Statistical Analysis Plan

Sponsor Name:	**Opioid Post-Marketing Requirement (PMR) Consortium**
Protocol Number and Title:	**2065-5: A Randomized, Double-Blind, Placebo-Controlled Clinical Trial of Structured Opioid Discontinuation versus Continued Opioid Therapy in Suboptimal and Optimal Responders to High-Dose Long-Term Opioid Analgesic Therapy for Chronic Pain**
Protocol Version and Date:	**Original Protocol: 10 Jan 2016** **Amendment 1: 07 Jul 2016** **Amendment 2: 02 Feb 2017**
INC Research Project Code:	**1004904A**
Author:	**Laura Truskowski** **Principal Biostatistician**
SAP Version:	**1.3**
SAP Version Date:	**11 April 2018**

Statistical Analysis Plan

Version: 1.3 **Version Date:** 11 April 2018

I confirm that I have reviewed this document and agree with the content.

APPROVALS	
INC Research	
Lead Biostatistician Laura Truskowski Principal Biostatistician	Date (dd-Mmm-yyyy): 09-May-2018
Senior Reviewing Biostatistician Kelley Wekheye Manager, Biostatistics	Date (dd-Mmm-yyyy): 09-May-2018
Opioid PMR Consortium	
Errol Gould, Ph.D. on behalf of the Opioid PMR Consortium	Date (dd-Mmm-yyyy): 04-MAY-2018
Diana Meske on behalf of the Opioid PMR Consortium	Date (dd-Mmm-yyyy): 03-MAY-2018

Statistical Analysis Plan

TABLE OF CONTENTS

Statistical Analysis Plan

1. GLOSSARY OF ABBREVIATIONS

Abbreviation	Description
AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
BPI-SF	Brief Pain Inventory – Short Form
BSODP	Blinded Structured Opioid Discontinuation Period
CI	confidence interval
CLBP	chronic low back pain
CMH	Cochran-Mantel-Haenszel
COWS	Clinical Opiate Withdrawal Scale
C-SSRS	Columbia-Suicide Severity Rating Scale
DHEAS	dehydroepiandrosterone sulfate
DSST	Digit Symbol Substitution Test
EQ-5D-5L	EuroQOL 5 dimensions (5 levels of response) instrument
ER	extended-release
FDA	Food and Drug Administration
FSFI	Female Sexual Function Index
ICH	International Conference on Harmonization
IIEF	International Index of Erectile Function
IR	immediate release
IRT	interactive response technology
ITT	Intent-to-Treat
LA	long-acting
LBP	lower back pain
LOCF	last observation carried forward
LSMeans	least squares means

Statistical Analysis Plan

Abbreviation	Description
MADDERS	Misuse Abuse Diversion Drug Event Reporting System
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model repeated measures
MOS	Medical Outcomes Study
MSER	morphine sulfate extended-release
MSIR	morphine sulfate immediate-release
NRS	numerical rating scale
OCER	oxycodone extended-release
OCIR	oxycodone immediate-release
OMER	oxymorphone extended-release
OMIR	oxymorphone immediate-release
OPC	Opioid PMR Consortium
PCS	potentially clinically significant
PGIC	Patient Global Impression of Change
PHQ-8	Patient Health Questionnaire Depression Scale
PI	pain intensity
PID	pain intensity difference
PMR	Post-Marketing Requirement
PP	Per-protocol
PQAS	Pain Quality Assessment Scale
PRN	pro re nata (as needed)
PT	preferred term
QST	quantitative sensory testing
RMDQ	Roland-Morris Disability Questionnaire
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan

Statistical Analysis Plan

Abbreviation	Description
SD	standard deviation
SE	standard error
SI	International System of Units
SOC	System Organ Class
SOWS	Subjective Opiate Withdrawal Scale
TEAE	treatment emergent adverse event
UDT	urine drug testing
VAS	visual analog scale
WHODD	World Health Organization Drug Dictionary
WPAI	Work Productivity and Activity Impairment scale

2. PURPOSE

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables and figures which will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

2.1. Responsibilities

INC Research will perform the statistical analyses and is responsible for the production and quality control of all tables, figures and listings.

2.2. Timings of Analyses

The analysis of safety and efficacy is planned after all subjects complete the final study visit or terminate early from the study. No interim analysis is planned.

3. STUDY OBJECTIVES AND HYPOTHESES

3.1. Primary Objective

The primary objective is to evaluate the effect on pain intensity (PI) of structured discontinuation of long-term opioid analgesic therapy compared to continuation of opioid therapy in Suboptimal and Optimal Responders to high-dose, long-term opioid analgesic therapy for treatment of chronic low back pain (CLBP).

3.2. Secondary Objective

The secondary objective is to evaluate the effect on additional clinical outcomes measures of structured discontinuation of long-term opioid analgesic therapy compared to continuation of opioid therapy for CLBP in Suboptimal and Optimal Responders to high-dose, long-term opioid analgesic therapy at multiple time points.

3.3. Exploratory Objectives

The exploratory objectives are:

- To compare experimental pain sensitivity in the opioid discontinuation vs. continuation groups in both Suboptimal Responders and Optimal Responders (subjects in substudy)
- To determine subject characteristics that predict response to structured discontinuation of opioid therapy
- To determine the effect of opioid discontinuation on certain endocrine function tests, sexual function, and the relationship of changes in PI and endocrine function to changes in male and female sexual function
- To determine the effect of opioid discontinuation on neurocognitive function

3.4. Hypotheses

The study hypotheses for Suboptimal and for Optimal Responders are as follows:

Suboptimal Responders: In subjects with CLBP on high-dose opioid therapy for at least 1 year who have a suboptimal response to opioid treatment, structured discontinuation results in an improvement in pain compared to continuation of opioid treatment.
Optimal Responders: In subjects with CLBP on high-dose opioid therapy for at least 1 year who have an optimal response to opioid treatment, structured discontinuation of opioid treatment results in worsening of pain compared to continuation of opioid treatment.

4. STUDY DESCRIPTION

4.1. Subject Selection

Subjects enrolled in the study must have been taking extended release (ER) or long-acting (LA) opioids or doses of immediate-release opioids at least 4 times a day for at least 12 months for non-radicular chronic low back pain (CLBP). The enrolled subjects must also have been taking high doses as defined in the table below of one of the following "index ER opioids" for their CLBP: morphine sulfate ER (MSER), oxycodone ER (OCER), or oxymorphone ER (OMER) for at least 3 consecutive months before the Screening Visit. The goal is to have a minimum representation of each index ER opioid (a minimum of approximately 20% of randomized subjects for each). Specific inclusion and exclusion criteria are listed in the Protocol.

Index ER Opioid	Daily Dose Range
Morphine sulfate ER	120-540 mg
Oxycodone ER	80-360 mg
Oxymorphone ER	40-180 mg

A subject will be classified at the Screening Visit as an Optimal Responder if his daily Average PI score is ≤4 and he is satisfied with his pain and physical function. A subject will be classified as a Suboptimal Responder if his daily Average PI score is ≥6 and he is dissatisfied with his pain and physical function.

Approximately 820 subjects will be randomized into the Blinded Structured Opioid Discontinuation Period (BSODP). For each Responder type, 205 subjects per arm will be needed.

4.2. Brief Description

This multicenter, randomized, double-blind, placebo-controlled study will consist of a common Screening Visit for all subjects, then different schedules for Optimal and Suboptimal Responders, followed by a common schedule for the BSODP and Follow-up Period.

4.2.1. Suboptimal Responders

Subjects classified as Suboptimal Responders at the Screening Visit will enter a 1-week Run-in Period during which they will be required to discontinue all previously prescribed opioid medications for the duration of the study and will receive a standardized regimen of study medication consisting of the index ER opioid plus the immediate-release (IR) opioid matching the index ER opioid they were taking at screening.

At the end of the 1-week Run-in Period, Suboptimal Responders will return to the clinic for a Tolerability Visit to determine whether they tolerated the standardized opioid regimen.

Suboptimal Responders who tolerated the regimen will then enter a 1-week Baseline Period during which they will self-administer the standardized opioid regimen and record their PI scores (Average and Worst over the past 24 hours) on a 0-10 numerical rating scale (NRS) daily before bedtime. Each subject's 7-day daily PI scores will be averaged and used as the subject's baseline PI scores.

