronic pain. Daily ratings are preferable to ratings of recalled pain over longer time periods such as a week, as daily ratings minimize the influence of recall biases (Dworkin *et al.*, 2005). For the NRS, participants provide one-daily report (at the end of each day over the course of a week) of their average knee pain intensity on a 0-10 pain intensity NRS. Subjects will record their Daily Pain Intensity NRS 0-10 each day for one week prior to each clinic visit using NEForm (except before randomization, Daily Knee Pain Intensity on a 0-10 NRS will be collected at the end of the day).
- WOMAC-C (Function subscale) (Bellamy *et al.*, 1988). The WOMAC-physical function subscale contains 17 items assessing daily functioning, each using an 11-point (0 to 10) numerical rating scale. The total index score (0-170) is the sum of the items. A higher WOMAC function score represents worse functioning and less ability to engage in daily activities.

- HADS (Hospital Anxiety and Depression Scale) (Zigmond and Snaith, 1983). The HADS is a 14-item self- report questionnaire designed to assess symptoms of anxiety and depression in those with medical illness (Norton *et al.*, 2013). It has well-established reliability and validity in the assessment of symptoms of depression and emotional distress, and has been used in numerous clinical trials. It does not include somatic symptoms, such as fatigue and sleeplessness, which may otherwise be attributable to physical illness, and it has been standardized among large community samples. It has also been validated in several medical illness populations with good sensitivity and specificity for predicting DSM-IV major depression or generalized anxiety disorder diagnoses. The HADS has been recommended as a psychosocial phenotyping measure in clinical trials of treatments for chronic pain (Edwards *et al.*, 2016). This scale has 14 items, 7 related to anxiety and 7 to depression, rated on 4 points (0 to 3) in domains of intensity or frequency. Scoring is done separately for depression and for anxiety and interpreted as normal for scores of 0-7, borderline abnormal (borderline case) for scores of 8-10 and abnormal (case) for scores of 11-21. This scale is used to assess depression and anxiety in addition to HEAL/EPPIC-Net core data elements (CDEs) because of its higher sensitivity to change especially in participants with medical illnesses.
- PROMIS Sleep Disturbance – 6A (Yu *et al.*, 2011). Sleep disruption has a bi-directional relationship with chronic pain and is an important secondary outcome to measure in pain trials (Edwards *et al.*, 2016). The Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance short form is a convenient 6-item scale that correlates strongly with the longer forms. It shows greater measurement precision for assessing sleep disturbance than other commonly-used (and much longer) questionnaires such as the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale; its brevity and convenience are a major advantage for both research and clinical settings (Yu *et al.*, 2011). The PROMIS Sleep Disturbance Scale is expressed as a T-score, with a population mean of 50 and SD of 10. Possible T scores in this distribution range from 31.7 to 76.1 (`http://www.healthmeasures.net/images/promis/manuals/PROMIS_Sleep_Disturbance_Scoring_Manual.pdf`).
- Patient Global Impression of Change (PGIC) (Kroenke *et al.*, 2019). The PGIC is a single-item measure of patient-reported improvement that is widely used as a general outcome measure in studies of chronic pain participants, including OA participants (Salaffi *et al.*, 2003). It is often used as an index of treatment- associated change, and patient-reported improvements in the form of PGIC scores correlate robustly with significant improvement in pain intensity, pain interference with activities of daily living, mood, and quality of life (Perrot and Lantéri-Minet, 2019).

The following measures are part of the HEAL/EPPIC-Net CDEs and are considered general pain-related outcomes. They will be assessed at baseline and at week 24/ET.

- Pain Catastrophizing Scale – Short Form 6 (Sullivan *et al.*, 1995). Catastrophizing is a pain-specific psychosocial construct comprising cognitive and emotional processes such as helplessness, pessimism, rumination about pain-related symptoms, and magnification of pain reports. The short-form Pain Catastrophizing Scale (PCS), is a 6-item, self-report measure of catastrophic thinking associated with pain (McWilliams *et al.*, 2015). Scores range from 0-24, with higher scores indicating more catastrophizing.
- PROMIS Physical Functioning Short-Form 6b (Schalet *et al.*, 2016). The PROMIS Physical Functioning short form is a 6-item scale that is widely used in pain research. It is a unidimensional scale that shows broad coverage of the physical function construct, good construct validity, and high levels of temporal stability (Schalet *et al.*, 2016). The PROMIS Physical Function Scale is expressed as a T-score, with a population mean of 50 and SD of 10. Higher scores represent better physical functioning; possible T scores in this distribution range from 21 to 59.
- Sleep Duration Question. A single-item scale measuring the duration of actual sleep a participant has gotten, on average, over the past month. Numerical responses will be provided in hours and minutes.
- Pain Health Questionnaire (PHQ) – Depression (Arroll *et al.*, 2010). The 2-item PHQ-2 is a brief depression screening tool that correlates strongly with PHQ-9 scores and shows good sensitivity for identifying individuals with depressive disorders in the general population and in a variety of medical

samples (Arroll *et al.*, 2010). Scores range from 0-6, with higher scores indicating more depressive symptomatology.

