STATISTICAL ANALYSIS PLAN

Protocol Number: PCORI-OPD-1610-37006
NCT03454555
Integrated Health Services to Reduce Opioid Use While Managing Chronic Pain

SAP VERSION: Final, Version 1.0

SAP DATE: October 2, 2023

SPONSOR: RTI International

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

AUTHOR (S): Shawn Hirsch, MPH, Lead Statistician
Sonia Thomas, DrPH, Senior Statistician

SAP HISTORY

Version	Date	Update	Reviewed by
0.1	5/15/19	Drafted	--
0.2	3/12/21	1. Added objective of examining difference in intervention effectiveness by delivery mode 2. Added objective of determining number of completed intervention visits that would effect opioid change 3. Updated per revised protocol	--
0.3	3/24/21	1. Updated per revised protocol and approval of changes by PCORI 2. Edits by Lauren McCormack and Mark Edlund	Lauren McCormack and Mark Edlund
0.4	5/10/21	1. Accepted edits from Lauren McCormack and Mark Edlund	Sonia Thomas

		2. Added more details about databases and EHR data extraction timeline, cleaning, and transfer 3. Edits by Sonia Thomas	
0.5	10/5/22	1. Proposed change of the Analysis Population from an ITT to MITT 2. Added exploratory analyses evaluating predictors of positive outcome response within each intervention arm 3. Added Brief Pain Inventory (BPI) outcome measures to Section 9.2 4. Highlighted the planned sensitivity and supportive analyses 5. Other minor edits and updates	Sonia Thomas and Shawn Hirsch
0.6	2/13/23	1. Changed back from MITT to ITT 2. Changed time to opioid discontinuation analysis from using EHR data to T3 Survey data 3. Added a few proposed sensitivity analyses 4. Updated derivation of Health Literacy measure	Sonia Thomas and Shawn Hirsch
1.0	10/2/23	Finalized	Sonia Thomas and Shawn Hirsch

Contents

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADHD	Attention Deficit Hyperactivity Disorder
AE	adverse event
AHRQ	Agency for Healthcare Research and Quality
BPI	Brief Pain Inventory
BMI	Body Mass Index
CBT-CP	Cognitive Behavioral Therapy for Chronic Pain
CDM	common data model
CDRN	Clinical Data Research Networks
CI	confidence interval
CNCP	chronic noncancer pain
CONSORT	Consolidated Standards of Reporting Trials
COT	chronic opioid therapy
CRF	case report form
DSMB	Data and Safety Monitoring Board
ED	emergency department
EHR	electronic health record
GCC	guideline-concordant care
GLMM	Generalized Linear Mixed Model
ICD-10	International Classification of Diseases, 10th Revision
INSPIRE	Integrated Services for Pain: Interventions to Reduce pain Effectively
IOM	Institute of Medicine
IRB	Institutional Review Board
ITT	intent to treat
LTF	Lost to follow-up
MAR	missing at random
MCID	minimal clinically important differences
MED	morphine-equivalent dose
MI	motivational interviewing
MID	minimally important difference
MITT	Modified intent-to-treat
OUD	Opioid Use Disorder
PCORI	Patient-Centered Outcomes Research Institute
PCORNet	National Patient-Centered Clinical Research Network
PCP	primary care provider
PI	principal investigator
PROMIS	Patient-Reported Outcomes Measurement Information System
PROMIS-PF	Patient-Reported Outcomes Measurement Information System—Physical Functioning

Abbreviation	Definition
PROMIS-PI	Patient-Reported Outcomes Measurement Information System—Pain Interference
REDCap	Research Electronic Data Capture
RTI	RTI International
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SDM	Shared Decision Making
SUD	substance use disorder
UNC	University of North Carolina Health System
VUMC	Vanderbilt University Medical Center

1. BACKGROUND AND PROTOCOL HISTORY

Chronic noncancer pain (CNCP) is common, and the societal and clinical burden is high. A recent Institute of Medicine (IOM) report estimated that up to one-third of Americans suffer from CNCP. With current treatment approaches, 13% of headache patients and 18% of back pain patients remain unable to work full time because of pain. CNCP costs the United States up to $635 billion annually, more than the cost for cancer or diabetes or heart disease. The etiology of a patient's CNCP is often poorly understood, and no physiological test can objectively identify its presence or intensity. Mental health disorders and substance use disorders (SUDs) are often comorbid with CNCP. Complete resolution of CNCP symptoms is uncommon.

Primary care and specialty pain clinics have relied disproportionately on pharmacologic approaches, typically opioids, to treat pain. Dating back to the 1990s, authoritative bodies recommending the expansion of opioid use in patients with CNCP had advised that this occur concomitantly with a robust mental health component. However, this balanced approached to CNCP management did not occur. Prescribed opioid use increased threefold in the past 2 decades, although it now seems to be declining. Whereas 50% of all the prescriptions written by pain specialists are for opioids, only 6% of the prescriptions written by primary care physicians (PCPs) are for opioids. However, PCPs write 56% of all opioid prescriptions.19 For some patients, opioids may be the most effective, or only effective, analgesic for pain management. Most opioid use is chronic opioid therapy (COT) for CNCP. Patients initiating COT typically remain on this treatment for years20,21 and individuals on high-dose COT are the least likely to discontinue COT.

Despite the dramatic increase in use, opioids are a challenging treatment modality for CNCP. Their efficacy is incomplete, and some individuals may receive little or no pain relief from COT. A recent Agency for Healthcare Research and Quality (AHRQ) systematic review concluded that evidence for the effectiveness of COT is insufficient.22 No studies have evaluated the effects of discontinuing opioids on pain level, function, quality of life, or withdrawal symptoms. Despite the lack of evidence for the effectiveness of COT for chronic pain, it is commonly used in practice.

Further, COT has substantial potential for harm. Opioid overdose deaths are an epidemic in the United States, with more than 42,000 deaths in 2016.27-29 Men have almost twice the mortality rate from opioid use than women.30 In one study, most opioid overdose deaths were attributed to diversion, illustrating that opioid risks are not limited to individuals receiving opioid prescriptions.

Additionally, individuals on higher doses of COT are at greater risk for overdose death32 and for the development of misuse or abuse. The annual societal cost of opioid abuse was estimated at $55.7 billion in 2007. In younger persons, new marijuana use and new misuse of prescription opioids are roughly equivalent. Other substantial unintended consequences include opioid-induced decreases in quality of life.

The Centers for Disease Control and Prevention (CDC), the Federation of State Medical Boards, and individual state medical boards have issued guidelines to promote the safe and effective prescribing of opioid analgesics. These guidelines advise that opioids only be used when other pharmacological and nonpharmacological treatment modalities are not effective. Clinical guidelines for opioid prescribing emphasize patient selection for opioid initiation, monitoring, and reducing opioid misuse and abuse. However, the guidelines and scientific literature generally provide only limited discussion of opioid reduction or discontinuation if an opioid-based approach is not working for the patient. That is, guidelines address only the technical aspects of reducing medication dosages for COT (i.e., how quickly to decrease the dosage).36-40 Importantly, the existing literature does not discuss how to motivate individuals to discontinue or decrease their opioid dosage or to reduce the risk of opioid misuse; nor how to manage pain symptoms in a patient-centered fashion during and after opioid reduction or discontinuation. This evidence gap leads to a predictable clinical practice gap. In the absence of an evidence base to guide providers and patients on how to decrease opioids, reduction in opioid dosage or discontinuation of COT is infrequent.

This study focuses on the Mid-South area of the United States, where the opioid epidemic has had a disproportionate impact. The study includes opioid users from North Carolina and Tennessee. Individuals living in the Southeastern United States (especially in rural Appalachia), younger persons, and individuals with mental health disorders are at particularly high risk of serious opioid adverse effects. Appalachian areas have very high rates of opioid use. For example, using data from IMS Health, which collects a variety of healthcare information and is the largest vendor of U.S. physician prescribing data, McDonald and colleagues reported that Tennessee was 55% over and North Carolina was 18% over the national per-capita mean milligrams of opioids. Also using IMS Health data, the CDC reported that both Tennessee and North Carolina are among the 13 states with the highest rates of opioid prescribing, ranging from 96 to 143 opioid prescriptions per 100 residents. In 2014, the age-adjusted rates of drug overdose deaths were 19.5 in Tennessee and 13.8 in North Carolina.

Protocol History		
Date	**Ver.**	**Summary of Updates (only including those specific to analysis):**
09/06/2018	1.0	Original version submitted to UNC IRB
10/30/2018	2.0	• Added recruitment manual, voicemail script, informed consent form, intervention handouts, DSMB charters, and other study documents • Revised description of recruitment targets across sites • Added English language inclusion criterion • Clarified which exclusion criteria relate to past medical history • Changed extraction of EHR data from opt-out to opt-in consent if patient withdraws from study

		• Added satisfaction with care as an outcome measure and removed knowledge check • Removed interim analysis
11/29/2018	3.0	• Revised cancer-related exclusion criterion • Added Withdrawal Consent Addendum to seek permission to extract EHR data if patient withdraws from study • Clarified opioid management procedures for Arm 1
12/06/2018	4.0	• Changed opioid dose cutoff from 50 mg MED to 40 mg MED. This change will enhance enrollment and have no impact on study sample size or power. We also made administrative updates to power calculations in the protocol
02/21/2019	5.0	• Clarified inclusion criteria that patient must be on opioids for chronic non-cancer pain (CNCP) • Added that we will confirm in the EHR that the actual dosage prescribed is at least 40 mg daily MED. • Clarified that patients taking opioids for maintenance treatment of an opioid use disorder (OUD) will be excluded from the study • Clarified that lack of insurance is not an exclusion criterion • Added patient-centered communication measure to surveys • Added text messaging as a recruitment method • Changed length of MI session to 30 to 60 minutes
03/26/2019	6.0	• Removed text messaging as a recruitment method after receiving IRB guidance that text messaging is not allowable for recruitment purposes.
07/03/2019	7.0	• Added brochure as a recruitment method. • Added patient incentive receipt.
11/07/2019	8.0	• Updated the T2/T3 follow-up survey protocol • Changed the upper age limit for exclusion criteria from 75 to 85 years • Removed current CBT exclusion criterion • Clarified cancer exclusion criterion • Removed the exclusion criterion about visit scheduled within next 90 days.
12/13/2019	9.0	• Changed opioid dose cutoff from 40 mg MED to 20 mg MED. This change will enhance enrollment and have no impact on study sample size or power. We also made administrative updates to power calculations in the protocol. • Removed current CBT exclusion criterion as per PCORI request
4/16/2020	10.0	• Described remote intervention delivery measures • Clarified enrollment procedures for Arm 1
07/08/2020	11.0	• Described remote enrollment procedures
06/30/2021	12.0	• Decreased target sample size from 1,060 to 608. Updated power calculations. • Revised the study timeline and duration.