At the Randomization Visit, Suboptimal Responders will continue on study if they have (i) a baseline mean Average PI score (mean of average daily score) ≥6; (ii) compliance with their index ER opioid study medication between 80-120%; and (iii) are still dissatisfied with their pain and physical function. Subjects who had a baseline mean Average PI score of 10/10 (i.e., they recorded a 10 every day) will be discontinued.

4.2.2. Optimal Responders

Subjects classified as Optimal Responders at the Screening Visit will enter a 1-week Observation Period, where they will be told to continue their current medication (including non-index ER and IR opioid medications). At the end of the Period, subjects will return to the clinic and be confirmed as Optimal Responders if (i) their index ER opioid medication use was ≥120 to ≤540 mg morphine equivalents/day on average with 80-120% compliance over the Observation Period; (ii) their mean Average PI score was ≤5; and (iii) they are still satisfied with their pain and physical function.

Optimal Responders will then enter a Taper Period of up to 2 weeks. Subjects will be tapered with the same index opioid they were taking at screening (using study medication provided for the index ER opioid). No other ER opioid will be allowed, nor any IR opioid. Subjects with a mean Average PI score >5 over 3 or more consecutive PI scores (excluding missing values) , which has also increased by ≥1.5 points from the mean Average PI over the 7-day Observation Period ,will be allowed to continue. Subjects who reach 2 weeks with their index ER medication tapered to 0 without meeting these criteria will be discontinued.

Optimal Responders who meet the criteria from the Taper Period will proceed to an Open-Label Titration Period of up to 3 weeks. The index ER opioid (supplied as ER opioid plus matching IR opioid) will be titrated with increases as frequently as every 4 days until the mean Average PI score over at least 3 consecutive pain scores is ≤5 and the dose of index ER opioid is ≥120 mg and ≤540 mg morphine equivalents/day in order to qualify for randomization.

4.2.3. Blinded Structured Opioid Discontinuation Period and Follow-up

All subjects who qualify for randomization will be randomized to either continue or discontinue opioid therapy during the BSODP lasting 24 weeks. Subjects will return to the clinic at Weeks 1, 2, 4, 6, 8, 12, 16, 20, and 24 for a series of assessments at each visit.

Subjects in the discontinuation arms will be completed tapered off their opioid treatment over approximately 3-4 weeks, depending on the baseline dose of the index ER opioid.

Subjects who complete the BSODP will advance to a 4-week Follow-up Period. However, those subjects who discontinue from the BSODP before Week 12 will complete Week 12 study procedures and those subjects who discontinue during Weeks 13-23 will complete Week 24 study procedures before proceeding to the Follow-up Period. During the Follow-up Period, subjects will self-administer a blinded taper of placebo for subjects in the discontinuation arm and active drug taper for subjects in the continuation arm.

4.3. Determination of Sample Size

The primary efficacy endpoint is the change from baseline to Week 12 (mean Average PI over the week before the Week 12 visit) in 0-10 NRS Average PI score comparing the opioid discontinuation group to the continued opioid therapy group. To detect a between-group difference of 0.8 points with 90% power, assuming a standard deviation (SD) of 2.5, using a two-sided 5% significance level, 205 randomized subjects per arm for each Responder type will be needed, for a total of 820 subjects. The calculation is based on a 2-sample t-test with equal variances assumed between the two groups. It is assumed that approximately 60-70% of the subjects will complete the study. Assuming a 3:1 screen failure rate, approximately 3280 subjects will need to be screened. Subjects who discontinue from the study will not be replaced.

4.4. Treatment Assignment and Blinding

Subjects are randomized in a 1:1 ratio to either the continued opioid therapy arm or the structured discontinuation of opioid therapy arm. Central randomization will be used, stratified by Responder status (Optimal Responder, Suboptimal Responder) and Index ER opioid (morphine, oxycodone, or oxymorphone). A minimum of 20% of subjects will be randomized into each of the baseline ER opioid strata within each Responder cohort. The interactive response technology (IRT) system will assign a unique randomization number which will dictate the treatment assignment and matching study drug kit for the subject.

In case of emergency, the Investigator will have the ability to break the blind at the clinical site for the welfare of the subject. Breaking the blind will be accomplished by

using the IRT web interface. The Investigator should discuss the necessity of breaking the blind with the Medical Monitor first.

4.5. Administration of Study Medication

The study medications are Food and Drug Administration (FDA) approved and will be provided by the Opioid PMR Consortium (OPC) as follows:

- OCER: OxyContin® (oxycodone hydrochloride ER tablets)
- MSER: MS Contin® (morphine sulfate ER tablets)
- OMER: Opana® ER (oxymorphone hydrochloride ER tablets)

Placebos for each of the abovementioned ER drugs will be manufactured by the respective manufacturer of the active ER medication and will be identical to the respective drugs in aspect, size, and color.

Associated IR medications (oxycodone immediate-release (OCIR), morphine sulfate immediate-release (MSIR), and oxymorphone immediate-release (OMIR)) are commercially available and will be provided by the OPC in open-label fashion as study prescribed drugs:

- OCIR: Mallinckrodt brand of oxycodone IR tablets
- MSIR: West-Ward brand of morphine sulfate (IR tablets)
- OMIR: Opana® (Endo brand of oxymorphone hydrochloride IR tablets)

The dose of the IR opioids will be limited to no more than twice daily as needed (PRN) administration of OCIR 10 mg, MSIR 15 mg, or OMIR 5 mg, matching the index ER opioid each subject is receiving.

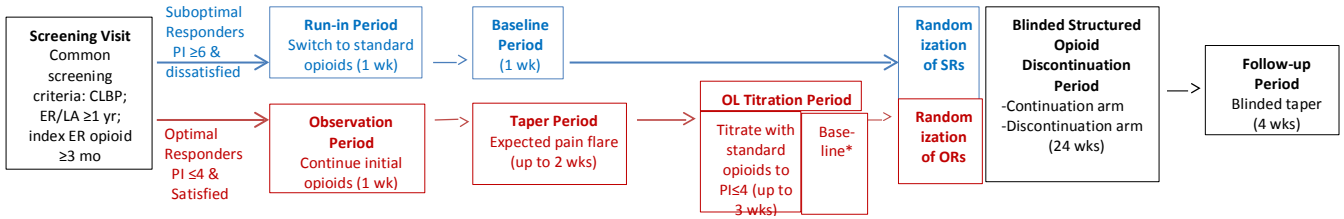
Subjects will also be provided acetaminophen as rescue medication (500 mg tablets, 1-2 tablets every 4-6 hours PRN, not to exceed 6 tablets/day). Subjects will be encouraged to take acetaminophen to manage their pain before they use the IR opioid.

Subjects in the continuation arm will receive tablets of their index ER opioid plus matching placebo tablets to maintain the opioid dose at a fixed regimen, but appear as if tapering is occurring.

Subjects in the structured discontinuation arm will receive placebo tablets matching the index ER opioid tablets plus index ER opioid tablets in decreasing dose to result in a tapering off the index ER opioid.

4.6. Study Procedures and Flowchart

Figure 1: Study Design



wk=week, mo=month, SR=Suboptimal Responder; OR=Optimal Responder.
ER/LA requirement can be satisfied by use of immediate-release opioid at least 4 times a day for at least 1 year
* Baseline PI scores=mean Average PI score including the Average PI scores during the Titration Period that meet the qualification criteria (Average PI score ≤5 for 3 consecutive non-missing values) plus any scores between the final qualification day and the Randomization Visit.

Note: In the tables below, blue shading denotes periods of the study that are identical for Optimal and Suboptimal Responders.