- Generalized Anxiety Disorder – 2 item scale (GAD-2) (Kroenke *et al.*, 2007; Plummer *et al.*, 2016). The GAD-2 is a 2-item screening tool that is widely used to screen for clinically significant anxiety symptoms and anxiety disorders in clinical settings. It shows good sensitivity and specificity as a screening tool for anxiety disorders (Kroenke *et al.*, 2007). Scores range from 0-6, with higher scores indicating more anxiety symptomatology.
- Tobacco, Alcohol, Prescription medication, and other Substance use Tool (TAPS-1) (Gryczynski *et al.*, 2017). The TAPS-1 is the screening component of the TAPS tool and consists of a single stem question with four items covering frequency of past-12-month use of tobacco, alcohol, and illicit drugs, and non-medical use of prescription medications. Scores range from 0-4; higher scores indicate a higher likelihood of problematic substance use. The TAPS-1 shows good sensitivity and specificity for identifying substance use disorders (Gryczynski *et al.*, 2017).
- Opioid Use Questionnaire (OUQ). The OUQ is an indicator of past or present use of any of the listed opioid medications. There are a total of three yes/no items where a yes indicates the use of such medications.

The following are secondary physiological outcomes related to pain and potential exploratory biomarkers for pain and treatment response.

- Staircase-evoked pain assessment. This procedure consists of stepping fully up and down onto a 20 cm high platform with both feet a total of 24 times. The lead leg is alternated between each up/down cycle. Subjects are instructed to use their normal gait for completing this task and are encouraged to complete the task despite increasing pain, without stopping if possible. The procedure is timed, and current knee pain intensity on a 0-10 Numeric Rating Scale (NRS) is assessed immediately before and following the procedure while the subject is in a seated, resting position. This staircase-evoked pain assessment has been well tolerated in multiple studies of knee OA participants to date, and the standardized effect size for anti-inflammatory treatments was shown to be considerably higher with the staircase measure than with other measures such as pain intensity ratings (Treister *et al.*, 2019).
- Temporal profile of OA pain will be assessed using the NRS (integer scale from 0-10) at the end of the day for the first seven days following the Screening visit in order to obtain a measure of pain variability **which will be used to screen out individuals with high variability (SD $>$ 1.2) in their pain**. Additionally, after Baseline, the NRS will be completed daily for the seven days immediately preceding all study visits.
- Serum levels of cytokines and chemokines. A specific list is provided in the Appendix of the protocol.
- Monocyte chemoattractant serum protein-1(MCP-1)/CCR-2 receptor binding inhibition by CNTX-6970. This test provides as a single score, expressed as a percentage, 0-100%, with higher scores indicating more binding inhibition.
- Monocyte chemoattractant protein-1 (MCP-1)/CCR-2 receptor binding inhibition by CNTX-6970 in synovial fluid.
- Synovial fluid levels of cytokines and chemokines (a specific list is provided in Appendix of the protocol), assessed at the end of the first treatment period (week 6) which will allow a comparison of subjects treated with CNTX-6970 300 mg BID to placebo-treated subjects.

5 Study Methods

5.1 General Study Design and Plan

This is a multi-crossover design with 2 blocks and 2 periods per block. Each period lasts 6 weeks. Patients will be randomized to one of the treatment sequences shown in Figure 1. The Schedule of Events Table in

the Protocol gives the timing of the outcome assessments. In general, we will be analyzing the data on an intent-to-treat basis (ITT).

5.2 Inclusion and Exclusion Criteria and General Study Population

Specifics on the inclusion/exclusion criteria are provided in the protocol. Subjects must complete at least 5 NRS's throughout the entire screening period. The following is blinded inclusion criteria not listed in the protocol:

- The standard deviation of these End of Day Pain Ratings (for average pain over the last 24 hours) from the screening period must be less than or equal to 1.2.
- Left Knee Average WOMAC Score range must be 20-45 (calculated average of pain scores across screening timepoints) *OR* Right Knee Average WOMAC Score range must be 20-45 (calculated average of pain scores across screening timepoints). These requirements will be reviewed and approved by CCC staff before the SAFER interview is scheduled.

The rationale for excluding individuals with high variability in their daily average pain measures, as determined by their NRS responses during the first week of screening, is to not include individuals with unreliable pain measures and not include individuals prone to high placebo response (Farrar *et al.*, 2014).

5.3 Randomization and Blinding

For entry into the study and randomization, all eligibility criteria MUST be met (i.e., all inclusion criteria and no exclusion criteria specified in the protocol). Patients who meet inclusion/exclusion criteria at the Baseline Visit (day 0) will be randomized. After documentation of written informed consent and the screening period, clinical sites will confirm eligibility and complete randomization by accessing the Almac's centralized web-based system for randomization and drug management - the Interactive Web Response System (IXRS). Subjects will be randomized equally between treatment sequences shown in Figure 1: Drug:Placebo:Placebo:Drug or Placebo:Drug:Drug:Placebo. A standard block randomization will be used, stratifying by K-L grade (stratum 1 K-L grade 1 or 2, and stratum 2 K-L grade 3 or 4) to achieve balance on randomization sequences. The randomization will be implemented by the IXRS vendor in consultation with the study statisticians. After Randomization, drug will be dispensed.