		• Updated recruitment and retention procedures (removed contacting alternate contacts; added option to contact participants during telehealth appointments; added option to send text message reminders to enrolled participant about follow-up surveys; added option to mail follow-up surveys to participants). • Updated study safety reporting procedures. • Updated background information about opioid use trends.
08/09/2021	13.0	• Revised to respond to IRB stipulations • Corrected minor errors
10/07/2021	14.0	• Corrected version numbers and dates in the list of attachments.
12/23/2021	15.0	Revised to include publicly available data as a data source. •
3/30/2022	16.0	• Revised to describe procedures related to requesting death certificates: maintaining the confidentiality of the data, procedures for ensuring the privacy of subjects, and how these data will be used.

2. PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analysis to be performed to assess whether CNCP patients who receive a MI+CBT-CP intervention have reduced daily morphine equivalent dose (MED), improved physical functioning, and lower pain interference compared to CNCP patients who receive SDM intervention. No formal interim analyses will be conducted. As such, all analyses described will be performed at the end of the study excluding any safety summaries provided for the DSMB which are described in Section 7.4.

3. STUDY OBJECTIVES AND OUTCOMES

3.1. Study Objectives

3.1.1 Primary Objective

1. Test whether CNCP patients who receive guideline-concordant pharmacotherapy with a MI+CBT-CP intervention have greater **opioid dose reduction** relative to their counterparts who receive guideline-concordant pharmacotherapy integrated with SDM.

3.1.2 Secondary Objectives

1. Test whether CNCP patients who receive guideline-concordant pharmacotherapy with an MI+CBT-CP intervention have improved **physical functioning** relative to their counterparts who receive guideline-concordant pharmacotherapy integrated with SDM.

2. Test whether CNCP patients who receive guideline-concordant pharmacotherapy with an MI+CBT-CP intervention have lower **pain interference** relative to their counterparts who receive guideline-concordant pharmacotherapy integrated with SDM.

3.1.3 Additional Objectives Not Specifically Listed in the Protocol Objectives section

1. Comparison of intervention groups for other self-reported outcomes of anxiety, pain intensity, depression, Brief Pain Inventory (BPI) pain intensity subscale and pain interference subscale, intent to taper, and discontinuation of opioid use.

2. While multiple subgroup analyses have an inherent risk of inflating Type I error, this trial offers a unique opportunity to generate hypotheses about the profile of patients most likely to benefit from the intervention. Consequently, planned *secondary* analyses will assess potential differential treatment effects (treatment by subgroup interaction) for two subgroups: 1) defined by participants with comorbid mental health conditions and 2) sex.

3. The study will also include descriptive *exploratory* analyses to explore the heterogeneity within each intervention arm separately on opioid reduction (predictors of response) according to covariates of interest including age, baseline pain score, comorbidities (including physical comorbidities and mental health disorders, and past or current alcohol or other substance abuse and related disorders), those taking other medications, patient health literacy level, BMI, and opioid dose used at baseline in 3 categories consistent with CDC guidelines: low (20-49 MED), moderate (50-89 MED), high (90 or more MED); we will also have a category for very low (1-19 Med); and intervention delivery mode (in-person vs telehealth). Testing for heterogeneity of treatment effects for these subgroups will be considered exploratory as opposed to confirmatory, and the results will be interpreted with appropriate caution

4. Determine if there is a threshold for the number of completed SDM visits or CBT sessions that results in a clinically significant decrease in opioid use from baseline, defined as a decrease of 10 MED daily dose.

3.2 Outcomes

3.2.1 Primary Outcome

The primary outcome is **change from baseline in average daily opioid dose in MED**. The daily MED for each prescription will be calculated by multiplying the quantity of each prescription by the strength of the prescription by the MED conversion factor and dividing it by the total days supply. The primary outcome will be examined at baseline and months **3, 6, 9, 12, 15 and 18** using opioid prescription information recorded in the EHR from 90 days prior to randomization until 18 months after randomization, as provided by the PCORnet common data model (CDM) data warehouse at each site. For each timepoint, the primary outcome will be derived including all opioid medications prescribed 90-days prior to the desired timepoint. The primary timepoint is 12 months; key secondary timepoints at 6 and 18 months.
Derivation of this outcome is provided in section 9.2.

3.2.2 Secondary Outcomes

1. Change in self-reported physical functioning, as collected via the **PROMIS-PF 8-item Short Form**, from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.

2. Change in self-reported pain interference, as collected via the **PROMIS-PI 8-item Short Form**, from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.

3.2.3 Other Self-Reported Outcomes

1. Change in self-reported pain intensity, as collected via the **PROMIS-PI 3-item Short Form**, from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.
2. Change in self-reported anxiety, as collected via the **PROMIS-Anxiety 4-item Short Form**, from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.
3. Change in self-reported depression, as collected via the **PROMIS-Depression 4-item Short Form**, from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.
4. Change in self-reported pain intensity and pain interference as measured by **BPI pain intensity subscale** and **BPI pain interference subscale**. from pre-intervention (baseline) to post-intervention (months 6 and 12). The primary timepoint is 12 months.
5. Incidence of **intent to taper** opioid medication during 12 months, defined by self-report on the 6 and 12 month surveys.
6. D**iscontinuation of opioid use at 12-months (exploratory outcome)**, defined as a 'No' response on the T3 survey question asking if the participant is currently taking an opioid medicine.

4. STUDY METHODS

4.1 Overall Study Design and Plan

We will conduct a multisite, randomized pragmatic trial to examine the comparative effectiveness of 2 interventions: a guideline-concordant opioid pharmacotherapy approach coupled with SDM (Arm 1) compared with a guideline-concordant opioid pharmacotherapy approach coupled with MI and CBT-CP (Arm 2). Participants will complete baseline, 6-month, and 12-month follow-up surveys, measuring self-report physical functioning and pain interference, along with other self-report measures. Additionally, participant's electronic health record (EHR) data will be extracted from the Common Data Model (CDM) from baseline to 18 months, which will be used to calculate average daily MED. The study design is presented in Figure 1, the study timeline is presented in Figure 2, and the data collected within each of the three study databases is presented in Figure 3. Section 4.4 below further describes the EHR data stored within the CDM, how the data is processed and transferred to RTI, and the rough time estimate for extracting and cleaning the EHR data.