Statistical Analysis Plan

Table 1: Schedule of Procedures for Suboptimal Responders

Period	Screening		Run-In Period (1 week)		Baseline Period (1 week)	Random-ization Visit	Blinded Structured Opioid Discontinuation Period[15] (24 weeks)									Follow-Up Period (4 weeks)	
Week			-2 to -1	-1	-1 to 0	0	1	2	4	6	8	12	16	20	24	26	28
Month						0		0.5	1	1.5	2	3	4	5	6	6.5	7
Visit Name[1]	Screening Visits[2]		--	SR1 Tolerability Visit	--	SR2	BDP1	BDP2	BDP3	BDP4	BDP5	BDP6	BDP7	BDP8	BDP9	FUV1	FUV2
Informed consent	X																
Demographics	X																
Medical history	X																
Physical exam[3]	X											X			X		X
Vital signs[4]	X			X		X	X	X	X	X	X	X	X	X	X	X	X
ECG	X																
Clinical laboratory	X											X			X		
Pregnancy test[5]	X					X											X
Quebec Task Force Classification for Spinal Disorders	X																
Daily 24-hr Average and Worst PI score in past 24 hours on NRS[6]	X		←----------X----------→			X	←----------X----------→										
RMDQ	X					X						X			X		
Daily administration of ER and IR opioids (phone)	X[7]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Check entry or continuation criteria[8]	X	X		X		X											
Dispense Study Drugs[9]		X		X		X	X	X	X	X	X	X	X	X	X	X	
Collect Study Drugs & Drug Accountability[9]				X		X	X	X	X	X	X	X	X	X	X	X	X
BPI-SF	X					X						X			X		
Online Patient Support Program		Access Granted	←----------Encourage use of online patient support program----------→														
Regional Pain Scale						X						X			X		
PQAS						X						X			X		
EQ-5D-5L						X						X			X		
C-SSRS[10]	X									X					X		
PHQ-8	X											X			X		
MOS Sleep Scale						X						X			X		
DSST						X						X			X		

Statistical Analysis Plan

Table 1: Schedule of Procedures for Suboptimal Responders (Continued)

Period	Screening		Run-In Period (1 week)		Baseline Period (1 week)	Random-ization Visit	Blinded Structured Opioid Discontinuation Period[15] (24 weeks)									Follow-Up Period (4 weeks)	
Week			-2 to -1	-1	-1 to 0	0	1	2	4	6	8	12	16	20	24	26	28
Month						0		0.5	1	1.5	2	3	4	5	6	6.5	7
Visit Name[1]	ScreeningVisits[2]		--	SR1 Tolerabil-ity Visit	--	SR2	BDP1	BDP2	BDP3	BDP4	BDP5	BDP6	BDP7	BDP8	BDP9	FUV1	FUV2
QST for OIH (for subjects enrolled at participating sub-study sites)						X						X			X		
Endocrine laboratory tests and sexual function questionnaire[11]	X											X			X		
Abuse-related events			←---------- X ---------- →														
Opioid withdrawal (SOWS)						X		X	X	X						X	
Opioid withdrawal (COWS)						X		X	X							X	
PGIC												X			X		
UDT[12]	X								X			X			X		
Work productivity (WPAI)						X						X			X		
AEs[13]		X	←---------- X ---------- →														
Telephone Calls[14]			←---------- X ---------- →														
Prior/concomitant medications	←---------- X ---------- →															X	X

1. Visit windows are ±3 days until BDP6, and ±5 days thereafter.
2. Screening procedures will be performed over 2 visits no more than 21 days apart.
3. Full physical exam at Screening and brief physical exam afterwards.
4. At Screening: Height, weight, BMI, pulse rate, respiratory rate, and blood pressure. At subsequent visit: pulse rate, respiratory rate and blood pressure only. If BMI > 40, complete STOPBang Questionnaire and record in source documents. Subject cannot continue if apnea risk is high.
5. Serum pregnancy test at Screening and urine pregnancy test at randomization and final visit.
6. Captured once daily by phone at bedtime except at screening when captured in the office.
7. Daily dose of opioid and rescue medications at screening will be taken from medical records or medical history.
8. At Tolerability Visit (SR1): Subjects who did not tolerate the standardized regimen will be discontinued. At Randomization Visit: Subjects will be included if they have an Average PI score ≥6 on 0-10 NRS over the Baseline Period (with a minimum compliance of 4 out of 7 daily PI scores), not have an Average PI score each day of the 7-day Baseline Period equal to 10/10 and be dissatisfied with their pain and physical function.
9. Three (3) drugs will be dispensed, collected, and counted: ER opioids, IR opioids, and acetaminophen.
10. C-SSRS "Lifetime" will be done at Screening and the C-SSRS "Since Last Visit" will be done at 6 weeks and 24 weeks after randomization.
11. Blood for endocrine function tests will be drawn at Screening and sent to lab but only tested on Randomized subjects. Sexual function questionnaires will be administered when endocrine function tests are drawn.
12. Quantitative UDT for illicit drugs, non-prescribed controlled substances (opioid and non-opioid), and alcohol. To prevent unblinding, after randomization, results for the index opioid and its metabolites will not be provided to the investigator.
13. Only serious AEs will be collected in the eCRF during Screening.
14. Phone call every week of the study where a visit is not conducted to check for wellbeing, study drug compliance and safety issues. Changes in medications and AEs obtained during phone calls will be captured in the IRT or eCRF as appropriate. All other information will be captured in source documents.
15. The Early Discontinuation Assessment Tool will be administered to any subject who discontinues during the BDP.

Statistical Analysis Plan

BDP=Blinded Discontinuation Period; FUV=Follow-up Visit

Table 2: Schedule of Procedures for Optimal Responders

Period	Screen-ing		Observation Period (1 week)		Taper Period (up to 2 weeks)		OL Titration Period (up to 3 weeks)[1]	Random-ization Visit	Blinded Structured Opioid Discontinuation Period[16] (24 weeks)									Follow-Up Period (4 weeks)	
Week			-6 to -5	-5	-5 to -3	-3	-3 to 0	0	1	2	4	6	8	12	16	20	24	26	28
Month								0		0.5	1	1.5	2	3	4	5	6	6.5	7
Visit Name[2]	Screen-ing Visits[3]		--	OR1	OR2 (wk -4)	OR3	Telephone or office (OR3.1 – OR3.6) visits every 4 days	OR4	BDP1	BDP2	BDP3	BDP4	BDP5	BDP6	BDP7	BDP8	BDP9	FUV1	FUV2
Informed consent	X																		
Demographics	X																		
Medical history	X																		
Physical exam[4]	X													X			X		X
Vital signs[5]	X			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
ECG	X																		
Clinical laboratory	X													X			X		
Pregnancy test[6]	X							X											X
Quebec Task Force Classification for Spinal Disorders	X																		
Daily 24-hr Average and Worst PI score in past 24 hours on NRS[7]	X		←-----X-----→		←-----X-----→		←---X----→	X	←----------X----------→										
RMDQ	X							X						X			X		
Daily dose or administration of ER and IR opioids and rescue meds taken (phone)	X[8]		←-----X[8]-----→		←-----X[8]-----→		←---X[8]---→	X[8]	←----------X[8]----------→									X[8]	X[8]
Check entry or continuation criteria	X	X		X[9]	X[9]	X[9]		X[9]											

Table 2: Schedule of Procedures for Optimal Responders (Continued)

Period	Screen-ing		Observation Period (1 week)		Taper Period (up to 2 weeks)		OL Titration Period (up to 3 weeks)[1]	Random-ization Visit	Blinded Structured Opioid Discontinuation Period[16] (24 weeks)									Follow-Up Period (4 weeks)	
Week			-6 to -5	-5	-5 to -3	-3	-3 to 0	0	1	2	4	6	8	12	16	20	24	26	28
Month								0		0.5	1	1.5	2	3	4	5	6	6.5	7
Visit Name[2]	Screen-ing Visits[3]		--	OR1	OR2 (wk -4)	OR3	Telephone or office (OR3.1 – OR3.6) visits every 4 days	OR4	BDP1	BDP2	BDP3	BDP4	BDP5	BDP6	BDP7	BDP8	BDP9	FUV1	FUV2
Dispense Study Drugs[10]				X	X	X		X	X	X	X	X	X	X	X	X	X	X	
Collect Study Drug & Drug Accountability[10]				X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
BPI-SF	X							X						X			X		
Online Patient Support Program		Access Granted			←----------Encourage use of online patient support program----------→														
Regional Pain Scale								X						X			X		
PQAS								X						X			X		
EQ-5D-5L								X						X			X		
C-SSRS[11]	X											X					X		
PHQ-8	X													X			X		
MOS Sleep Scale								X						X			X		
DSST								X						X			X		
QST for OIH (for subjects enrolled at participating sub-study sites)								X						X			X		
Endocrine laboratory tests and sexual function questionnarie[12]	X													X			X		
Abuse-related events			←---------- X ----------→																

Statistical Analysis Plan

Table 2: Schedule of Procedures for Optimal Responders (Continued)