Blinding Procedures

Eligible subjects will be randomized through the IXRS. Blinding will be assured by restricting access of site and Sponsor personnel and/or designees to the randomization sequences, and by providing identical tablets and packaging for the placebo and 300 mg CNTX-6970 tablets. For the final analysis, randomization sequences for all subjects will be released after all subjects have completed the study and the clinical database is locked. For the DSMB safety reviews, the randomization sequences will be released to an independent statistician and a programmer to produce unblinded reports. The Sponsor and the Investigators will remain blinded.

Unblinding of individual randomization sequences during the study is discouraged. The Investigator at a site may break the blind for a given subject in the event of a medical emergency, where knowledge of the subject's randomization sequence must be known in order to facilitate appropriate emergency medical treatment. The Investigator should attempt to contact the study Medical Monitor before unblinding a subject's randomization sequence identity in order to obtain concurrence that unblinding a subject's randomization sequence is necessary. If not reasonable prior to unblinding, investigators should notify the Medical Monitor as soon as possible after an unblinding event. Once unblinded, a subject must be permanently withdrawn from study participation.

6 Sample Size

The sample size for this study was determined based on the primary efficacy aim to detect clinically meaningful and realistic effects of CNTX-6970 (300 mg BID) compared to placebo in the multi-period crossover

design model. Power was computed for effect sizes ranging from 0.25 to 0.50 using Monte Carlo simulation. Data were simulated from multi-period crossover models with block, period, and carryover effects, and varying within and between subject variability.

Although carryover effects can be efficiently isolated and estimated in the analysis stage by using the mixed-effects models proposed in Section 7, power for the primary hypothesis was computed using only WOMAC-A assessed during the last 2 of 6 weeks per treatment period to minimize carryover effects. The power of the test depends on the relationship of the between-subject variability to the within-subject variability. In general, the between-subject variability is larger than the within-subject variability, but a range of Between/Within (B/W) variance ratios (5/1, 2/1) were investigated. With a sample size of $n = 55$ per arm, using a two-sided Type I error rate of $\alpha = 0.05$, the power to detect an effect size (for the active treatment vs. placebo) of Cohen's $d = 0.35$ exceeds 95% when the ratio $B/W = 2/1$.

Assuming a 15% early termination (e.g., dropout) rate during each of the 4 periods will yield a sample size of approximately $n = 29$ with complete observations at all treatment periods. If we conservatively assume that only subjects with complete data will be analyzed, the corresponding power with $n = 29$ exceeds 80% when the ratio $B/W = 2/1$ with an effect size as low as 0.30. Figure 2 shows the results of a simulation study to evaluate the power of this design. Because we will use all available data from all enrolled subjects in the analysis, the actual power in the case of 15% early termination per treatment period will be larger than what Figure 2 shows. If the ratio B/W is larger than 2, say B/W=3, 4 or 5, the power for detecting the treatment effects is much larger. For example, with only 29 participants, if B/W=5, the power to detect an effect of magnitude Cohen's $d = 0.35$ exceeds 0.95. Note that the reported test-retest reliability of the WOMAC-A measure is 0.86, which corresponds to a $B/W = 7$.

7 General Analysis Considerations

All demographic and baseline variables will be summarized by means, standard deviations, medians, and interquartile ranges (IQR) for continuous measures. Frequency and percentages will be reported for all categorical measures (including binary measures).

The general approach for statistical inference for aims associated with outcomes measured at multiple periods will utilize a mixed-effects multi-period crossover model with 2 blocks and 2 treatment periods per block (of length 6 weeks). Maximum likelihood estimation will be used primarily for parameter estimation and likelihood ratio tests will be used to test hypotheses under the primary and secondary objectives. Significance of an active treatment effect versus placebo will be determined by testing if the treatment effect coefficient differs from zero in the mixed-effect model. When there is a possibility of carryover effects, the multi-crossover model will incorporate carryover effect terms. Only carryover effects from active to placebo will be formally modelled. For outcomes recorded only at baseline and week 24, an appropriate generalized linear model will be used to estimate and test for the active treatment effect, controlling for covariates noted below and also the baseline value of the outcome.

All hypothesis tests will be two-sided using $\alpha = 0.05$ significance level, and p-values unadjusted for multiple testing will be reported. The justification for this decision is that this is Phase II investigation and all assessed outcomes are of specific and unique interest. Much more important than testing hypotheses is the efficient estimation of all quantities of interest, and this will be carried out throughout the analyses. Treatment effect point estimates with 95% confidence intervals will be reported along with estimates of other model parameters, regardless of whether the null hypotheses were rejected or not.

The secondary and exploratory analyses proposed here are contingent on sufficient data availability after database lock. If the data for any outcome is insufficient for the proposed analysis, then the analysis that will be performed will be considered exploratory in nature. If data is insufficient to fit complex models in these cases (e.g., due to singularities or convergence issues), then the model will be simplified as necessary, e.g., fitting the model without the site random effect if that solves convergence or singularity issues. If this does not solve the problem, then summary statistics will be reported for the outcome in order to compare the active treatment to the placebo treatment.