Figure 1. Study Design

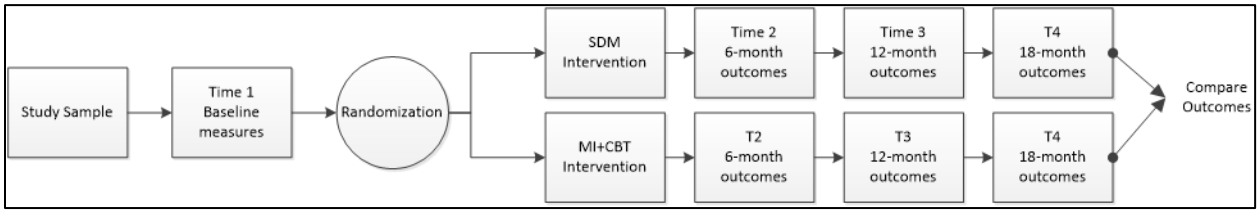


Figure 2. Study Timeline

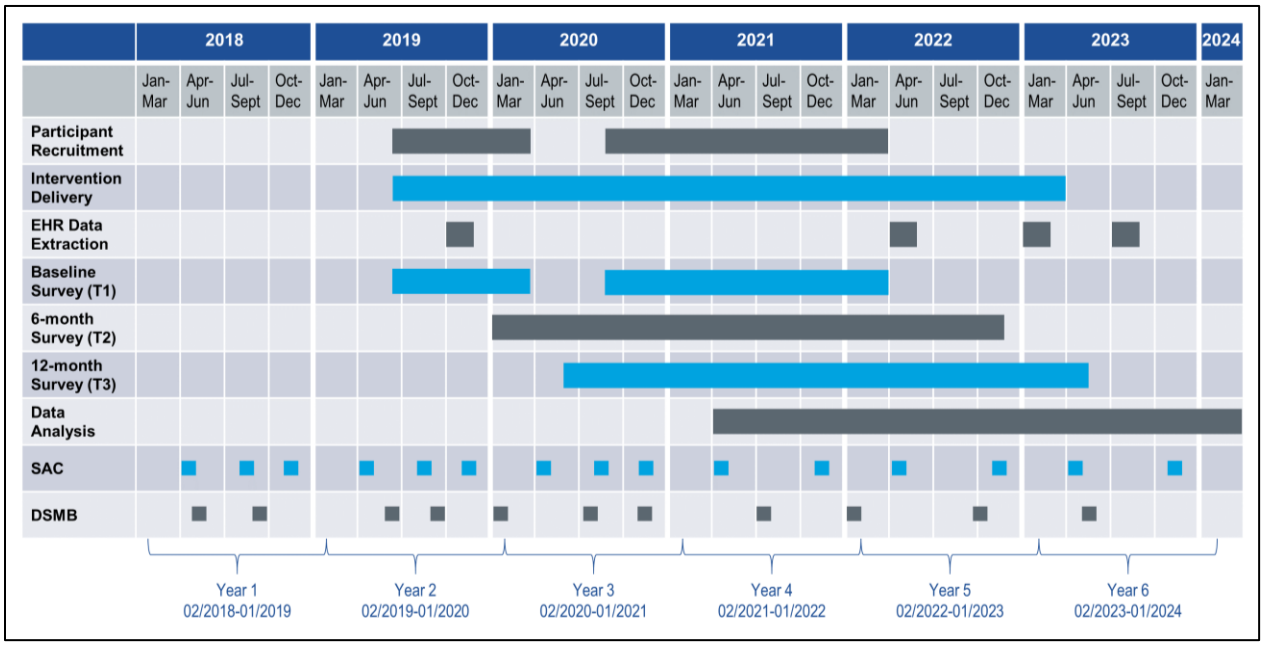
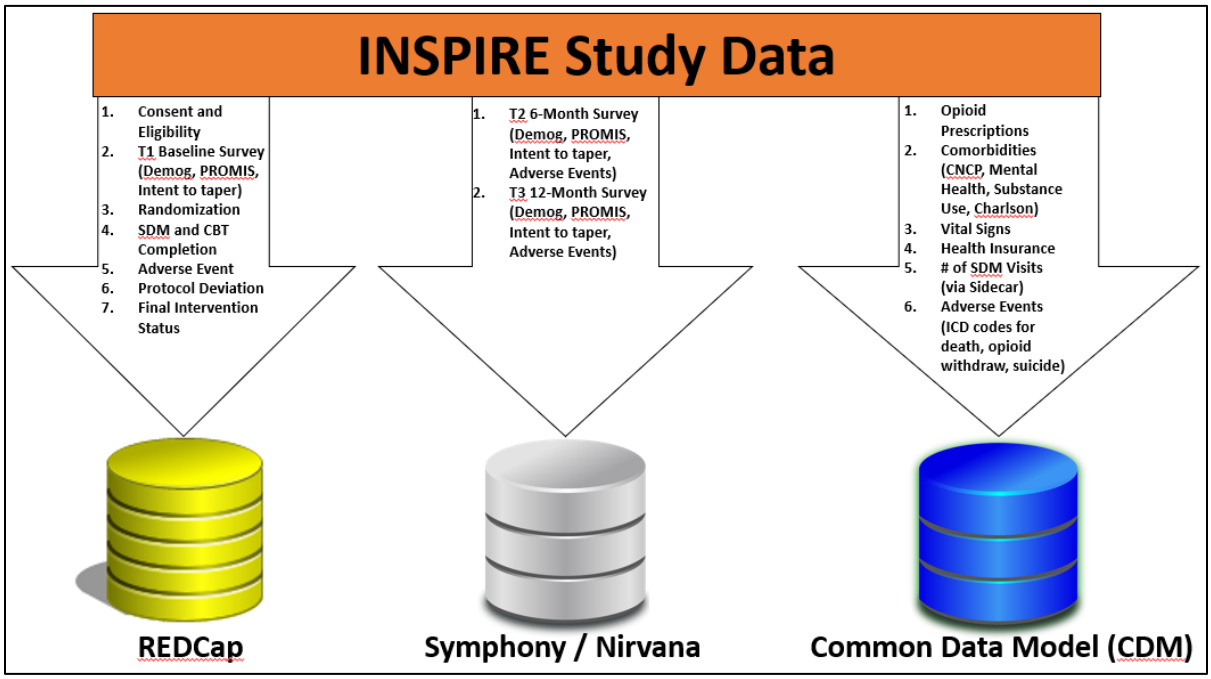


Figure 3. Data within Each of the 3 Study Databases



4.2 Study Population

The trial will be conducted in primary care and specialty pain clinics at 3 university health systems: UNC, Duke, and VUMC.

4.2.1. Participant Characteristics

The study population is defined by the following eligibility criteria.

Inclusion Criteria:

1. Aged 18 to 85 years
2. History of CNCP
3. Receiving moderate-dose COT for CNCP as evidenced by current or most recent prescription of an average daily MED of 20 mg or greater.
4. Receiving care at a participating clinic from a participating provider, as evidenced by at least 1 in-person visit within the past 12 months

Exclusion Criteria:

1. Opioid use is for pain directly related to an active cancer diagnosis
2. Opioid use is for maintenance treatment of an opioid use disorder (OUD)
3. Currently receiving CBT
4. Non-English speaking
5. Suicide attempt within the past 3 years or active suicidal ideation.
6. Other reason at the discretion of the investigator.

4.3 Study Arm Assignment and Randomization

Total anticipated enrollment is 608, with 304 participants in each arm. We expect to enroll about 203 participants per institution. This study will use real-time 1:1 ratio randomization to limit participant loss prior to treatment. Eligible participants will be randomized using a stratified, permuted-block design with random block sizes of 4 and 6 because this constrained randomization approach ensures balance between treatment groups within each of the 3 institutions (our only stratification factor) at the completion of each block. Consequently, throughout the trial, the intervention arms are expected to have approximately equal sample sizes both within an institution and across the study. We will not be able to blind participants or providers to intervention arm assignment because of the nature of the intervention.

4.4 Electronic Health Record (HER) Data in the CDM

4.4.1 Data Stored in the CDM

The EHR data extracted from the CDM contains the study's primary outcome (average daily opioid dose in MED), which is derived from each participant's prescription data. This data source also contains important subgroup variables including vital signs and CNCP, and substance use and mental health diagnoses. The study will also be extracting potential adverse events based on ICD-10 codes and insurance provider information. Each site will also perform a separate 'sidecar' request to pull out the total number of SDM visits per participant (based on attendance at a healthcare visit with a trained SDM provider, clinic notes and smartphrases), including the delivery mode of the visit (in-person vs. telehealth). A sidecar request is necessary for this type of data since this data is not included in the CDM.

4.4.2 Data Extraction and Transfer to RTI

To extract EHR data for INSPIRE participants, each participant's INSPIRE Subject ID needs to be mapped to their Medical Record Number (MRN). To accomplish this, the RTI statistician sends an Excel Trial Table to each site coordinator that contains each participant's INSPIRE Subject ID and the date in which they either completed or withdrew from the study. The site coordinator then populates each participant's MRN in the Trial Table (note: RTI is not provided MRNs) and sends it to the site programmer. Prior to this, the site programmer is provided an INSPIRE Query Package from RTI that contains SAS programs designed to extract the appropriate EHR data from their CDM. The site programmer will run the SAS program, which calls in the Excel Trial Table and extracts all pre-specific EHR data for each INSPIRE participant. The resulting SAS datasets are uploaded to the RTI SFTP server by the site programmer after they have cross-walked the output with the expected data dictionary (for a brief QAQC check).

4.4.3 Timeline and Data Cleaning for EHR Data Extraction

Each study site's CDM is refreshed with new EHR data 4 times a year: January, April, July, and October. Additionally, there is up to a 3-month lag time when data from a participant's clinic visit is entered into some institutions' EHR research database (which feeds into the CDM). For each CDM data transfer, the initial transfer will occur immediately after the scheduled CDM data refresh. After RTI receives the initial transfer, data managers and statisticians will review the EHR data for quality checks including, but not limited to: receipt of all expected data tables per participant, missing prescription data fields, participants with an average baseline MED of < 20, out of range fields, and potential repeat/duplicate data. The RTI data manager will query the site regarding identified data quality issues. Based on the site's response regarding queries, a final cleaned extracted EHR dataset is expected approximately 2-months after the initial data extraction. Figure 4 below depicts the timing and processing for each planned EHR data extraction.

Figure 4. EHR Timeline and Processes

	2019			2020	2021	2022					2023								2024
	Oct	Nov	Dec	Jan-Dec	Jan-Dec	Jan-Mar	Apr	May	Jun	Jul-Dec	Jan	Feb	Mar	Apr-Jun	Jul	Aug	Sep	Oct-Dec	Jan-Mar
I. EHR Data Extraction [1,2]																			
a. Initial Extraction																			
b. Data Review and Clean																			
c. Final Extraction (clean data)																			
II. Data Analysis																			

[1] There is roughly a 3-month lag between a participant's clinic visit and when that EHR data will be available within the CDM
[2] The CDM is updated 4 times a year: January, April, July, and October. Each CDM data extraction above is to be scheduled *after* the CDM database has been updated.

4.5 General Masking Procedures

Masking procedures are not applicable for this study since it is an unmasked behavioral intervention trial. However, interim reporting of summary statistics by intervention arm (such as reports to the DSMB), will be blinded and only accessible to the study statistician. The statistical lead, project PI, and all project staff and site PIs and staff remain blinded to cumulative data.

4.6 Database Lock

As noted above, 3 different study databases will provide data. The last participant is expected to be enrolled in March 2022 and the last intervention visit may go until March 2023. The last T2 (6-month) survey is expected in September 2022 and the last T3 (12-month survey) is expected in March 2023. The T2 and T3 survey database will be locked approximately one month after the last completion of the T3 survey. The REDCap database lock will occur after the last participant's final intervention status CRF has been completed (i.e., completed the 12-month intervention phase) and all data queries have been resolved by the study sites. As mentioned above in section 4.4.3, the EHR data extraction takes places 3 times across the study. *Although the final EHR data extraction is planned to take place in July or August 2023, or 18-months after the last participant was enrolled into the study, it will likely not contain 18-months of EHR data for the last participant since there is up to a 3-month lag between clinic visit and data inclusion in the CDM. However, all participants should contain 12-months of EHR data in the final extraction (the study's primary timepoint).* The final EHR data extraction will be considered final and locked after the sites have addressed any potential RTI queries and transferred a clean dataset (expected to occur ~2-months after the initial data transfer). The locked datasets from each of the 3 databases will be used to derive the final analysis dataset which will be used to perform the analyses described within this SAP. Analysis data specifications are further described in section 9.2.

The study is designed as intention-to-treat (ITT). If a participant discontinues further intervention or participation in the study, every attempt will be made to continue to perform the required study-related follow-up surveys and EHR data extraction. However, surveys and EHR data will not continue to be collected from those who are determined ineligible, withdrew from all study activities, investigator withdrew and other (depending on the reason).