Period	Screen-ing		Observation Period (1 week)		Taper Period (up to 2 weeks)		OL Titration Period (up to 3 weeks)[1]	Random-ization Visit	Blinded Structured Opioid Discontinuation Period[16] (24 weeks)									Follow-Up Period (4 weeks)	
Week			-6 to -5	-5	-5 to -3	-3	-3 to 0	0	1	2	4	6	8	12	16	20	24	26	28
Month								0		0.5	1	1.5	2	3	4	5	6	6.5	7
Visit Name[2]	Screen-ing Visits[3]		--	OR1	OR2 (wk -4)	OR3	Telephone or office (OR3.1 – OR3.6) visits every 4 days	OR4	BDP1	BDP2	BDP3	BDP4	BDP5	BDP6	BDP7	BDP8	BDP9	FUV1	FUV2
Opioid withdrawal (SOWS)								X		X	X	X						X	
Opioid withdrawal (COWS)								X		X	X							X	
PGIC														X			X		
UDT[13]	X										X			X			X		
Work productivity (WPAI)								X						X			X		
AEs[14]		X	←	----------	----------	----------	----------	X	----------	----------	----------	----------	----------	----------	----------	----------	→		
Telephone Calls[15]			←	----------	----------	----------	----------	X	----------	----------	----------	----------	----------	----------	----------	----------	→		
Prior/concomitant medications	←	----------	----------	----------	----------	----------	----------	X	----------	----------	----------	----------	----------	----------	----------	----------	→	X	X

1 If the dose of the index ER opioid needs to be titrated above the baseline dose, the Titration Period may be extended by 1 additional week (+3 days).
2. Visit windows are ±3 days until BDP6, and ±5 days thereafter.
3. Screening procedures will be performed over 2 visits no more than 21 days apart.
4. Full physical exam at Screening and brief physical exam afterwards.
5. At Screening: Height, weight, BMI, pulse rate, respiratory rate, and blood pressure. At subsequent visit: pulse rate, respiratory rate, and blood pressure only. If BMI > 40, complete STOPBang Questionnaire and record in source documents. Subject cannot continue if apnea risk is high.
6. Serum pregnancy test at Screening and urine pregnancy test at randomization and final visit.
7. Captured once daily by phone at bedtime except at Screening when captured in the office.
8. Daily dose of opioid and rescue medications at screening will be taken from medical records or medical history. Continuation of daily dose of opioid and rescue medications during the Observation Period will be confirmed by phone. Subsequently, confirmation of daily administration of ER and IR opioid study medications will be obtained by phone.
9. At end of Observation Period: Subjects must have had an index ER opioid medication use of ≥120 mg and ≤540 mg morphine equivalents/day (see table in protocol) on average and have had 80% to 120% compliance over the 1-week Observation Period; (ii) have a mean Average PI score over the 1 week of the period ≤5 (with a minimum compliance of 4 out of 7 daily PI scores); and (iii) still be satisfied with their pain and physical function. During or at the end of the Taper Period: Subjects must have a mean Average PI score >5 over ≥3 consecutive scores (non-missing values), where this mean Average PI score has also increased by ≥1.5 points from the mean Average PI score over the 7-day Observation Period. Note that if these criteria are satisfied at Visit OR2, subject may skip OR2 and advance directly to Visit OR3. At the end of Titration Period: Subjects must have completed their titration and have achieved a mean Average PI score over at least 3 consecutive scores (non-missing values) ≤5 with acceptable side effects and a minimum dose of ≥120 mg and ≤540 mg morphine equivalents/day to continue on study.
10. The non-blinded ER opioid taper dose will be dispensed at OR1 and OR2; Open-label ER opioids, IR opioids, and acetaminophen will be dispensed starting at OR3 (for titration in the Titration Period); Double-blind ER opioid and open-label IR opioid and acetaminophen will be dispensed at OR4. Collection and counting of study medication will be done at all visits except OR1.
11. C-SSRS "Lifetime" will be done at Screening and the C-SSRS "Since Last Visit" will be done at 6 weeks and 24 weeks after randomization.

Statistical Analysis Plan

12. Blood for endocrine function tests will be drawn at Screening and sent to the lab but tested only on randomized subjects. Sexual function questionnaires will be administered when endocrine function tests are drawn.
13. Quantitative UDT for illicit drugs, non-prescribed controlled substances, and alcohol. To prevent unblinding, after randomization, results will not be provided to the investigator for the index opioid and its metabolites.
14. Only serious AEs will be collected in the eCRF during Screening.
15. Phone call every week of the study where a visit is not conducted to check for wellbeing, study medication compliance, and safety issues. Changes in medication and AEs obtained during phone calls will be captured in the IRT or eCRF as appropriate. All other information will be captured in source documents.
16. The Early Discontinuation Assessment Tool will be administered to any subject who discontinues during the BDP.

BDP=Blinded Discontinuation Period; FUV=Follow-up Visit

5. ENDPOINTS

5.1. Primary Efficacy Endpoint

The primary efficacy endpoint is the change in the mean Average PI score on the 0-10 NRS from baseline to the 1 week period before the Week 12 visit.

5.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

For Suboptimal Responders:

- Changes from baseline to Weeks 4, 8, 16, 20, and 24 in 0-10 NRS mean Average PI scores over 1 week before each visit
- Changes from baseline to Weeks 4, 8, 12, 16, 20, and 24 in 0-10 NRS mean Worst PI scores over 1 week before each visit
- Cumulative response function in percent improvement in PI score at Weeks 12 and 24 compared to baseline over 1 week before each visit
- Changes from baseline to Weeks 12 and 24 in physical function as measured by the Roland-Morris Disability Questionnaire (RMDQ) score
- Changes from baseline to Weeks 12 and 24 in the impact of pain on function as measured by the Brief Pain Inventory – Short Form (BPI-SF) score
- Changes from baseline to Weeks 12 and 24 in pain quality as measured by the Pain Quality Assessment Scale (PQAS) score
- Changes from baseline to Weeks 12 and 24 in pain spread as measured by the Regional Pain Scale
- Changes from baseline to Weeks 12 and 24 in sleep quality as measured by the Medical Outcomes Study (MOS) Sleep Scale
- Changes from baseline to Weeks 12 and 24 in mood as measured by the Patient Health Questionnaire Depression Scale (PHQ-8)
- Changes from baseline to Weeks 12 and 24 in quality of life as measured by the EuroQOL 5 dimensions (5 levels of response) instrument (EQ-5D-5L)
- Changes from baseline to Weeks 12 and 24 in work productivity as measured by the Work Productivity and Activity Impairment (WPAI) scale
- Patient Global Impression of Change (PGIC) at Weeks 12 and 24

- Proportion of subjects with overall clinical benefit at Weeks 12 and 24 defined as a composite measure of key clinical endpoints: ≥30% improvement in mean Average PI score, ≥20% improvement in RMDQ, and PGIC of moderately or better improvement

For Optimal Responders:

- Changes from baseline to Weeks 4, 8, 16, 20, and 24 in 0-10 NRS mean Average PI score over 1 week before each visit
- Changes from baseline to Weeks 4, 8, 16, 12, 20, and 24 in 0-10 NRS mean Worst PI scores
- Cumulative response function in percent worsening in PI score at Weeks 12 and 24 over 1 week before each visit compared to baseline
- Changes from baseline to Weeks 12 and 24 in physical function as measured by the RMDQ score
- Changes from baseline to Weeks 12 and 24 in the impact of pain on function as measured by the BPI-SF score
- Changes from baseline to Weeks 12 and 24 in pain quality as measured by the PQAS score
- Changes from baseline to Weeks 12 and 24 in pain spread as measured by the Regional Pain Scale
- Changes from baseline to Weeks 12 and 24 in sleep quality as measured by the MOS Sleep Scale
- Changes from baseline to Weeks 12 and 24 in mood as measured by the PHQ-8
- Changes from baseline to Weeks 12 and 24 in quality of life as measured by the EQ-5D-5L
- Changes from baseline to Weeks 12 and 24 in work productivity as measured by the WPAI scale
- PGIC at Weeks 12 and 24

5.3. Exploratory Measurements

Exploratory measurements include:

- Change in sensitivity to thermal stimuli on quantitative sensory testing (QST) in substudy subjects

- Identification of baseline factors predictive of benefit of opioid discontinuation among Suboptimal Responders
- Changes from baseline in endocrine function and sexual function as characterized in the male and female questionnaires; relationship of changes in endocrine function to changes in PI; relationship of changes in endocrine function to changes in male and female sexual function; and relationship of baseline endocrine status to efficacy of opioid discontinuation
- Changes from baseline in neuropsychological function as characterized by the Digit Symbol Substitution Test (DSST)

5.4. Assessment of Safety

Safety will be assessed by:

- Adverse events (AEs)
- Abuse-related events (using the Misuse Abuse Diversion Drug Event Reporting System [MADDERS™])
- Opioid-specific side-effects
- Vital signs (heart rate, blood pressure, and respiratory rate only)
- Clinical laboratory parameters
- Physical examination

Other safety measures will include:

- Urine drug testing (UDT)
- Columbia-Suicide Severity Rating Scale (C-SSRS)
- Opioid withdrawal effects (measured by the Subjective Opiate Withdrawal Scale [SOWS], the Clinical Opiate Withdrawal Scale [COWS], and withdrawal AEs)

6. ANALYSIS SETS

Separate analyses will be conducted for Suboptimal Responders and Optimal Responders since these are considered two different study populations. Therefore, all analysis sets will be defined for Suboptimal Responders and Optimal Responders separately.