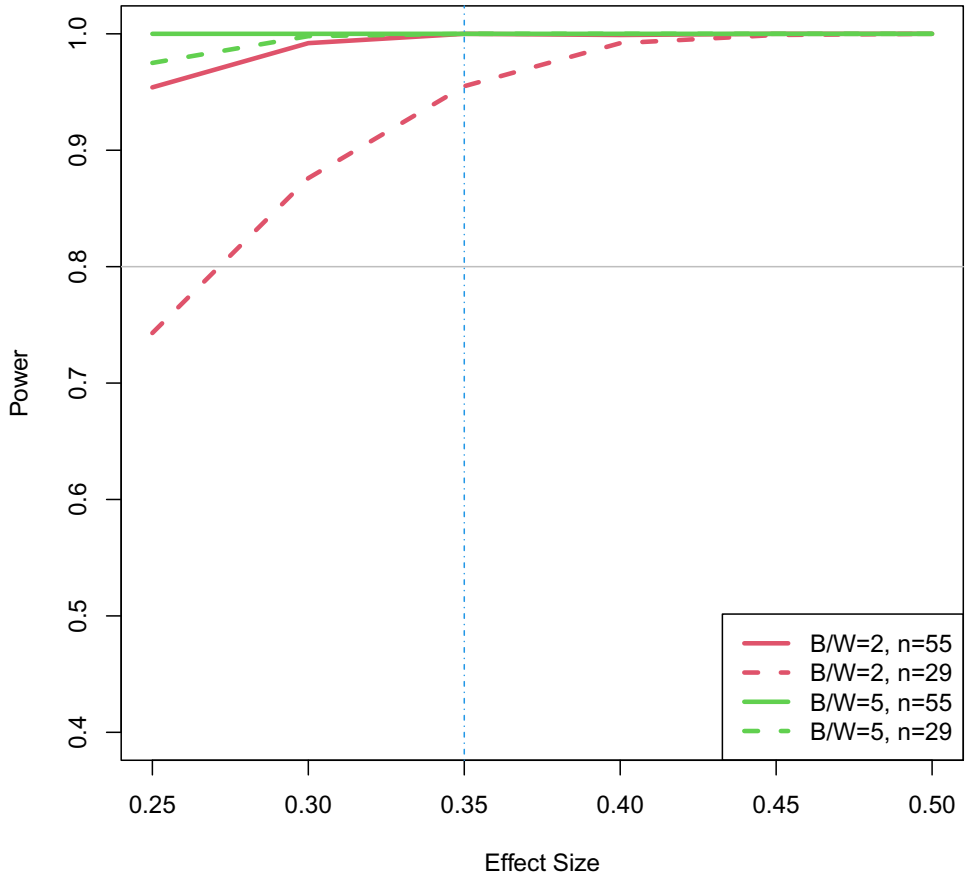

Figure 2: Power versus Cohen's d effect size. B/W indicates the ratio of between-subject to within-subject variance of the WOMAC-A measure. The proposed sample size is $n = 55$. With 15% dropout per treatment period, there will be $n = 29$ patients with complete observations at all treatment periods of this multi-crossover design. Assuming that only data from completers will be used in the analysis gives a lower bound of the power that can be achieved for the given range of effect sizes.

7.1 Timing of Analyses

There are no planned interim analyses for this trial. The final analysis will occur when the study finishes (after data lock).

7.2 Analysis Populations

This study is designed to evaluate changes in pain related to primary OA of the knee using an ITT principle. Eligible subjects will have evidence of chronic knee OA with a history of moderate to severe pain in the designated index knee for a minimum of 6-months prior to screening. Although this condition commonly affects multiple joints and subjects may have bilateral OA of the knee(s), the efficacy assessments will be conducted only on the index knee.

7.3 Covariates and Subgroups

All analyses will adjust for K-L grade, age, sex, and site (as a random effect if the data is sufficient to model a site random effect).

7.4 Missing Data

A patient who remains in the study for the first two treatment periods or for all four treatment periods will receive a balanced exposure to both treatments. Because the main analysis is based on maximum likelihood estimation using a mixed-effects multi-period crossover model, the inferences will be unbiased under missing at random (MAR) assumptions. No imputation of missing data is planned for this study, unless intensive investigations indicate that reliable and valid data imputation is possible, in which case multiple imputation will be performed and the analysis of the imputed data sets will serve as a sensitivity check of the results from the primary analysis based on the mixed-effects models (van Buuren and Groothuis-Oudshoorn, 2011).

If a patient discontinues treatment prior to completing the entire 24 weeks of study, all their data (including weekly assessments) from their time in the study up to time of discontinuation will be used in the analyses.

Should a patient discontinue study medication and decide to withdraw consent, an end of study visit should be scheduled at the next of the closest following clinic visit. Should a patient discontinue study medication, but not withdraw consent, best efforts will be made by the site to collect all study data at the regular study assessment times through the next of the closest following clinic visits. At this clinic visit, all assessments outlined for the end of study visit should be conducted. For subjects who continue to be followed for safety to the end of the period, adverse events should continue to be reported.

7.5 Interim Analyses and Data Monitoring

No interim analysis is planned for this study.

8 Safety Analyses

A primary objective of this trial is to evaluate the safety of CNTX-6970 300 mg BID for the treatment of pain related to OA of the knee compared to placebo. The primary safety endpoint is the incidence of treatment emergent adverse events (TEAEs), reported between the administration of study drug on Day 1 and the completion of the study at Week 24.