4.7 Study Flow Chart of Assessments and Evaluations

.Measurement	Screening (Month -12 to 0)	Baseline Survey (Month 0)	Baseline CDM Extraction (Month -3 to 0)	Intervention visits (Months 1-12)	CDM Extraction (Months 6, 12, 18)	T2 survey (Month 06)	T3 survey (Month 12)
Informed consent	●						
Demographics	●					●	●
Health literacy level		●					
Randomization		●					
Average daily MED			●		●		
Medical encounters			●		●		
Charlson Comorbidity Index (ICD-9/10)			●		●		
Chronic noncancer pain conditions (ICD-9/10)			●		●		
Mental health diagnoses (ICD-9/10)			●		●		
Alcohol/substance use diagnoses (ICD-9/10)			●		●		
Body mass index (based on height and weight)			●		●		
Health insurance payer			●		●		
Opioid withdrawal medication prescriptions			●		●		
Administer study intervention				●			
EHR template note				●			
Adverse Event Case Report form (as indicated)				●			
Brief Pain Inventory		●				●	●
PROMIS self-report scales: (Pain Interference, Physical functioning, Pain Intensity, Anxiety, Depression)		●				●	●
Current psychotherapy		●				●	●
Self-reported opioid use and intent to taper		●				●	●
Use of intervention content; knowledge check						●	●
Hospitalization or ED visits related to opioid use (overdose or withdrawal) or suicide attempt			●		●	●	●
Death					●		
Adverse Event review and evaluation	●	●	●	●	●	●	●
Fidelity review and evaluation				●		●	●

5. ANALYSIS POPULATIONS

5.1 Intention-to-Treat (ITT)

The primary and secondary analyses will be based on the intention-to-treat (ITT) principle, with data from all participants analyzed according to the arm to which they were randomized irrespective of the amount of intervention received. Participants who are randomized but do not complete the study will be used in all analyses for which data are available. This approach ignores nonadherence, protocol deviations, withdrawal and lost-to-follow-ups. However, in the event a participant is accidentally randomized twice, all data for this participant will be included in the CONSORT diagram; however, the participant's first randomization assignment will only be included in the ITT analyses. The participant's second randomization assignment will be removed from the analysis dataset and subsequent data analyses. Analysis of the ITT population avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers, accepting that protocol deviations occur in actual clinical practice (Heritier, Gebski, & Keech, 2003).

5.2 Per-Protocol Population

As in most pragmatic clinical trials, some participants may not adhere to the intervention they were randomized to receive, reducing fidelity to the intervention as designed and potentially changing the effectiveness of the intervention. The most likely form of nonadherence will be absence from group therapy sessions in the MI+CBT-CP arm or with patients in the SDM arm ceasing interaction with the assigned provider. The potential impact of this nonadherence is underestimating the magnitude of the true treatment effect. To assess this, in addition to the ITT analyses, we will also conduct secondary analyses using a per protocol population consisting of participants that met all inclusion/exclusion criteria, and received a substantial portion of the randomized intervention, defined as at least 4 completed SDM visits and at least 4 completed MI or CBT sessions (out of a total of 1 MI and 8 CBT). Note that the inclusion criterion for baseline MED is based on the most recent prescription, while the baseline value for MED in the analysis is the average MED for the past 90 days. Having a baseline value < 20 is not a protocol violation when the sites indicated that inclusion criterion was met based on the most recent prescription when the patient was assessed. A sensitivity analysis on the per-protocol population will also be conducted by defining this population as those who met all inclusion/exclusion criteria and completed at least 3 SDM visits and at least 3 MI or CBT sessions. This sensitivity analyses are included as many SDM participants may only complete 3 visits with their provider throughout the 12-month intervention phase in the event their opioid prescriptions can be renewed over a 3-month period without an actual clinic visit.

The per-protocol analysis provides a better estimate of the true efficacy of an intervention (i.e., among those who completed the treatment as planned) (Ranganathan, Pramesh, & Aggarwal, 2016). No other analysis populations are defined for this protocol.

6. SAMPLE SIZE DETERMINATION

Sample size estimates were generated to provide robust power to detect minimal clinically important differences (MCID) in reduction of opioid use between the 2 study arms.

A total sample size of 506 participants (253 per study arm) will provide 80% power to detect an effect size of a difference of 10 MED between the SDM and MI+CBT intervention arm using an analysis of covariance model with α=0.05 and assuming a baseline model-adjusted standard deviation for change from baseline of 40 MED. The sample size was inflated by 20% (608 total participants; 304 per study arm) to account for potential variance inflation associated with CBT group delivery (based on a group

size of 8 and an interclass correlation coefficient of 0.01) and a potential loss of follow-up opioid prescription data on as many as 12% of participants.

For the PROMIS secondary outcome measures, the minimally important difference (MID) for the physical function scale is 2 units, whereas the MID for the pain interference scale is 3.5 units. For both instruments, we assume the standard deviation for change from baseline is 10 (see section 9.2 for a more thorough description of these secondary measures). With the planned total of 608 randomized participants and assuming 20% of those have missing PROMIS scale responses, the power for the PROMIS Pain Interference score is 96%, while the power for the PROMIS Physical Functioning scale is 57% (Table 6.1).

Note that the enrollment was slower than planned, and due to timeline constraints was stopped prior to reaching 608 randomized participants. Final study enrollment was 526 randomized participants. As of June 2023 (DSMB report), study discontinuation was at 20%;, but survey completion rate was tracking around 60% (40% missing, higher than the planned levels). Approximate post-hoc power based on the final sample size is provided in Table 6.1.

Table 6.1. Power for Primary and Key Secondary Outcomes

Total sample size	Primary aim: Power for opioid MED[1]	Secondary aim: power for PROMIS Pain Interference[2]	Secondary aim: Power for PROMIS Physical Functioning[2]
Planned 608	80%	96%	57%
Approximate post hoc power based on randomized (n=526)	75% based on protocol-assumed level of missing opioid prescription data	85% Based on number of completed surveys	40% Based on number of completed surveys

7. STATISTICAL / ANALYTICAL ISSUES

7.1 General Rules

All statistical computations will be performed and data summaries will be created using SAS 9.4 or higher. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of participants (by intervention where relevant); continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by median and range or frequency and percentage. Model-based analyses described in section 9 will also be used to obtain point estimates and associated confidence intervals for the primary and secondary outcomes as well as p-values for comparisons of data between intervention arms. P-values presented will be based on two-sided tests unless otherwise specified. For continuous outcomes, checks of normality will be performed and if substantially violated, transformations or non-parametric tests will be employed.

7.2 Subgroups and Heterogeneity of Intervention Effects

Treatment by subgroup interactions and exploration of baseline predictors of response are described as additional objectives. Statistical methods are described in 9.5.4 and 9.5.5.

7.3 Handling of Dropouts and Missing Data

Analysis of the primary and secondary repeated measurement endpoints will be based on a repeated measures mixed model using data from all time points to account for missing data and thereby maximize information used for the analyses. The correlation of measurements between visits for a participant at earlier time points (including those who discontinue early) contribute to the model by reducing the standard error of the outcome over time, allowing us to have a tighter confidence interval at the primary timepoint (i.e., at 12-months). These models treat missing data as ignorable missing, assuming any missing data are missing at random (MAR); see Section 9.3 for more detail.

In the event of partially missing medication information for a prescription, we will calculate the daily MED and days supply based on other available prescription data and a series of assumptions vetted for reasonableness by prescribing study clinicians and by evaluation of consistency with other prescription data. All assumptions for calculating daily MED and days supply are described further in Section 9.2.

7.4 Interim Analyses and Data Monitoring

The INSPIRE Leadership has established a Data Safety Monitoring Board (DSMB) to oversee this study. The DSMB will meet 1-2 times a year as specified in the DSMB charter to review the study, although may be convened between planned meetings to discuss study issues related to adverse events/safety. This protocol was approved by the DSMB prior to initiation of recruitment. The DSMB will monitor study progress and can recommend that the trial be stopped for safety or futility. A description of the report to be provided to the DSMB to review the study are included in Appendix A.

7.5 Multicenter Studies

This is a 3-center study including Duke University, University of North Carolina at Chapel Hill (UNC), and Vanderbilt University Medical Center (VUMC). Although we expect the number of enrolled participants to be variable across each enrolling clinic, we expect the total number of enrolled participants to be roughly equal across the three Universities. Since randomization was stratified based on study site, and since this is a pragmatic trial with potential site differences in intervention implementation, site will be included as a term in most model-based analyses.

7.6 Multiple Comparisons and Multiplicity

The comparison of the intervention arms for the primary outcome at 12-months is the primary formal hypothesis test (0.05 level of significance, two-tailed). Statistical significance of the 2 key secondary outcomes (PROMIS physical functioning and PROMIS pain interference at 12 months) will be assessed with adjustment for multiple comparisons using the Hochberg modification to the Bonferroni adjustment (discussed further in section 9.5.1). Primary and secondary hypothesis tests resulting in nonsignificant p-values will be interpreted as inconclusive. All other statistical comparisons between intervention arm will be considered descriptive in nature with no adjustment for multiple comparisons, and all confidence intervals will be generated using 95% bounds.

8. STUDY PARTICIPANT CHARACTERIZATION

8.1 Participant Disposition

Participant eligibility status and disposition will be summarized and described using a standard CONSORT diagram. The diagram will include the number of participants randomized, the

number completing any intervention visits, the number that did not actively withdraw from the study, survey completion rates, and participants with EHR data extraction. Due to the pragmatic nature of the study, many participants may have been marked as completing the intervention phase of the study; however, they did not actually complete any intervention visits and could not be contacted by research coordinators. However, these participants were still included in the study as they received survey invitations and EHR extractions. For CONSORT reporting purposes under the study follow-up section, we will report the number and percentage of those that did NOT actively withdraw or discontinue early and the summary statistics for those that did withdraw or discontinue early. Among those that definitively ended the study early, reasons for study discontinuation or withdrawal will be listed.