6.1. Safety Sets

Different Safety Sets will be defined to summarize data pre- and post-randomization.

- For Suboptimal Responders, the Safety Set Pre-Randomization will include all subjects who were enrolled in the study and were switched to the standardized regimen (i.e., who enter the Run-in Period) and took at least one dose of study drug.
- For Optimal Responders:
  - The Safety Set Pre-Randomization (Observation Period) will include all Optimal Responder subjects who entered the Observation Period.
  - The Safety Set Pre-Randomization (Taper Period) will include all Optimal Responder subjects who were enrolled in the study and took at least 1 dose of study medication during the Taper Period.
  - The Safety Set Pre-Randomization (Titration Period) will include all Optimal Responder subjects who were enrolled in the study and took at least 1 dose of study medication during the Titration Period (i.e., who enter the Open-label Titration Period)

For both Suboptimal Responders and Optimal Responders, the Safety Set Post-Randomization will include all subjects who were randomized and received at least one dose of double-blind treatment.

Safety will be analyzed separately for these analysis sets. Subjects who are randomized to one treatment group but mistakenly assigned the medication kit of the other treatment group for over half their time in the study will be reported “as treated”.

6.2. Randomized Set

The Randomized Set will include all subjects who were randomized into the BSODP. The analysis set will be summarized in the disposition tables and will be the analysis set for the summary of protocol deviations to indicate subjects who were randomized but did not receive study drug in the BSODP.

6.3. Intent-to-treat Set

The Intent-to-treat (ITT) Set will include all subjects who were randomized and received at least one dose of study drug after randomization. All efficacy analyses will be performed using this analysis set. Subjects will be reported in the treatment group to which they were randomly assigned.

6.4. Per-protocol Set

The Per-protocol (PP) Set is a subset of the ITT set. It will include subjects who have at least 6 weeks of valid mean Average PI scores, not including baseline, but including a valid Week 12 mean Average PI score (a valid mean is calculated from 2 or more daily Average PIs).

In addition, the subjects must have satisfied the following inclusion criteria:

- Have a clinical diagnosis of non-radicular CLBP for a minimum of 12 months and
  - For Suboptimal Responders, pain must have been present for at least several hours a day and have an Average PI score of 6-9 within the past 24 hours of screening
  - For Optimal Responders, subjects must have an average PI score of 1-4 within the past 24 hours of screening;

- Have been taking ER/LA opioids (or immediate release opioids at least 4 times a day) for at least 12 months

- Have been taking one of the 3 index opioid drugs around-the-clock at a twice-a-day frequency for at least 3 consecutive months

Also, subjects in the PP Set must not have violated the following exclusion criteria at the time of randomization:

- Have a primary diagnosis of fibromyalgia, complex regional pain syndrome, neurogenic claudication due to spinal stenosis, spinal cord compression, acute nerve root compression, severe or progressive lower extremity weakness or numbness, bowel or bladder dysfunction as a result of cauda equina compression, diabetic amyotrophy, meningitis, diskitis, back pain because of secondary infection or tumor, or pain caused by a confirmed or suspected neoplasm

- Have undergone a surgical procedure for back pain within 6 months prior to the Screening Visit

- Have had a nerve or plexus block, including epidural steroid injections or facet

blocks, within one month prior to the Screening Visit or botulinum toxin injection in the lower back region within 3 months prior to screening

The subjects must also not have more than 2 instances of medication compliance outside of the 80%-120% range at each compliance assessment after randomization.

Protocol deviations and other criteria for inclusion in the PP Set will be assessed before unblinding.

6.5. Protocol Deviations

Protocol deviations will be categorized and graded as important or not important. Protocol deviations will be summarized for the Randomized Set. The summary will show the frequency and percentage of subjects with any protocol deviation by treatment group and for all subjects. Each category of deviation will also be summarized, counting subjects with at least one deviation in the category. A comprehensive listing of protocol deviations will be provided.

7. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

7.1. General Methods

All summary tables will be presented separately for Suboptimal and Optimal Responders, as determined at Screening.

All efficacy analyses will be conducted separately for Suboptimal and Optimal Responders, without adjustments for multiplicity. Safety analyses will be presented according to the defined Safety Sets Pre- and Post-Randomization.

Summary statistics for continuous variables will include number of subjects, mean, SD, median, first and third quartiles, minimum, and maximum by treatment group. Summary statistics for categorical variables will include the number of subjects with data, and frequencies and percentages for each category by treatment group. Unless stated otherwise, percentages will be based on the number of subjects with nonmissing observations at that time point.

All analyses will be performed using SAS® version 9.3 or higher unless otherwise specified.

All demographic, background, efficacy, and safety data collected as part of this study will be included in subject data listings. Listings will be by Responder type and randomized treatment group within Responder type.

7.2. Key Definitions

7.2.1. Definition of Baseline

For PI scores, baseline will be defined differently for Suboptimal and Optimal Responders.

For Suboptimal Responders, the mean of the available Average PI scores over the 7-day Baseline Period will count as the subject's baseline Average PI score for statistical analysis. Similarly, the mean of the Worst PI scores over the 7-day Baseline Period will count as the subject's baseline Worst PI score for statistical analysis. At least 4 out of 7 scores must be present.

For Optimal Responders, the mean of the Average PI scores during the Titration Period that meet the qualification criteria (Average PI score ≤5 for 3 consecutive non-missing values) and any scores between the last qualification score and the Randomization Visit will count as the subject's baseline Average PI score for statistical analysis. Similarly, the mean of the Worst PI scores over the 3 qualification days (as defined above) during the Titration Period and any scores between the qualification score and the

Randomization Visit will count as the subject's baseline Worst PI score for statistical analysis.

The baseline mean PI scores (Average and Worst) will be rounded to one decimal place.

For all other assessments, the last non-missing observation before the first dose in the BSODP will be used as the baseline observation. For example, if an assessment is collected at screening and at randomization and the randomization assessment is missing, then the screening value will be used.

7.2.2. Mean Pain Intensity Scores

Subjects are asked to rate their pain daily before bedtime using the 0-10 NRS based on Worst pain and Average pain over the past 24 hours. For the scheduled post-randomization visits, mean Average PI and mean Worst PI at the visit will be defined as the means of the respective PI scores over the 7 days preceding the visit. If there is only one daily PI score available over the 7 days before a particular visit, the mean will not be calculated and the data point will be considered missing. The means will be rounded to one decimal place before analysis.

7.2.3. Treatment Emergence for Adverse Events

Each AE occurring before randomization into the BSODP will be classified into the Pre-Randomization period (defined by the Safety Sets) for which it was treatment emergent according to its start date and time relative to the start and end dates and times of each period. Any AE that was not present before the start date and time of each period, or was present before but increased in intensity during the period will be considered treatment emergent.

For all subjects, treatment-emergent adverse events (TEAEs) in the Post-Randomization Period will be those that started or worsened on or after the date and time of the first dose of double-blind treatment in the BSODP up to 28 days after the subject's last visit date.

Further detail on classifying AEs as TEAEs and the handling of partial or missing AE start dates and times is given in Section 12.3.

7.3. Visit Windows

Visit windows specified in the protocol are ±3 days up to the Week 12 visit of the BSODP and ±5 days thereafter. However, in general, data points for visits falling outside of these windows will still be mapped to the named scheduled visit, except for data recorded as a discontinuation visit in the BSODP. These will be mapped to the next scheduled visit for the assessment after the last scheduled visit.