Aim 1: Assess the safety CNTX-6970 (300 mg BID).

There will be no hypothesis testing under Aim 1. All AEs and SAEs observed during different treatment periods will be summarized based on severity and relation to the active treatment (CNTX-6970 300mg BID). The reports will include tabulation of the AEs and SAEs by period, block, and sequence; point estimates and 95% CIs will be reported. Similarly, we will summarize relevant laboratory measures which include chemistry and Complete Blood Count (CBC) with appropriate estimates.

The primary safety endpoint is the incidence of treatment emergent adverse events (TEAEs), reported between the administration of study drug on Day 1 and the completion of the study at Week 24 or Early Termination. Specifically, the primary safety endpoint for the study is the following: Percent of adverse events and serious adverse events in the final 2 weeks of each treatment period. Vital signs (heart rate, blood pressure, etc.) will be obtained regularly throughout the study at the time of study visits. Height and weight will be collected during the screening period and at the subject's final study visit. Blood chemistry and hematology panels will also be obtained regularly at study visits. All ECGs will be documented by recording date, time, heart rate, QRS duration, PR interval, RR interval, QT, and QTcF on the eCRF. Paper strips will be maintained in the patient file. If indicated, additional ECG assessments can be made at the discretion of the investigator.

The severity of each AE will be graded on a 4-point scale and reported in detail as indicated on the eCRF. Vital signs, systemic effects, and laboratory abnormalities will be graded using the Toxicity Grading Scale

for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials, available at https://www.fda.gov/media/73679/download (See MOP Appendix M).

- Grade 1 (Mild: No interference with activity)
- Grade 2 (Moderate: Some interference with activity, not requiring medical attention)
- Grade 3 (Severe: Prevents daily activity and requires medical intervention)
- Grade 4 (Potentially Life Threatening: Emergency Room visit or hospitalization).

9 Efficacy Analyses

This section describes the statistical analyses for the objectives defined by the aims and hypotheses described in Section 4.

9.1 Primary Efficacy Analysis

For the efficacy Aim 2, data from the WOMAC Part A (WOMAC-A) assessed only in the last two weeks of each 6-weeks treatment period will be used for this primary analysis in order to avoid any carryover effects, which are expected to vanish after 4 weeks. Hypothesis 1 in Section 4 will be tested using the following random-effects multi-period crossover model that employs block and period effects to compare the active treatment to placebo (Zucker *et al.*, 1997; Chen and Chen, 2014). Specifically, the model is:

$$y_{i(s)jkm} = \beta_0 + \gamma_j + \delta_k + \beta_1 X_{i(s)jk} + \boldsymbol{\beta}_2' \boldsymbol{w}_{i(s)} + b_s + b_{i(s)} + \epsilon_{i(s)m}, \tag{1}$$

where $i(s)$ is the index for patient i at site s; β_0 is the intercept; j is the index for block, $j = 1, 2$ and γ_j is the block effect; k is the index for period, $k = 1, 2$, and δ_k is the period effect; m is the measurement time; $y_{i(s)jkm}$ is the primary outcome for subject $i(s)$, block j, period k, time m; β_1 represents the treatment effect, and $X_{i(s)jk}$ is an indicator for active treatment ($X_{i(s)jk} = 1$) or placebo ($X_{i(s)jk} = 0$) for patient $i(s)$, in block j, during period k; $\boldsymbol{w}_{i(s)}$ is a vector of baseline characteristics (e.g., age, sex, K-L grade indicator) for patient $i(s)$ and $\boldsymbol{\beta}_2$ is a vector of coefficients for those baseline characteristics; $b_s \sim N(0, \sigma^2_{\text{site}})$ is a random site effect and $b_{i(s)} \sim N(0, \sigma_b^2)$ is a subject-level random effect (nested within site); $\epsilon_{i(s),m}$ is a random error with an $AR(1)$ structure: $\epsilon_{i(s),m} = \phi \epsilon_{i(s),m-1} + u_{i(s)m}$, with $u_{i(s)m} \sim N(0, \sigma_u^2)$. The subject-level random effects will model the repeated measures within subjects. The block and period effect terms will accommodate changes in mean pain scores across blocks and periods of the design.

The primary hypothesis of interest for this model is $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. Maximum likelihood estimation will be used to estimate the model parameters using the statistical software R and likelihood ratio tests will be performed to test and obtain p-values for the hypotheses of treatment effects (active versus placebo) for the active treatment. Under the null hypothesis, the likelihood ratio test will have an approximate chi-square distribution with 1 degree of freedom. The hypothesis testing proposed here is based on the WOMAC-A assessed only in the last 2 weeks of each 6-week treatment period. This provides a more than adequate 4-week washout interval for CNTX-6970. The adequacy of this 4-week period is based on extensive pharmacokinetics (PK) and pharmacodynamics (PD) of the study agent CNTX-6970 provided in the Centrexion dossier.

Additionally, for the analysis under Aim 3 described below, we explicitly model the potential carryover effects using model (2),

The proposed model (2) utilizes a standard approach to adjust for carryover effects in a multi-period crossover design when estimating the treatment effect, and allows for testing for the existence of carryover effect (Chen and Chen, 2014).