In the event a participant is accidentally randomized twice, the participant's first randomization assignment will only be displayed and included in the ITT analyses. The participant's second randomization assignment will be removed from the analysis dataset and subsequent data analyses.

8.2 Protocol Deviations

Protocol deviations will be summarized by total number of deviations and by type of deviation.

8.3 Study Interventions and Adherence

Participants will be randomized (based on a 1:1 ratio) to receive either a SDM or MI+CBT intervention. SDM seeks to explore and compare treatment options, assess a patient's values and preferences, and reach a shared decision between a patient and provider regarding current pain management options. CBT is an empirically based behavioral pain management therapy intervention which has been found effective for chronic pain. MI will be used to enhance motivation for active participation in the CBT sessions. This arm includes 1 individual MI session and 8 CBT-CP weekly group therapy sessions. Throughout the study, participants in each of the study arms will receive guideline-concordant care (GCC), as based on CDC guidelines. The guidelines give guidance about medication selection, dose and duration, when and how to assess progress, and discontinue medication if needed.

The study will incorporate three separate Fidelity Case Report Forms (CRFs) to assess adherence for each study intervention. More information on intervention fidelity are included in a separate Fidelity Assessment Plan. The number of completed SDM visits and CBT group sessions will be reported. CBT session attendance is recorded by the study site in the REDCap database. The number of SDM visits are obtained by a sidecar data extraction from the EHR as described in Section 4.4.1.

8.4 Demographic and Baseline Characteristics

All data collected at baseline (either from the T1 survey or from the CDM) will be summarized descriptively for the ITT population. Variables of interest include age, sex, race, ethnicity, education, marriage status, employment status, health insurance coverage, prescribed average daily opioid dose (MED), and PROMIS pain interference and physical functioning scores. No statistical tests will be run to determine differences in demographic characteristics between the intervention arms since participants were randomized into the different interventions. NOTE: many of the demographic data collected on the T1 baseline survey are duplicated within the EHR data. If any discrepancies in the demographic data are identified between the two data sources and cannot be resolved, data collected directly from the participant on the T1 survey will be considered the gold standard and included in the baseline analyses.

9. EFFICACY AND SAFETY ANALYSES

9.1 Overview of Analysis Methods

Statistical analyses for this study have been designed to compare (1) the effectiveness of the 2 interventions (MI+CBT-CP and SDM) in reducing opioid dosage among patients with CNCP, and (2) the effects of the 2 interventions on PROMIS pain interference and physical function. The analyses will use data collected at baseline through 18 months to evaluate these effects, and will formally test the primary hypothesis that the change in opioid dose (absolute change in average daily MED) differs between the 2 intervention arms at 12 months and will describe differences in absolute change in opioid dose at 6 and 18 months and percent change in dose at each timepoint. The analysis also will formally test intervention arm differences for 2 secondary outcomes: change in pain interference and physical function at 12 months. Change in outcomes at 6 months will also be described. Other self-reported outcome measures will also be compared between interventions. Primary and secondary analyses will use model-based approaches that take advantage of the longitudinal structure of the outcome data to address missing data caused by patient loss to follow-up or nonresponse and take into consideration correlated data collected across time. Additional analytic details are in the sections below.

9.2 Variable Definitions

The table below provides a description of the study endpoints and subgroup categories, including data type and derivation. More information regarding these variables is provided in the analysis dataset specification file.

Variable	Type	Definition
Primary Outcome		
Change in opioid dose (based on average daily MED) from baseline to 12 months post-randomization	Continuous	The primary outcome will be derived from EHR data from the Mid-South CDRN data warehouse. Total morphine equivalents for each prescription will be calculated by multiplying the quantity of each prescription by the strength of the prescription (milligrams of opioid per unit dispensed). The quantity-strength product is then multiplied by conversion factors to estimate the milligrams of morphine equivalent to the opioids dispensed in the prescription. The total average dose in morphine equivalents per day supplied is calculated by summing the morphine equivalents for each prescription filled during a given period and dividing by the number of days supplied. The below algorithms further describe derivation of the primary outcome measure: **One Prescription's MED Daily (MEDD) = (Strength per unit) x (Quantity dispensed) x (MME Conversion factor) / (Days Supply)** **Avg. MEDD per Timeframe = (Multiply each RX's MEDD by Days Supply) then (Sum all RX's Total MEDD for those prescriptions within timeframe) / Total number of days in timeframe** Opioid dose will be calculated as the prescribed milligrams of daily MED averaged over the 90 days prior to randomization and averaged over 90-days for the time periods of 3, 6-, 9-, 12-, 15- and 18-months post randomization. We expect no missing prescriptions, yet some participants might not have complete data because (a) they discontinued and removed consent to use data after that date, (b) they were lost to follow-up and may have moved out of the clinic's care, (c) the last CDM transfer did not contain their data up to 18 months past baseline or (d) they discontinued taking opioids. Cases of (d) will be identified by the response to the T2 and T3 survey item G20 "Are you currently taking an opioid medicine now?". If the response indicates "no" and opioid prescriptions last dose date ends before month 18, then an MED of zero will be assumed for all days after the last dose date to Month 18 (or date participant removed informed consent, if applicable). Otherwise, if there is not a "no" response

Variable	Type	Definition
		to this item at the applicable T2 or T3 survey, or the survey was not completed, then the data after the last dose date ends will be considered missing in the calculation of the average MED across the 3-month time intervals. NOTE: Section 5.3.1 in the protocol states the time periods of 4-, 6-, 8-, 10-, 12-, 14-, 16-, and 18- would be used; however, that was an inadvertent typo and it is correctly stated in protocol section 5.2.1. For each of the post-randomization periods, change in daily opioid dose will be computed as the difference between the dose calculated during that period and the dose from the baseline period. Further documentation regarding the medications included in the derivation of average daily opioid dose (MED) can be found in the supplemental file titled "INSPIRE_SAP_Suppl_Opioid_Meds"
Average percentage change in MED (based on average daily MED) from baseline period	Continuous	Percent change is calculated as the change from baseline divided by the baseline value. For example, if average baseline is 30 and average post baseline is 20, percent change is (20-30)/30 = decrease of 33%. Alternatively, if average baseline is 30 and average post baseline is 40, average percent change is (40-30)/30 = increase of 33%.
Average relative amount of MED from baseline (based on average daily MED)	Continuous	**If warranted based on evaluation of the data prior to database lock**, we may evaluate average relative change rather than actual change – defined as the value in the post baseline timepoint divided by the baseline value. For example, if average baseline is 30 and average post baseline is 20, average relative change is 20/30 = 2/3rds of the average baseline MED. If determined not warranted, this variable will not be calculated.
Indicator for change from BL of at least 10 MED	dichotomous	Calculated at 3, 6, 9, 12, 15, 18
Secondary Outcomes		

Variable	Type	Definition
Change in PROMIS pain interference from baseline to 6- and 12-months post-randomization	Continuous	Pain interference is a measure of the extent to which pain interferes with patient physical, mental, and social activities. It is an outcome that has been identified by patients in the target population as a critical patient-centered outcome for evaluating the potential effects on an opioid-reduction strategy. Pain interference will be measured using a standardized instrument, the 8-item PROMIS-PI scale. The PROMIS-PI scale is a T-score based on PROMIS normative data; a score of 50 represents the average score for the normative population and 10 is the standard deviation (SD) of that population. A higher T-score indicates higher pain interference and worse health. **The T score is calculated via online scoring provided by PROMIS (https://www.assessmentcenter.net/ac_scoringservice/templates/UserManual.pdf)** The PROMIS Pain Interference scoring manual is found at the below link: https://staging.healthmeasures.net/images/PROMIS/manuals/PROMIS_Pain_Interference_Scoring_Manual.pdf
Change in PROMIS physical functioning from baseline to 6- and 12-months post-randomization	Continuous	Physical functioning measures one's upper extremities (dexterity), lower extremities (walking and mobility), central regions (back and neck), and instrumental activities of daily living. We will measure physical functioning using the 8-item PROMIS-PF instrument. The PROMIS-PF scale is a T-score based on PROMIS normative data; a score of 50 represents the average score for the normative population and 10 is the standard deviation (SD) of that population. A higher T-score indicates higher physical functioning and better health. **The T score is calculated via online scoring provided by PROMIS (https://www.assessmentcenter.net/ac_scoringservice/templates/UserManual.pdf)** The PROMIS Physical functioning scoring manual is found at: https://staging.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Physical_Function_Scoring_Manual_26May2022.pdf
Other Outcomes		
Change in PROMIS pain intensity from baseline to 6-	Continuous	We will measure pain intensity using the 3-item PROMIS instrument. The PROMIS scale is a T-score based on PROMIS normative data; a score of 50 represents the average score for the

Variable	Type	Definition
and 12-months post-randomization.		normative population and 10 is the standard deviation (SD) of that population. A higher T-score indicates higher pain intensity and worse health. **The T score is calculated via online scoring provided by PROMIS (https://www.assessmentcenter.net/ac_scoringservice/templates/UserManual.pdf)** The PROMIS pain intensity scoring manual is found at: https://staging.healthmeasures.net/images/PROMIS/manuals/PROMIS_Pain_Intensity_Scoring_Manual.pdf
Change in PROMIS anxiety from baseline to 6- and 12-months post-randomization.	Continuous	Anxiety measures self-reported fear, anxiety, hyperarousal, and somatic symptoms related to arousal. Anxiety is best differentiated by symptoms that reflect autonomic arousal and experience of threat. We will measure anxiety using the 4-item PROMIS instrument. The PROMIS scale is a T-score based on PROMIS normative data such that a score of 50 represents the average score for the normative population and 10 is the standard deviation (SD) of that population. A higher T-score indicates higher anxiety and worse health. **The T score is calculated via online scoring provided by PROMIS (https://www.assessmentcenter.net/ac_scoringservice/templates/UserManual.pdf)** The PROMIS anxiety scoring manual is found at: https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Anxiety_Scoring_Manual_03June2022.pdf
Change in PROMIS depression from baseline to 6- and 12-months post-randomization.	Continuous	Depression measures self-reported negative mood, views of self, social cognition, and decreased positive affect and engagement. We will measure depression using the 4-item PROMIS instrument. The PROMIS scale is a T-score based on PROMIS normative data such that a score of 50 represents the average score for the normative population and 10 is the standard deviation (SD) of that population. A higher T-score indicates higher depression and worse health. **The T score is calculated via online scoring provided by PROMIS (https://www.assessmentcenter.net/ac_scoringservice/templates/UserManual.pdf)**