If 2 or more data values have the same visit designation, the one closest to the target day will be used. If the 2 values are equidistant from the target day, the earlier value will be used.

For PI scores, weekly means will be calculated for each week using the 7 days up the seventh day of each week (i.e., Day 1-7 for Week 1, Day 8-14 for Week 2) as long as there are 2 or more scores present during the week. The mean score for each week may not coincide with scheduled visits. However, for Week 12, the mean Average and Worst PI scores will be calculated from the scores taken on the day of the Week 12 visit and the preceding 6 days. Daily PI scores from days used in the Week 11 calculation will not be used in the Week 12 mean, so there must be 2-7 days of scores present to calculate a valid Week 12 mean. The Week 24 mean PI scores will be similarly determined.

7.4. Pooling of Study Sites

No pooling of study sites will be specified for analysis, as terms for site will not be included in the statistical models.

7.5. Subgroups

Study data will be summarized and analyzed separately for Optimal Responders and Suboptimal Responders. Descriptive summaries of the values and mean changes from baseline to Week 12 in mean Average PI will be presented by index ER opioid within each Responder group.

The primary efficacy analysis will be repeated for subgroups defined by sex and (separately) by age (<=65 years, >65 years) within each Responder group.

8. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION

8.1. Subject Disposition and Withdrawals

Subject disposition will be summarized separately for the Pre- and Post-Randomization periods, as well as separately for Optimal and Suboptimal Responders.

Pre-randomization, the separate summaries of subject disposition for each Responder type will be based on all subjects entering each period who were classified to each Responder type. This count will be the denominator for percentages in each section of the table. For Suboptimal Responders, the frequencies and percentages of subjects randomized into the BSODP, not randomized into the BSODP, and the reasons for discontinuation before randomization will be presented.

For Optimal Responders, the frequencies and percentages of subjects entering each period Pre-Randomization (Observation, Taper, and Open-label Titration) will be presented. The frequencies and percentage of subjects discontinuing during the period and reasons for discontinuation will be also be presented for each period, including discontinuations before randomization.

The summaries of subject disposition Post-Randomization for each Responder type will be based on all randomized subjects and will be by treatment group and overall. For each column, the number of subjects randomized into each treatment group and overall will be the denominator for percentages and will be tabulated first, without percentages.

The frequencies and percentages of subjects in each post-randomization analysis set will be tabulated below the frequencies of randomized subjects.

The frequencies and percentages of subjects who completed the BSODP and who prematurely discontinued during the BSODP will be presented. The specific reasons for early discontinuation will be shown, also with frequencies and percentages. “Other” reasons for discontinuation will be summarized as one category, rather than by separate text categories.

The frequencies and percentages of subjects randomized into the BSODP who enter and do not enter the Follow-up Period will be summarized, with reasons for not entering also summarized. Subjects completing and discontinuing from the Follow-up Period will be summarized with frequencies and percentages including reasons for discontinuation. The denominator for percentages in this section of the disposition summary will be the number of subjects who enter the Follow-up Period.

A flow chart showing the numbers of subjects entering each period and moving on to

the next or discontinuing will be provided.

For subjects who discontinue from the BSODP, the Early Discontinuation Assessment Tool will be summarized with frequencies and percentages of subjects responding yes and no to each question. For each question, the level of importance will also be summarized with frequencies and percentages of subjects responding yes to the question.

The categories of change in pain will be summarized for severity of pain, area of pain, and sensitivity of pain.

8.2. Demographics

Demographics include age, sex, race, ethnicity, baseline height and weight, and body mass index (BMI).

Age in years will be calculated from the date of informed consent and the date of birth.

Height and weight values in inches and pounds, respectively, will be used in summaries and listings. Units for BMI will be kg/m^2.

The demographic summaries will be by treatment group, including an all subjects column. Age, height, weight and BMI will be summarized with descriptive statistics for quantitative assessments. Sex, race, and ethnicity will be summarized with frequencies and percentages. Age will also be categorized as <25 years, 25-49 years, 50-64 years, and >=65 years and summarized.

Separate demographic summaries will be produced for the Safety Set Post-Randomization, ITT, and PP Sets within each Responder type.

8.3. Baseline Characteristics

Baseline characteristics to be summarized by treatment group and overall are:

- Baseline Average PI
- Baseline Worst PI
- Index ER Opioid
- Duration of Use of Index ER Opioid
- Quebec Task Force on Spinal Disorders Classification
- Time Since Diagnosis of CLBP

The PI scores, duration of index ER opioid use, and time since diagnosis of CLBP will be summarized with descriptive statistics for quantitative assessments. The remaining baseline disease characteristics are qualitative and will be summarized with frequencies and percentages of subjects in each category.

Separate summaries will be produced for the Safety Set Post-Randomization, ITT, and PP Sets within each responder type.

8.4. Medical History

Medical history will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA). Summaries by System Organ Class (SOC) and Preferred Term (PT) will be produced for the Safety Set Post-Randomization. The summaries will be in alphabetical order by SOC and by PT within each SOC. Subjects will be counted once on the overall level (any medical history), SOC level, and PT level.

8.5. Prior and Concomitant Medications

Medications taken before and during the study will be coded according to the World Health Organization Drug Dictionary (WHODD). Summaries of medications will be by Anatomical Therapeutic Chemical (ATC) level 3 and preferred drug name within ATC level 3. If the level 3 term is not available, the level 2 term will be used. Subjects taking a medication multiple times will be counted only once on the ATC level and on the preferred drug name level. Drugs will appear in alphabetical order by ATC level 3 term and preferred drug name within the ATC term.

Medications will be classified as prior to enrollment (started before the signing of informed consent, regardless of stop date), as prior within each of the periods defined by the Safety Sets Pre-Randomization (continued past the start and end dates of the period or started on the start date of the period or before the end date), or concomitant within the BSODP. A particular medication may be classified as being in more than one period.

If the start date is partial, then it will be imputed using the same rules as for AEs listed in Section 12.3. If the stop date is partial and the medication was not classified as prior or concomitant in a period using the start date alone, then the stop date will be imputed as follows for determination of concomitance within a period only:

- If the day is missing, then the last day of the month and year will be used. If this date is after the last date of the latest period the subject entered where the months are the same, then the last such date will be used. This causes the medication to be considered concomitant in the latest period which ends with the same medication stop month.

- If the day and month are missing, then the last day of the year will be used. If this date is after the last date of the latest period the subject entered and the years are the same, then the last such date will be used. This causes the medication to be considered concomitant in the latest period which ends with the same medication stop year.
- If the stop date is completely missing, then it will be set to the last date of the latest period into which the subject entered, such that it will be considered concomitant in that period.

8.5.1. Pre-randomization Prior Medications

Prior medications taken before randomization and before study enrollment will be summarized.

Medications will be classified as prior to enrollment (started before the signing of informed consent, regardless of stop date), or as prior within each of the periods defined by the Safety Sets Pre-Randomization as described in Section 8.5. A particular medication may be classified as being in more than one period.

8.5.2. Post-randomization Concomitant Medications

Post-randomization concomitant medications will be summarized by treatment group.

The classification of medications as post-randomization concomitant is described in Section 8.5. A particular medication may be classified as being in more than one period.

9. EFFICACY

All efficacy analyses will present descriptive statistics and, when appropriate, statistical tests to compare the 2 treatment groups to which subjects were randomized. All statistical comparisons will be 2-sided at the 0.05 alpha level, with separate analyses for the Suboptimal Responders and the Optimal Responders. There will be no adjustments for multiplicity.

9.1. Primary Efficacy Endpoint and Analysis

The primary efficacy endpoint is the change from baseline to the 1 week period before the Week 12 visit in the mean Average PI. The primary analysis will be performed using the ITT Set.

The analysis of the primary efficacy endpoint will be in two stages. First, multiple imputation will be used for missing data based on the reason for missing data. Second, the multiple datasets will be analyzed using mixed model repeated measures (MMRM), and the individual imputation dataset results will be combined using standard methods (e.g., Little & Rubin, 2002). The imputation and analysis are described below.

Missing data occurring between observed data points for a given subject will be imputed using Markov Chain Monte Carlo (MCMC) methods as provided by the SAS MI procedure to create monotone missing data patterns for each subject.