9.2 Secondary Efficacy Analyses

The outcomes for Aims 3, 4 and 5 in Section 4 are assessed at multiple time points throughout the study.

Aim 3. Outcome (a) WOMAC-C is assessed at baseline and each study visit through week 24. Outcome (b) HADS (anxiety and depression scales which will be modeled separately) is assessed at in-person visits except baseline. Outcome (c) PGIC is assessed at week 3 and week 24 only. Outcome (d) PROMIS Sleep Disturbance Scale – 6A is assessed at baseline and at in-person visits through week 24. Outcome (e) Sleep Duration Question is assessed at baseline and week 24 only. Outcome (f) Daily NRS is collected daily for the week prior to each study visit and these daily NRS values will be modeled in the analysis. The analysis plan for outcomes (a), (b), and (d), and (f) will be similar to the analysis for aim 2 except that the model will adjust for possible carryover effects in placebo-treated periods that are preceded periods with the active drug treatment. To incorporate the modeling of carryover effects, we will use a modification of model (1):

$$y_{i(s)jkm} = \beta_0 + \gamma_j + \delta_k + \beta_1 X_{i(s)jk} + \lambda_{drg} Z_{i(s)jk,drg} + \boldsymbol{\beta}_2' \boldsymbol{w}_{i(s)} + b_s + b_{i(s)} + \epsilon_{i(s)m}, \quad (2)$$

where λ_{drg} is used to model carryover effects with $Z_{i(s)jk,drg} = 1$ if the treatment at block j, period $k-1$ is drug for patient $i(s)$ in site s; all other terms in the model are defined as in (1). Hypothesis 4 for these measures correspond to $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$ from model (2).

For PGIC, this instrument asks if the overall pain since the start of the study is "Very much improved" (score 0), "Much improved" (score 1), "Minimally improved" (score 2), "No Change" (score 3), "Minimally worse" (score 4), "Much worse" (score 5), "Very much worse" (score 6). The analysis for outcome (c) PGIC will use a proportional odds regression with outcome PGIC at week 24 (V9), controlling for treatment, age, sex and baseline WOMAC-A; this model will not control for week 3 PGIC since the instrument is asking about the change in pain. Note also that the treatment in the last period is the same treatment as in the first period, so a comparison of the active to the placebo in terms of change in pain can be done using week 24 PGIC outcomes for the two different treatment sequences used in this trial. The analysis plan for outcome (e) will follow the same approach as in aim 4 below.

Aim 4. Outcomes (a)-(f) are assessed at baseline and week 24. Testing hypotheses for these outcomes will not benefit from the multi-period crossover design because they are assessed only at baseline and week 24. Instead, we will test the hypotheses by comparing subjects taking active drug versus placebo at the last treatment period ending on week 24. Regressions will be used to model the outcome at week 24 as a function of treatment, the baseline level of the outcome, age, sex and K-L Grade. A standard linear regression will be used for continuous outcomes; a proportional odds model will be used for the non-binary ordinal outcomes defined by discrete categories (e.g., PHQ-2, GAD) and logistic regression models will be used for binary outcomes (e.g., OUQ).

Aim 5: The hypotheses under aim 5 are for staircase-evoked pain assessment, serum and synovial fluid levels of chemokines and cytokines; and synovial monocyte chemoattractant protein-1/CCR-2 receptor binding inhibition in blood and synovial fluid. Hypotheses for Staircase-Evoked Pain Assessment, will be tested with the same approaches as for testing the hypotheses under Aim 3 which can accommodate outcomes on participants measured at more than one occasion. For synovial fluid, which is only collected once, the hypothesis testing will use the same approach as for hypotheses under Aim 4 - we will model the week 6 assessment of measures as functions of treatment in the first period, adjusting for sex, age and K-L Grade. The analysis for serum chemokines/cytokines will be more complicated since these measures are obtained at multiple occasions and we want to assess if their impact on pain differs due to treatment in this multi-crossover study. Analytical methods for this complex type of mediation analysis still need to be developed. We will evaluate the impact of the treatment sequence on the change in the WOMAC-A from baseline to V9, modeling serum and synovial fluid levels of chemokines and cytokines as potential mediators of treatment effect by investigating a generalization of the 2×2 crossover mediation modeling approaches introduced by (Josephy *et al.*, 2015) and generalized to this multi-crossover trial design.

9.3 Exploratory Analyses

Aim 6. If CNTX-6970 is more effective than placebo, we will evaluate the following characteristics of its effect: (a) onset of action; (b) carryover effect after treatment discontinuation. We will study the change of symptoms over time using a time trend in our statistical models within each of the 6-week treatment periods to model the weekly WOMAC-A measure. Under this aim, the interest is not in hypothesis testing, but

rather in understanding the progression and magnitude of onset of action and the carryover effects of the CNTX-6970 compound. This approach will allow an elucidation of the nature of action of the compound. Specifically, we will study the change of symptoms over time using the following mixed-effect multi-period crossover model that incorporates a time trend within each of the 6-week treatment periods to model the weekly WOMAC-A measure:

$$\begin{aligned} y_{i(s)jkm} &= \beta_0 + \gamma_j + \delta_k + \beta_1 X_{i(s)jk} + \alpha_0 t_{i(s)jk} + \alpha_1 X_{i(s)jk} t_{ijkm} + \lambda_{0.drg} Z_{i(s)jk,drg} + \lambda_{1.drg} Z_{i(s)jk,drg} t_{i(s)jkm} + \\ &\quad + \boldsymbol{\beta}_2' \boldsymbol{w}_{i(s)} + b_s + b_{i(s)} + \epsilon_{i(s)m}, \end{aligned} \tag{3}$$

where $t_{i(s)jkm}$ is the patient's $i(s)$ time measure (at site s) within block j and period k at the mth assessment and the common notations are defined as in equation (1). The coefficient α_0 is the slope for change of symptoms during a placebo treatment period, when that period was preceded by none or a placebo treatment period; α_1 (the slope of the change of symptoms during an active treatment period) is a measure of the speed of the onset of action of the active drug; counting week 6 of the previous period as the baseline for the next period, $\lambda_{0.drg}$ is the active treatment's carryover effect at the start of a placebo treatment period that follows an active treatment period; $\lambda_{1.drg}$ is the difference in the slopes of treatment change during placebo treatment between two conditions: when the placebo treatment period is preceded by active treatment versus when the placebo treatment period is preceded by no or placebo treatment. Thus $\lambda_{1.drg}$ will measure the speed of carryover effect disappearance after the active treatment is discontinued.

Aim 7. The biomarker-related objectives are to identify physiological and behavioral measures that are related to established indices of pain from OA of the knee (e.g., WOMAC-A), as well as general pain-related measures (e.g., depression, anxiety, sleep). All results from these discovery analyses will be reported as exploratory and will be subject to independent validation. The interpretation of the results will also be in the context of all analyses that were performed and will emphasize the discovery nature of the work. Statisticians on the study team have developed and published methods for analysis of complex multimodal data for developing optimal treatment decision rules (e.g., Petkova *et al.*, 2016; Ciarleglio *et al.*, 2018; Park *et al.*, 2017, 2020a,b) which is a primary objective of precision medicine. Although this previous work has been primarily motivated by mental health research (e.g., depression), these methodologies are very general and are applicable to pain and other medical conditions. As necessary, new methodologies will be developed for the analysis of the data, collected simultaneously over several weeks. The results of these investigations are expected to be features (for example, the variability of pain across days as measured by daily NRS scores) or combinations of features that constitute biosignatures for specific response to CNTX-6970 and placebo response.

Aim 8: This aim will explore sociodemographic and clinical predictors of response to CNTX-6970 by examining the relationship between treatment with CNTX-6970 and placebo, and sociodemographic and clinical covariates. No hypotheses will be tested under this aim. Rather various research questions will be addressed, such as:

Question 1. Do women show a greater reduction in pain ratings following CNTX-6970 compared to placebo, relative to men?

Question 2. Do patients with high levels of psychological distress show a greater reduction of pain with CNTX-6970, relative to patients who do not have high levels of psychological distress?

References

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K., and Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine* **8**:348–353.

Bellamy, N., Buchanan, W., Goldsmith, C., Campbell, J., and Stitt, L. (1988). Validation study of womac: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *The Journal of rheumatology* **15**:1833—1840.

Chen, X. and Chen, P. (2014). A comparison of four methods for the analysis of N-of-1 trials. *PLOS ONE* **9**:1–12.

Ciarleglio, A., Petkova, E., Ogden, T., and Tarpey, T. (2018). Constructing treatment decision rules based on scalar and functional predictors when moderators of treatment effect are unknown. *Journal of the Royal Statistical Society. Series C, Applied statistics* **67**:1331.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., Kerns, R. D., Stucki, G., Allen, R. R., Bellamy, N., Carr, D. B., Chandler, J., Cowan, P., Dionne, R., Galer, B. S., Hertz, S., Jadad, A. R., Kramer, L. D., Manning, D. C., Martin, S., McCormick, C. G., McDermott, M. P., McGrath, P., Quessy, S., Rappaport, B. A., Robbins, W., Robinson, J. P., Rothman, M., Royal, M. A., Simon, L., Stauffer, J. W., Stein, W., Tollett, J., Wernicke, J., and Witter, J. (2005). Core outcome measures for chronic pain clinical trials: Immpact recommendations. *Pain* **113**:9–19.

Edwards, R. R., Dworkin, R. H., Turk, D. C., Angst, M. S., Dionne, R., Freeman, R., Hansson, P., Haroutounian, S., Arendt-Nielsen, L., Attal, N., Baron, R., Brell, J., Bujanover, S., Burke, L. B., Carr, D., Chappell, A. S., Cowan, P., Etropolski, M., Fillingim, R. B., Gewandter, J. S., Katz, N. P., Kopecky, E. A., Markman, J. D., Nomikos, G., Porter, L., Rappaport, B. A., Rice, A. S. C., Scavone, J. M., Scholz, J., Simon, L. S., Smith, S. M., Tobias, J., Tockarshewsky, T., Veasley, C., Versavel, M., Wasan, A. D., Wen, W., and Yarnitsky, D. (2016). Patient phenotyping in clinical trials of chronic pain treatments: Immpact recommendations. *Pain* **157**:1851–1871.