Variable	Type	Definition
		The PROMIS depression scoring manual is found at: https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Depression_Scoring_Manual_03June2022.pdf
Change in Brief Pain Inventory (BPI) pain severity summary measure from baseline to 6- and 12-months post-randomization.	Continuous	The BPI assess pain at its "worst", "least", "average", and "now". Although clinical trials often use pain at "worst" and "average" to represent single pain severity, the BPI's developers recommend that all four severity items be used, because the models for validation of the BPI included all four items. A composite pain severity score is derived by taking a mean severity score of the four severity questions, if all four items are answered. The BPI severity score is also used to define the participant's baseline pain severity. The BPI User Guide is found at: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf
BPI pain severity categories	Categorical	BPI pain severity sub score categories at baseline are defined as (Boonstra 2014): Mild: < 3.5 Moderate: 3.5 – 7.4 Severe: 7.5+
Change in Brief Pain Inventory (BPI) pain interference summary measure from baseline to 6- and 12-months post-randomization.	Continuous	The BPI measures how much pain has interfered with seven daily activities: general activity, walking, work, mood, enjoyment of life, relation with others, and sleep. BPI pain interference is scored as the mean of the seven pain interference items, This mean can be derived if more than 50% (or 4 of 7) of the total items have been completed. The BPI User Guide is found at: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf

Variable	Type	Definition
Discontinuation of opioid medications (from T3 Survey) self-report	Binary	Discontinuation of opioid medication at 12-months will be defined as a response of "No" to the G20 question on the T3 survey that asks: "Are you currently taking an opioid medicine now? Commonly prescribed opioids include hydrocodone, oxycodone, codeine, morphine, and fentanyl." If a participant responds "No" to the survey but has opioid prescriptions in the EHR within 15-days prior to the 12-month timepoint throughout the 18-month timepoint, the participant will be considered as still taking opioid prescriptions. A response of "don't know" will be lumped with a response of yes, both indicate the patient has not discontinued their opioid.
Intent to Taper	Ordinal and Binary	A question on the T1, T2, and T3 survey includes: Please say how much you agree with this statement: "Reducing the amount of opioid medicines I take is a goal of mine" with response options including Strongly Agree, Agree, Uncertain, Disagree, and Strongly Disagree. In addition to reporting the frequency of initial response options, responses of Strongly Agree and Agree will be re-classified as an intent to taper, while Uncertain, Disagree, and Strongly Disagree will be re-classified as no intent to taper.
Relative opioid use self-report (from T3 survey)	Ordinal	A question on the T2, and T3 survey includes my overall use of opioids has increased, stayed the same, decreased.
Safety Outcomes		
Adverse Events (AEs); 1) Opioid Withdrawal Overdose, 2) Suicidality Risk, and 3) Death	Count and proportion	Three different types of AEs will be collected throughout the intervention phase of the study (from randomization to 12-months post-randomization): 1) opioid withdrawal or overdose, 2) suicidality risk, and 3) death. AEs will be identified and reported from the REDCap study database, the 6 and 12-month follow-up surveys, and deaths from the electronic medical records (based on ICD-10 codes). To identify duplicate AE reporting and corroborate AE details, we collect AE type and onset date in each database. If the same AE is reported in multiple different databases, it will only be counted as one AE. The total number of AEs as well as any documented AE will be examined.
Subgroups and Other Covariates		

Variable	Type	Definition
Mental Health Disorder Flag(s)	Binary	Includes an indicator for any mental health disorder recorded as a diagnosis in the EHR, as well as binary indicators for the presence in the EHR of each of the following: adjustment disorder, anxiety, ADHD, dementia, developmental disorder, childhood disorders, impulse control disorder, mood disorder, personality disorder, schizophrenia or other psychotic disorder, suicidal ideation, and other miscellaneous mental health disorders. The ICD-10 codes used to identify Mental Health Disorders was pulled from the Healthcare Cost and Utilization Project Agency for Healthcare Research and Quality (H-CUP-AHRQ) website (https://hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp#download)
Substance Use Disorders Flag(s)	Binary	Includes an indicator for any substance use disorder recorded as a diagnosis in the EHR, as well as binary indicators for the presence in the EHR of each of the following: alcohol use disorder and substance use disorder. The ICD-10 codes used to identify Substance Use Disorders was pulled from the Healthcare Cost and Utilization Project Agency for Healthcare Research and Quality (H-CUP-AHRQ) website (https://hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp#download)
Age	Continuous and categories	As reported on the T1 (baseline) survey and captured in REDCap. Age category will also be calculated as: < 50 50-64 >= 65
Intervention Delivery Mode	Categorical and Binary	Due to the COVID-19 pandemic, the study began offering telehealth intervention visits beginning in early-May 2020. The intervention visit Case Report Forms (CRFs) were modified to document if a specific intervention visit was administered via telehealth or in-person visit. Since the CRFs were not modified until early-August 2020, all visits up-to-that point will be defined as in-person visits. Descriptive statistics of intervention delivery mode by intervention will be displayed (in-person only, telehealth only, and mixed mode); however, for adjusted analyses, mixed mode will be combined with telehealth to create a binary classification variable.

Variable	Type	Definition
Baseline Opioid MED categories	Ordinal	See the first row in this table regarding the derivation of average daily opioid prescription (in MED). Prescriptions provided 3-months, or 90-days prior to randomization will be considered a participant's baseline opioid MED, classified as None (0 MED), Very low (1-<20 MED) Low (20 - <50 MED), Moderate (50 - <90 MED), and High (90+ MED).
BMI	Continuous and categories	Derived based on the participant's weight and height obtained from the EHR data (pulled within 90-days of the randomization date). BMI category will also be calculated as: underweight: <18.5 normal 18.5 - < 25 Overweight 25 - < 30 Obese 30 - <40 Extremely obese >=40 A dichotomous flag will also be calculated for BMI obese vs not obese.
Sex	Binary	As collected from the T1 (baseline survey); based on birth certificate.
Health Literacy	Ordinal	A shortened version of the Health Literacy Skills Instrument (HLSI) is given on the baseline (T1) survey (https://www.rti.org/impact/health-literacy-skills-instrument-hlsi). The HLSI was developed to assess four (4) domains of health literacy skills: print literacy (reading and writing), numeracy skills, oral literacy skills (listening), and information seeking (navigation of Internet and facilities). Users are presented health information stimuli on the survey which represent health related issues across the life course for health promotion and disease prevention, health care maintenance and treatment, and health system and health information navigation. The short version of the HLSI implemented in the INSPIRE study included a total of 5 questions. Health Literacy was derived into 3 categories: High Health Literacy (answering all 5 questions correctly), Average Healthy Literacy (answer 3 or 4 questions correctly), and Low Health Literacy (answering 0 to 2 questions correctly), with missing responses counted as incorrect.

Variable	Type	Definition
		Note, participants that skipped all 5 questions were not derived a total score; however, those that at least answered one question, but skipped or didn't answer other questions were counted as having an incorrect response.
Charlson Comorbidity Index	Continuous	A count of the number of diseases that make up the Charlson Comorbidity Index (see list below). Score ranges from 0 to 17, with higher scores meaning greater comorbidity. Diseases include: 1) Myocardial Infarction; 2) Congestive Heart Failure; 3) Peripheral Vascular Disease; 4) Cerebrovascular Disease; 5) Dementia; 6) Chronic Pulmonary Disease; 7) Rheumatic Disease; 8) Peptic Ulcer Disease; 9) Mild Liver Disease; 10) Diabetes without complications; 11) Diabetes with complications; 12) Paraplegia and Hemiplegia; 13) Renal Disease; 14) Cancer; 15) Moderate or Severe Liver Disease; 16) Metastatic Carcinoma; and 17) AIDS/HIV. The ICD-10 codes for each disease can be found in the specifications document.
Number of Unique Chronic non-cancer pain (CNCP) Diagnoses	Count	A count of the total number of unique CNCP diagnoses from the EHR data (diagnosis codes were pulled 3-months prior to the randomization date); a list of all CNCP diagnosis codes can be found in the specifications file. The ICD-10 codes used to classify the different CNCP ICD-10 codes was pulled from the Healthcare Cost and Utilization Project Agency for Healthcare Research and Quality (H-CUP-AHRQ) website (Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses (ahrq.gov)
Site	Categorical	3 different study sites – Duke, UNC, and Vanderbilt; Site is captured in the REDCap database
Number of Completed MI and CBT Sessions	Count and categories	The total number of completed MI and CBT sessions will be derived from the REDCap database. Per the protocol, there is 1 MI session and 8 different group CBT sessions. Categories will be defined as: Number of intervention sessions received: MI/CBT: 0, 1-2, 3-4, >4)

Variable	Type	Definition
Number of Completed SDM visits	Count and categories	The participant's total number of SDM visits will be derived from the SDM sidecar request from each study site. Note: in the SDM sidecar dataset received from study sites, each documented SDM visit from a trained SDM clinician is supplied with the associated SDM visit date and delivery mode (in-person or telehealth). Only SDM visits within the 12-month intervention phase will be counted in the total number of completed SDM visits. Categories will be defined as: Number of intervention sessions received: SDM: (0, 1-2, 3-4, >4)
Incidence of other types of pain treatments used	Binary and count	These are reported as collected on the T1-T3 survey (no derivations required) and include: non-opioid medications, over the counter pain reliever, topical pain relievers, herbal pain reliever, cortisone injections, acupuncture, hypnosis, meditation, yoga, massage, psychotherapy, CBT, chiropractic treatment, relaxation training, physical therapy, hydrotherapy, ice or heat therapy, Transcutaneous Electrical Nerve Stimulation (TENS), and other.