Multiple imputation (MI) will then be used to create 10 complete datasets from the monotonic datasets for the mean Average PI data using regression models with terms for treatment group and each subject's previous weekly mean Average PI scores. This step generates imputed values for subjects who discontinue the study for any reason using all subjects in the dataset.

Values imputed by multiple imputation in these 10 datasets will then be replaced by alternate imputed values specifically for subjects who discontinue due to AEs and for subjects who discontinue due to Lack of Efficacy.

For subjects who discontinue due to AEs, missing weekly mean Average PIs after discontinuation will be imputed by Responder type as follows:

- For Suboptimal Responders: a randomly generated response from a normal distribution with a mean equal to the subject's baseline Average PI score and variance as calculated from all subjects at the corresponding time point of missing data within the Suboptimal Responders.
- For Optimal Responders: a randomly generated response from a normal distribution with a mean equal to the subject's mean Average PI score over the

first 3 days at the end of the Taper Period which qualified the subject for entry into the Titration Period (i.e., the mean Average PI score is >5 for ≥3 consecutive days and has increased by ≥1.5 points from the mean Average PI over the 7-day Observation Period). The variance for the randomly generated score will be calculated from all subjects at the corresponding time point of missing data within the Optimal Responders. This mean Average PI score is used as a conservative imputation, as it will be worse than the baseline mean Average PI score for an Optimal Responder.

For missing mean Average PI scores caused by discontinuation due to lack of efficacy, data will be imputed using the subject's last non-missing weekly mean Average PI score.

All imputed PI scores will be rounded to one decimal place. Any imputed values >10 will be set to 10, and any imputed values <0 will be set to 0.

The analysis of the individual imputation datasets for the primary endpoint will use an MMRM analysis which includes the fixed effects of treatment group, index ER opioid (morphine, oxycodone, or oxymorphone), Week, and the interaction between treatment group and Week as factors, and baseline Average PI as a covariate. Data up to Week 12 will be included in this analysis. The Restricted Maximum Likelihood estimation approach will be used, and the default covariance structure will be unstructured. If there are convergence problems for this model, other variance-covariance matrix structures will be considered, such as Toeplitz, auto-regressive, and compound symmetry.

The MMRM results from each dataset will be combined using the SAS MIANALYZE procedure to generate a single test statistic and p-value for the primary test of significance. Model-based least squares means (LSMeans) of the changes from baseline for the 2 treatment groups and for the difference in means will be presented with their standard errors (SE) and 95% confidence intervals (CI), as estimated from MIANALYZE.

Summary statistics obtained from the MI results will be provided for the mean Average PI values at each visit and changes from baseline to each visit.

9.2. Secondary Efficacy Endpoints and Analyses

9.2.1. Secondary Analyses of the Primary Endpoint

The primary efficacy endpoint of change from baseline to Week 12 in mean Average PI will be summarized separately for each of the 3 index ER opioids with descriptive statistics. These summaries will include the imputed values.

The analysis described in Section 9.1 will be repeated for the PP Set. The same 10

imputation datasets created for the ITT Set analysis will be used for this analysis, but restricted to the PP Set.

Two sensitivity analyses of the primary efficacy endpoint using the ITT Set will be performed as follows:

i) Missing mean Average PI values will be imputed using the last observation carried forward (LOCF). Analysis of covariance (ANCOVA) will be used to assess the treatment difference in change from baseline to Week 12. The model will include terms for baseline Average PI, index ER opioid, and treatment group.

ii) The same MMRM model as used for the primary analysis will be applied to the observed data up to Week 12 without imputation of missing values. LSMeans with SEs and 95% CIs will be presented for the changes from baseline to Week 12 by treatment group and for the treatment difference.

9.2.2. Analysis of Pain Intensity Secondary Endpoints

Change from baseline to Weeks 4, 8, 16, 20, and 24 in mean Average PI score and change from baseline to Weeks 4, 8, 12, 16, 20, and 24 in mean Worst PI score will be summarized with descriptive statistics, without imputation of missing values. An MMRM model as described for the primary efficacy analysis (without imputation of missing values) will be fitted and used to generate LSMeans, SEs, and 95% CIs for each time point by treatment group and for the treatment difference. P-values will be provided for the treatment difference at each week.

9.2.3. Analysis of Scale Score Secondary Endpoints

The changes from baseline to Weeks 12 and 24 in the scale scores will be analyzed using ANCOVA with terms for baseline score as a covariate, index ER opioid, and treatment group. There will be no imputation of missing values. LSMeans with SEs and 95% CIs will be presented for the changes from baseline to each visit by treatment group and for the treatment difference.

The scales and scoring algorithms are described in the following sections.

9.2.3.1. Roland-Morris Disability Questionnaire

The RMDQ used in this study is a series of 24 statements that can be used to describe the subject's back pain on the day the questionnaire is administered. The score for the RMDQ is the sum of the marked statements, so ranges from 0-24, where 0 = no disability and 24 = maximum disability.

9.2.3.2. Brief Pain Inventory – Short Form

The BPI-SF is a 9-item self-administered questionnaire used to evaluate the severity of a subject's pain and the impact of this pain on the subject's daily functioning. The subject is asked to rate their worst, least, average, and current pain intensity (0-10), list current treatments and their perceived effectiveness, and rate the degree that pain interferes with general activity, mood, walking ability, normal work, relations with other persons, sleep, and enjoyment of life on a 0-10 point scale (0 = does not interfere, 10 = completely interferes).

The 0-10 pain intensity scores will be summarized separately for worst, least, and average over the past 24 hours, and for pain right now. A composite Pain Index will be calculated as the mean the 4 pain intensities, rounded to one decimal place.

Each of the 7 pain interference scores will be summarized separately. If at least 4 of the 7 components are answered, then a composite Pain Interference Index will be calculated as the mean of the non-missing individual pain interference scores, rounded to one decimal place.

9.2.3.3. Pain Quality Assessment Scale

The PQAS is a 20-item instrument developed to quantify the quality and intensity of pain associated with all types and categories of pain problems, including nociceptive and neuropathic pain. Seventeen items measure intensities of various aspects or types of pain on 0-10 scales (0=not intense, 10 = most intense). The other 3 items quantify the unpleasantness of pain and the intensity of deep pain and surface pain, also on 0-10 scales. All items and changes from baseline will be summarized separately, without any overall score.

In addition, the final question on the PQAS addresses time quality of pain, described as intermittent, variable, and stable. Type of pain will be summarized with a shift table with frequencies and percentages showing baseline type across the columns and type at Week 12 and Week 24 across the rows. Total columns and rows will be included to show the distribution at each visit. Each shift table by Responder type, treatment group, and week will be restricted to subjects who have a response at baseline and at the visit, which will be the denominator for percentages.

9.2.3.4. Regional Pain Scale

The Regional Pain Scale measures the location and intensity of pain in 38 articular and nonarticular regions. The intensity of pain or tenderness is indicated on a scale of None, Mild, Moderate, and Severe. All items are expected to have a response, even if the response is None. The measure for each visit and change from baseline is the number of regions with pain of Mild or higher.

In addition, each region will be summarized with shift tables from baseline to each visit (baseline severity across the columns by severity at each visit down the rows). Rows and columns for totals will be included. The denominators for percentages in each shift table will be the number of subjects with a response at the visit and at baseline for the region.

9.2.3.5. Medical Outcomes Study Sleep Scale

The MOS Sleep Scale is a 12-item questionnaire which measures sleep quality in 7 scales over the past 4 weeks: sleep disturbance, snoring, sleep short of breath or headache, sleep adequacy, sleep somnolence, and 2 sleep problems indexes. In addition, the average hours of sleep over the past 4 weeks is recorded as a raw measure and also coded as an optimal sleep index.

The MOS will be scored and the sleep scales calculated according to the MOS Sleep Scale User's Manual v1.0 (Spritzer and Hays, 2003).

9.2.3.6. Patient Health Questionnaire Depression Scale (PHQ-8)

The PHQ-8 is an 8-item questionnaire that aims at assessing the level of depression of a subject. Each item is scored from 0 = “not at all” to 3= “nearly every day”; the total score, which is the sum of the scores for each item, can be from 0 to 24. A score ≥10 is considered major depression and ≥20 is severe major depression.

The total score will be considered missing if more than 1 item is missing. If only 7 of the 8 questions are answered, the sum will be weighted up by multiplication by 8/7.