Farrar, J. T., Troxel, A. B., Haynes, K., Gilron, I., Kerns, R. D., Katz, N. P., Rappaport, B. A., Rowbotham, M. C., Tierney, A. M., Turk, D. C., and Dworkin, R. H. (2014). Effect of variability in the 7-day baseline pain diary on the assay sensitivity of neuropathic pain randomized clinical trials: An ACTTION study. *Pain* **155**:1622–1631.

Gryczynski, J., McNeely, J., Wu, L., Subramaniam, G. A., Svikis, D. S., Cathers, L. A., Sharma, G., King, J., Jelstrom, E., Nordeck, C. D., Sharma, A., Mitchell, S. G., O'Grady, K. E., and Schwartz, R. P. (2017). Validation of the taps-1: A four-item screening tool to identify unhealthy substance use in primary care. *Journal of General Internal Medicine* **32**:990–996.

Josephy, H., Vansteelandt, S., Vanderhasselt, M.-A., and Loeys, T. (2015). Within-subject mediation analysis in AB/BA crossover designs. *Int. J. Biostat.* **11**:1–22.

Jung, S. Y., Jang, E. J., Nam, S. W., Kwon, H. H., Im, S. G., Kim, D., Cho, S. K., Kim, D., and Sung, Y. K. (2018). Comparative effectiveness of oral pharmacologic interventions for knee osteoarthritis: A network meta-analysis. *Modern Rheumatology* **28**:1021–1028.

Kroenke, K., Krebs, E. E., Turk, D., Korff, M. V., Bair, M. J., Allen, K. D., Sandbrink, F., Cheville, A. L., DeBar, L., KA, K. A. L., and Kerns, R. D. (2019). Core outcome measures for chronic musculoskeletal pain research: Recommendations from a veterans health administration work group. *Pain Medicine* **20**:1500–1508.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., and Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine* **146**:317–325.

McWilliams, L. A., Kowal, J., and Wilson, K. G. (2015). Development and evaluation of short forms of the pain catastrophizing scale and the pain self-efficacy questionnaire. *European Journal of Pain* **19**:1342–1349.

Norton, S., Cosco, T., Doyle, F., Done, J., and A.Sacker (2013). The hospital anxiety and depression scale: a meta confirmatory factor analysis. *Journal of Psychosomatic Research* **74**:74–81.

Park, H., Petkova, E., Tarpey, T., and Ogden, R. T. (2017). Projection pursuit for characterizing interactions between a treatment and a large number of covariates. *In preparation* .

Park, H., Petkova, E., Tarpey, T., and Ogden, R. T. (2020a). A constrained single-index regression for estimating interactions between a treatment and covariates. *Biometrics* .

Park, H., Petkova, E., Tarpey, T., and Ogden, R. T. (2020b). A sparse additive model for treatment effect-modifier selection. *Biostatistics. doi:10.1093/biostatistics/kxaa032* .

Perrot, S. and Lantéri-Minet, M. (2019). Patients' global impression of change in the management of peripheral neuropathic pain: Clinical relevance and correlations in daily practice. *European Journal of Pain* **23**:1117–1128.

Petkova, E., Tarpey, T., Su, Z., and Ogden, R. T. (2016). Generated effect modifiers in randomized clinical trials. *Biostatistics* **18**:105–118.

Plummer, F., Manea, L., Trepel, D., and McMillan, D. (2016). Screening for anxiety disorders with the gad-7 and gad-2: a systematic review and diagnostic meta-analysis. *General Hospital Psychiatry* **39**:24–31.

Salaffi, F., Leardini, G., Canesi, B., Mannoni, A., Fioravanti, A., Caporali, R., Lapadula, G., and Punzi, L. G. (2003). Onorthrosis and quality of life assessment (goqola). reliability and validity of the western ontario and mcmaster universities (womac) osteoarthritis index in italian patients with osteoarthritis of the knee. *Osteoarthritis Cartilage* **11**:551–560.

Schalet, B., Hays, R., Jensen, S., Beaumont, J., Fries, J., and Cella, D. (2016). Validity of promis physical function measured in diverse clinical samples. *Journal of Clinical Epidemiology* **64**:112–118.

Sullivan, M. J. L., Bishop, S., and Pivik, J. (1995). The pain catastrophizing scale: development and validation. *Psychological Assessments* **7**:432–524.

Treister, R., Suzan, E., Lawal, O. D., and Katz, N. P. (2019). Staircase-evoked pain may be more sensitive than traditional pain assessments in discriminating analgesic effects: A randomized, placebo-controlled trial of naproxen in patients with osteoarthritis of the knee. *Clinical Journal of Pain* **35**:50–55.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software* **45**:1–67.

Yu, L., Buysse, D. J., Germain, A., Moul, D. E., A, A. S., Dodds, N. E., Johnston, K. L., and Pilkonis, P. A. (2011). Development of short forms from the promis sleep disturbance and sleep-related impairment item banks. *Behavioral Sleep Medicine* **10**:6–24.

Zigmond, A. S. and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* **67**:361–370.

Zucker, D. R., Schmid, C. H., McIntosh, M. W., D'Agostino, R. B., Selker, H. P., and Lau, J. (1997). Combining single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of Clinical Epidemiology* **50**:401–410.