9.3 Analysis of the Primary Outcome

The intervention effect on opioid use over time will be assessed using a linear mixed model for repeated measures (MMRM) using categorical time effects (ref: S Davis). The model will be used to generate point and interval estimates and to test differences in mean changes in opioid dose between the 2 intervention groups at 12 months (primary timepoint) and 6 and 18 months (secondary timepoints). While the primary and secondary time assessments are at 6, 12 and 18 months, all available opioid prescription data from baseline (3-month average before randomization) through 18 months will be included in the model. The model will include fixed effects for the intervention group, time interval (as a categorical variable for **3, 6, 9, 12, 15 and 18**), intervention-by-time interaction, baseline opioid dose, and the stratification effect of institution (see model statement below). For each of the post-baseline timepoints, the daily MED is the average of the time from the end of the prior interval (i.e., 12 months is the 3- month average between 9 and 12 months). (Note: an inconsistency in the protocol inaccurately lists these intervals in protocol section 5.3.1, but they are accurately listed in Protocol Section 5.2.1).

This approach will provide consistent estimates and valid inferences under missing at random (MAR) data assumptions while accounting for correlation among multiple measures on the same participant. This mixed model will improve the power of the study and the precision of all estimates by allowing all available measures for an individual to be incorporated in the analysis, even if other timepoints are missing. An intervention-arm–specific Toeplitz correlation structure will be assumed for the within-participant measures, to reflect the changing correlation of these measures over time and permit the correlation structure to differ between study arms. Note: if this model does not converge, then a Toeplitz (not intervention-specific) structure will be fit. This model is further described below:

$$Y_{ij}=\alpha+\beta_1 Time_{ij}+\beta_2 Site_{ij}+\beta_3 Intervention_{ij}+\beta_4 BaselineOpioidDose_{iji}+\beta_5(InterventionXTime)_i+\varepsilon_{ij}$$

Where Y_{ij} is the jth measure of opioid dose in subject i, α is the intercept, $\beta 1$ through $\beta 5$ are coefficients, and ε_{ij} is the residual error term, and the variance-covariance matrix of the residual error terms for an individual is structured with the Toeplitz pattern such that all timepoints have the same variance estimate and all timepoints the same distance apart have the same correlation (i.e., the correlation between months 6 and 9 is assumed to be the same as the correlation between 9 and 12). Note that the protocol specified a random subject effect, but adding a random subject effect does not modify the model or model results in any way and so is excluded.

Statistical analysis modeling will be carried out utilizing SAS/STAT PROC MIXED (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006) with the general structure of the SAS code for this model shown below. Adjusted estimates of the change in opioid dose at each timepoint between the two interventions and corresponding 95% confidence intervals will be produced. The study's primary hypothesis will be tested by estimating the difference in the change in MED between the two intervention arms at 12 months. The study's secondary hypothesis will also be tested and associated effects estimated with the difference in LS means which estimates the average difference in the response outcome between the two intervention arms at each timepoint.

```
proc mixed data = INSPIRE_Primanaly method=REML;
class Unique_ID Intervention timepointN clinic
model change_OpioidDose_MED = Intervention timepointN clinic baseopioid
intervention*timepointN /s residual DDFM=KR;
   repeated timepointN / type=toep group=intervention subject=Unique_ID
R RCORR;
```

```
   LSMestimate 'MI-CBT vs SDM at 12 Mon'
         intervention*timepointN 0 0 0  1 0 0
                                 0 0 0 -1 0 0 / CL;
   lsmeans intervention*timepointN/ diff pdiff CL;
run;
NOTE: Variable names may differ in the actual data.
```

Note: As specified in the protocol, the distribution of opioid doses will be evaluated prior to finalizing the analysis plan, and if opioid dose change from baseline is determined to be highly skewed and substantially non-normal, then opioid doses will be transformed to the natural log scale prior to subtracting baseline. The difference of logs will be modeled, the resulting estimates will be exponentiated back to the regular scale, and thus the primary analysis will compare the intervention arms based on the *relative* dose change, interpreted as the ratio of the post-baseline dose to the baseline dose. The analysis will be completed on either the change from baseline or the ratio relative to baseline, but not both. An evaluation of the distribution of change from baseline in MED was completed after the second CDM transfer. Data was found to be essentially normally distributed, and so analysis based on the ratio scale will not be completed.

Residual plots will be output to assess model assumptions/diagnostics.

9.4 Sensitivity and Supportive Analyses of the Primary Outcome

Supportive Analyses

- The primary analysis will be repeated for the per-protocol population.
- As an *exploratory* analysis, the primary analysis may also be repeated for the modified intent to treat (MITT) population (defined below).
    - The MITT population in this study is defined as any randomized participants that received at least some of the study treatment ("randomized and treated"), defined as at least one SDM, MI, or CBT visit.
- Percentage change in daily MED from the baseline period also will be evaluated in a supportive analysis of the primary outcome. In this analysis, rather than change from baseline, each timepoints value will be calculated as the percent change, calculated as the change divided by the baseline score. The statistical model is the same.
- Compare the percentage of participants across intervention group that have met the clinically significant decrease in opioid use from baseline, defined as a decrease of 10 MED daily dose.
    - This analysis will be performed with a logistic mixed model for repeated measures (Proc GLIMMIX), similar to the method described in section 9.6.2 below.

Sensitivity Analyses - Assessment if Baseline Characteristics are Associated with Missing Outcome Data

This assessment will only be conducted if the percentage of participants with missing MED data at the 12-month timepoint is more than 20% in the ITT population.

A logistic regression model will be used to compare the demographic characteristics of participants who provided 12 months of EHR (prescription) data as compared with participants who were lost to follow-up within the EHR or withdrew consent for study participation to identify possible differential attrition. These characteristics will include but are not limited to: age, race, sex, baseline pain, baseline opioid use, co-

morbid health conditions, and others. In the event this logistic regression model indicates that a specific baseline characteristic is associated with a higher rate of attrition (or missing data), it can lead to biased results for both the primary and secondary outcomes.

If results of the logistic regression model indicate a specific baseline characteristic is associated with missing data, a sensitivity analysis on both the primary and secondary outcome measures will be implemented that run a similar model to those proposed in Sections 9.3 and 9.5 respectively, but the model will also include the baseline characteristic (i.e.: covariate) found to be associated with missing data.

Sensitivity Analyses - Multiple Imputation

This sensitivity analysis will only be conducted if the primary analysis identifies a statistically significant treatment group difference at 12 months in the ITT population and if the percentage of participants with missing MED at the 12-month timepoint is more than 20% in the ITT population. The MMRM analysis approach assumes missing data due to early study discontinuations are missing at random (MAR). As a sensitivity analysis to guard against bias if the MAR assumption is incorrect, a sensitivity analysis will be conducted on the primary outcome to assess the robustness of the primary analysis results using methods that assume that data are missing not at random (MNAR). Specifically, we will use multiple imputation of missing data to generate multiple imputed complete datasets, analysis of covariance on each dataset, and Rubin's rule to compare treatment groups across the imputed datasets (Ratitich, 2014). We will explore one sensitivity analysis called a tipping point analysis: impute missing outcomes in the intervention found to have a weaker change from baseline at varying percentages of the observed data (i.e., decreasing the observed MED) until a statistically significant treatment group difference is no longer identified (O'Kelly and Ratitch, 2014). This analysis tells us how much difference in the data could be present in order to still identify a statistically significant difference.

9.5 Analysis of Secondary Outcomes

9.5.1 PROMIS Pain Interference and Physical Functioning

Secondary analyses will generate point and interval estimates of mean change in the PROMIS physical function and PROMIS pain interference T-scores from baseline to 6 and 12 months between the 2 intervention arms. The 12-month timepoint is primary. These secondary analyses will use appropriate linear mixed model-based approaches analogous to those described for the primary outcome analysis (section 9.3), to take advantage of the data from 6-month assessments for participants who did not complete the 12-month assessment as well as the correlation between the timepoints for participants with both assessments to estimate the change at 12-months for all participants with at least one post-baseline assessment.

Since comparisons at 12 months for these outcomes are two protocol-specified secondary analyses, statistical significance will be assessed with adjustment for multiple comparisons using the Hochberg modification to the Bonferroni adjustment. The Bonferroni adjustment is a simple function of the raw p-values and is computationally

quick but can be too conservative. Step-down methods remove some conservativeness, as do the step-up methods described by Hochberg (1988). The modification procedure rejects all hypotheses with smaller or equal p-values to that of any one found less than it's critical value, defined as (i/m)/Q (where I = the individual p-value rank; m = the number of tests; and Q = the false discovery rate). Essentially, if the largest p-value for these two comparisons is < 0.05, then both tests are considered statistically significant. However, if the largest p-value is > 0.05 then the next largest p-value will be assessed for statistical significance (adjusting for multiple comparison) relative to the number of tests being performed (i.e.: 0.05/2 = 0.025).

9.6 Analysis of Other Self-Report Outcomes

9.6.1 PROMIS Anxiety, Depression, and Pain Intensity, BPI Pain Interference and Pain Severity

These analyses will generate point and interval estimates of mean change in score from baseline to 6 and 12 months between the 2 intervention arms. These analyses will use appropriate linear mixed model-based approaches analogous to those described for the secondary outcome analysis (section 9.4). Since these outcomes are exploratory in nature, no adjustments for multiple comparisons will be made.

9.6.2 Intent to Taper and Self-Report Opioid Reduction

Self-reported intent to taper opioid medication is collected at baseline, 6 and 12 months, while self-reported opioid reduction since the start of the study is only collected at 6 and 12 months. As mentioned above in section 9.2, for statistical testing, intent to taper will be categorized into a binary measure (Strongly Agree and Agree will be classified as an intent to taper while Uncertain, Disagree, and Strongly Disagree will be classified as no intent to taper). Incidence of intent to taper will be compared between intervention arm using both chi-square test of the original ordinal variable and a logistic generalized mixed model for repeated measures for the dichotomized variable. Self-report opioid reduction will be compared between intervention arm using only a chi-square test of the ordinal variable.

Chi-square tests separately at 6 months and 12 months will stratify by site and will treat the categories as ordinal with no specified numerical spacing between ordered categories using a site-stratified Mantel-Haenszel chi-square test with standardized midrank scores (SAS PROC FREQ scores=modridit option).

For the dichotomous intent to taper outcome, the model will test differences in proportions of the intention to taper opioid use between interventions at 6 and 12 months. The model will include fixed effects for the intervention arm, timepoint, site, intervention-by-time interaction, baseline intent to taper (original 5 levels, treated as a continuous covariate), and baseline opioid MED. The correlation between months 6 and 12 for a participant will be modeled with a Toeplitz structure (i.e. an "RSIDE" random effect) (ref S Davis). Odds ratio and 95% confidence interval for the odds of observing intent to taper will be estimated comparing MI+CBT to SDM).

Statistical analysis modeling will be carried out utilizing SAS/STAT PROC GLIMMIX (SAS Institute Inc. 2017) with the general structure of the SAS code for this model is shown below.

```
Proc glimmix data=Anlydata;
   class patient intervention time site;
   model Taper = intervention time site intervention*time baseopioid
btaper / solution link=logit dist=binary oddsratio DDFM=KR;
   repeated time / type=TOEP subject=patient RSIDE;
   lsmeans intervention*time / ILINK CL; /*estimates event rates*/
   lsmeans intervention*time / oddsratio diff CL; /*estimates Ors*/
   estimate 'MI-CBT vs SDM at 12 Mon' Intervention 1 -1
   intervention*time 0 1 0  -1  / EXP CL;
Run;
```

9.6.3 Discontinuation of Opioid Medications at 12-Months

Opioid discontinuation is an exploratory outcome. Opioid discontinuation at 12 months is identified on the T3 survey with supporting evidence from the prescription record (see definition above in chart). Note the protocol describes an analysis of time to opioid discontinuation, yet this was modified during the analysis planning stage prior to database lock as based on the T3 survey due to complications of not being able to identify absent prescriptions in the electronic record as being due to opioid discontinuation or a participant switching medical care outside the site's health care system.

Contingency table methods will be used to generate point and interval estimates e by intervention arm. Treatment groups will be compared using an unadjusted chi-squared test. Event rates are expected to be too small to stratify by clinical site. If event rates are large enough for the risk difference to be relevant, the risk difference 95% CIs will be based on Wald-type CIs under the null hypothesis as specified by Sato. If the expected cell counts < 5, groups will be compared with a Fisher's exact test, and the exact risk difference 95% CI will be obtained using the score statistic based on Chan and Zhang (ref Chan, Zhang 1995). Where specified, Binomial 95% CIs for proportions are based on the exact Clopper-Pearson method (ref Clopper, Pearson).

In SAS, using PROC FREQ:

The chi-square test and corresponding risk difference 95% CI are obtained by:

```
table Intervention*Opid_discont /riskdiff(method=wald equal
         var=null)chisq ;
```

For small expected cell counts, the exact test and risk difference 95% CI are obtained by:

```
 table Intervention* Opid_discont / fisher; exact riskdiff
(method=score);
```

9.6.4 Pre-Specified Subgroup Tests of Heterogeneity and Interaction with Intervention Type on Opioid Use

We will test the interaction of intervention and 2 pre-defined subgroups (having at least one comorbid mental health condition (excluding substance use disorders) and sex) at the p=0.10 significance level. Model based analyses to evaluate heterogeneity of intervention effects by the specific subgroups will use a model analogous to those described above for the primary analysis (section 9.3); however, a subgroup type and the interaction between subgroup type and intervention arm will be added to the model. This addition will allow the study team to estimate difference between intervention for each subgroup category separately, and determine if there is a substantial difference in opioid change at 12 months between the subgroups for the SDM and MI+CBT intervention arms (interaction: does one intervention have substantially better or worse outcome than the other intervention for one of the subgroups but not the other subgroup). The general structure of the SAS code is shown below (example is for sex).

```
proc mixed data = INSPIRE_Primanaly method=REML;
     class Unique_ID intervention timepointN Sex clinic;
     model change_OpioidDose_MED = Intervention timepointN clinic
     baseopioid intervention*timepointN  Sex intervention*Sex
     intervention*Sex*timepointN /s residual DDFM=KR;
     repeated timepointN / type=toep group=intervention
     subject=Unique_ID R RCORR;
     lsmeans Intervention*Sex*timepointN / diff pdiff CL;
     lsmestimate intervention*Sex*timepointN "CBT vs. SDM for Females at
     12-months"
                          0 0 0  1 0 0 0 0 0 0 0 0
                          0 0 0 -1 0 0 0 0 0 0 0 0/ CL;
     lsmestimate intervention*Sex*timepointN "CBT vs. SDM for males at
     12-months"
                          0 0 0 0 0 0 0 0 0 1 0 0
                          0 0 0 0 0 0 0 0 0 -1 0 0/ CL;

     contrast "trt by sex interaction at 12m"
              Intervention*Sex 1 -1 -1 1
              intervention*Sex*timepointN
                          0 0 0  1 0 0 0 0 0 -1 0 0
                          0 0 0 -1 0 0 0 0 0  1 0 0/ CL;
     run;
     NOTE: Variable names may differ in the actual data.
```

9.6.5 Exploratory Evaluation of Heterogeneity of Intervention Response to Opioid Use

We will explore treatment heterogeneity for a set of 10 potential covariates listed below, as specified in the protocol and defined in section 9.2 above:

- presence of substance use disorders (yes/no)
- presence of comorbid diseases (Charlson comorbidity index none vs 1 or more),
- Age (< 50, 50-64, 65+) ,
- baseline pain (BPI pain severity score of mild, moderate, severe),

- health literacy (high, average, and low),
- BMI categories (obese vs. not obese),
- Taking non-opioid pain medications prescribed (yes/no) per T1 survey
- Baseline opioid dose categories (very low, low, moderate, and high).
- Intervention delivery mode (in-person vs. telehealth),
- Number of intervention visits/sessions received: 0, 1-2, 3-4, >4

Testing for heterogeneity of intervention effects for these subgroups will be considered *exploratory*, and the results will be interpreted with appropriate caution. The outcome of interest for these analyses will be change from baseline to 12-months in average prescribed daily opioid dose (in MED). Analyses for these 10 covariates will be conducted the same as for the 2 specified subgroups above (Section 9.6.4), except that p-values will be interpreted descriptively only.

Each covariate will be added individually to a repeated measures MMRM models similar to the primary model described above while intervention arm is still adjusted for in the model .

9.6.6 Percent of Participants Who Experienced an Increase in Pain and Decrease in Opioid Dose

We will explore the percentage of participants per intervention arm who experienced an increase in pain and a decrease in opioid dose. For this analysis, we will create a binary variable for both the PROMIS Pain Interference and Physical Functioning outcomes that identifies participants that experienced an increase in these measures at 12-months since baseline (i.e.: an increase in the Pain Interference T-Score at 12-months and a decrease in the Physical Functioning T-Score at 12-months). We will then categorize their opioid use at 12-months in comparison to baseline as: unchanged (within +/- 10 MED), decreased (<10 MED), or increased (>10 MED).

We will provide descriptive statistics via contingency table that displays the count and proportion of participants, per intervention arm, that experienced an increase in pain but a decrease in opioid prescriptions. No hypothesis or statistical testing will occur.

9.7 Safety and Adverse Events Analysis

AEs will be listed and summarized overall and by event term (hospitalization or emergency department visit due to opioid withdraw/overdose, suicidality, and death). The number and proportion of individuals experiencing an adverse event, any SAE, and AE by relationship to intervention will also be summarized.

Contingency table methods will be used to generate point and interval estimates of the risk of each AE type by intervention arm. Treatment groups will be compared using an unadjusted chi-squared test. Event rates are too small to stratify by clinical site. For events with expected cell counts < 5, groups will be compared with a Fisher's exact test, where specified, Binomial 95% CIs for proportions are based on the exact Clopper-Pearson method (ref Clopper, Pearson).

In SAS, using PROC FREQ:

The chi-square test and corresponding risk difference 95% CI are obtained by:

```
table Intervention *AEvar /riskdiff(method=wald equal var=null)chisq ;
```

For small expected cell counts, the exact test and risk difference 95% CI are obtained by:

```
table Intervention*AEvar / fisher; exact riskdiff (method=score);
```

10 REFERENCES

a. Boonstra, A., Schiphorst Preuper, H., Balk G., Stewart, R. (2014). Cut-off points for mild, moderate, and severe pain on the visual analogue scale for pain in patients with chronic musculoskeletal pain. Pain, 12, 2545-2550.

b. Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800–803.

c. Davis SM. Chapter 5: Mixed models for repeated measures with categorical time effects (MMRM). In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

d. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). SAS® for Mixed Models, Second Edition, Cary, NC: SAS Institute Inc.

e. Ratitch, B. Chapter 6: Multiple Imputation. In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

f. O'Kelly M, Ratitch, B. Chapter 8: Analyses under missing-not-at-random assumptions. In Clinical Trials with Missing Data: a Guide for Practitioners. O'Kelly, M and Ratitch B. Wiley. 2014.

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

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

i. Chan IS, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics. 1999;55(4):1202-1209.

j. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika, 1934 26(4): 404–413

k. Sato T. On the variance estimator for the Mantel-Haenszel risk difference. Biometrics. 1989; 45:1323-1324.

11 TABLE SHELLS AND DISPLAYS

All table shells and displays are documented in a separate supplemental file to this SAP. Data displays may be added, deleted, rearranged or the structure may be modified after finalization of the SAP. Such changes require no amendment to the SAP as long as the change does not contradict the text of the SAP.