9.2.3.7. Work Productivity and Activity Impairment

The WPAI used in this study is for the specific health problem of lower back pain (LBP). The questionnaire consists of 6 questions regarding the past 7 days:

1. Currently employed
2. Hours of work missed due to LBP
3. Hours of work missed for other reasons
4. Hours actually worked
5. Degree LBP affected productivity while working (0-10)
6. Degree LBP affected regular activities (0-10)

Questions 2-5 are skipped if the subject was not employed during the past week (7 days).

Outcomes from the WPAI:LBP are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity. Referring the responses to

questions 2-6 as Q2-Q6, respectively, the impairment percentages are calculated as:

% work time missed due to LBP = 100*Q2/(Q2+Q4)

% impairment while working due to LBP = 10*Q5

% overall work impairment due to LBP = 100*{Q2/(Q2+Q4) + [(1-(Q2/(Q2+Q4)))*(Q5/10)]}

% activity impairment due to LBP = 10*Q6

In addition to the ANCOVA and summaries of actual and changes from baseline, the 4 percentages will be categorized as 0%, >0-5%, >5-10%, >10-25%, >25-50%, >50-75%, >75%-<100%, 100% and summarized with frequencies and percentages. Differences between the treatment groups in the distributions will be assessed using a Cochran-Mantel-Haenzsel (CMH) test stratified by index ER opioid with row mean scores, appropriate for ordered categories.

9.3. Additional Efficacy Analyses

Additional efficacy analyses stated in the protocol are described in the following sections. No imputation of missing data will be employed for any of these analyses.

9.3.1. Cumulative Response Function - Mean Average PI

For the Suboptimal Responders, the cumulative response function in percent improvement in mean Average PI score at Weeks 12 and 24 relative to baseline over 1 week before each visit will be shown by treatment group. Since reduction in mean Average PI represents improvement, the pain intensity difference (PID), defined as baseline Average PI – mean Average PI at visit, expressed as a percentage of baseline, will be used. The categories used will include percent PID relative to baseline of >0% (any improvement), and ≥10% to ≥50% in steps of 10% as a decreasing function. If sufficient subjects show improvement ≥50%, additional cut points will be added.

For the Optimal Responders, the worsening from baseline represented by the actual changes from baseline to Weeks 12 and 24 as percentages of baseline will also be shown (including >0% (any worsening), and ≥10% to ≥50% in steps of 10%), similarly to the summary for improvement.

9.3.2. Cumulative Response Function - RMDQ

For the Suboptimal Responders, the cumulative response function in percent improvement in RMDQ score (0-24) at Weeks 12 and 24 relative to baseline over 1 week before each visit will be shown by treatment group. Since reduction in RMDQ score represents improvement, the negative of change from baseline will be expressed as a

percentage of baseline and summarized similarly to mean Average PI.

Likewise, for Optimal Responders, worsening in RMDQ score as a percentage of baseline will be summarized.

9.3.3. Overall Clinical Benefit – Suboptimal Responders

The overall clinical benefit for the Suboptimal Responders is defined as ≥30% improvement in mean Average PI and ≥20% improvement in RMDQ and PGIC of moderately or better improvement. The proportion of subjects in the Suboptimal Responder group with overall clinical benefit at Weeks 12 and 24 will be analyzed using logistic regression for binary endpoints with model terms for baseline Average PI, baseline RMDQ, index ER opioid, and treatment group. Subjects who are missing data for any of the components defining overall clinical benefit will be excluded from the analysis for the week with missing data.

9.3.4. Time to Discontinuation Due to Lack of Efficacy

The time to discontinuation due to lack of efficacy starting from Day 1 of the BSODP and for any reason will be summarized by treatment group using Kaplan-Meier estimates of the survival function. Median and quartile times will be shown with 95% CIs.

Subjects who complete the BSODP will be censored on the last day of the period (~24 weeks). For the summaries of subjects who discontinue to lack of efficacy, subjects who do not discontinue or discontinue for other reasons will be censored on their last day of contact within the BSODP.

9.3.5. Patient Global Impression of Change

The PGIC is a self-administered questionnaire that assesses the subject's level of improvement/worsening from the beginning to the end of treatment. Subjects are asked to select the category of change that most closely describes any change experienced in the pain in their painful areas from the beginning of the BSODP to Week 12 and to Week 24.

The scale has levels describing change as:

Very much improved
Much improved
Minimally improved
No change
Minimally worse
Much worse
Very much worse

The frequencies and percentages of subjects reporting each category of change will be summarized by treatment group at each visit. Differences in the distributions between treatment groups will be assessed with a CMH test stratified by index ER opioid using row mean scores, as is appropriate for ordered categories.

9.3.6. EuroQOL 5 Dimensions (5 Levels of Response) Instrument

The EQ-5D-5L measures quality of life in 5 dimensions:

Mobility
Self-care
Usual activities
Pain/discomfort
Anxiety/depression

Each is rated in 5 levels from no problems/pain/anxiety to being unable/extreme pain/extreme anxiety. The responses for each category will be summarized by treatment and visit with frequencies and percentages reporting each level.

In addition, a visual analog scale (VAS) rates the subject's health on a 0-100 scale from the worst imaginable health state to the best imaginable health state. The VAS values and changes from baseline to each visit will be summarized with descriptive statistics.

9.4. Exploratory Efficacy Analyses

Exploratory analyses are:

- Change in sensitivity to thermal stimuli on QST (subjects participating in the substudy). This will be reported in a separate report of the substudy.
- Univariate and multivariate regression analyses with each subject characteristic as an independent variable to determine the subjects' characteristics that predict suboptimal or optimal response to structured discontinuation of opioid therapy
- Changes from baseline in endocrine function and sexual function as characterized in the male and female questionnaires; relationship of changes in endocrine function to changes in PI; relationship of changes in endocrine function to changes in male and female sexual function; and relationship of baseline endocrine status to efficacy of opioid discontinuation
- Changes from baseline in neurocognitive function as characterized by the number of correct responses from the DSST

9.4.1. Subject Characteristics Predictive of Response

The baseline characteristics listed in Section 8.3 will be used as covariates and factors in regression or ANCOVA models against the change from baseline to Week 12 in mean Average PI. Each baseline characteristic will be fit individually into the model including terms for treatment and treatment-by-baseline characteristic interaction and the terms tested for significance. A forward-stepwise regression approach will also be used to fit the best predictive model. Only suboptimal responders in the ITT Set with a Week 12 mean Average PI will be used in these analyses.

MMRM models using observed data for subjects in the discontinuation arms (without multiple imputation), including Week as a factor will also be used to assess the predictive ability of each baseline characteristic.

9.4.2. Digit Symbol Substitution Test

The DSST assesses overall neuropsychological function. The test is sensitive to brain damage, dementia, age, and depression, and is a widely used instrument for measuring the neuropsychological effects of opioid therapy. The digits (1-9) are paired with symbols, and the test consists of matching the symbol for a series of digits as fast as possible. The measure is the number of correct symbols in 90 seconds, which will be analyzed using ANCOVA with model terms of baseline score, stratification parameters, and treatment group using observed data (without multiple imputation)

9.4.3. Changes in Endocrine and Sexual Function

Endocrine function will be assessed at Screening, Week 12, and Week 24.

The endocrine endpoints will be:

- HPG axis: Total testosterone, free testosterone (assayed using the equilibrium dialysis method), FSH, LH, and, in women estradiol
- HPA axis: ACTH, cortisol, dehydroepiandrosterone sulfate (DHEAS)
- Growth hormone axis: IGF-1

Questionnaires regarding sexual function (International Index of Erectile Function [IIEF] in males and Female Sexual Function Index [FSFI] in females) will be obtained at the same time point that endocrine blood samples are drawn.

The IIEF is a 15-item questionnaire which encompasses 5 domains of sexual function in

males: Erectile Function, Orgasmic Function, Sexual Desire, Intercourse Satisfaction, and Overall Satisfaction. Each question is scored on a scale of 0 to 5 or 1 to 5. Missing responses are scored as 0. The score for each domain is obtained by summing the scores for each domain as follows:

Domain	Items	Maximum Score
Erectile Function	1, 2, 3, 4, 5, 15	30
Orgasmic Function	9, 10	10
Sexual Desire	11, 12	10
Intercourse Satisfaction	6, 7, 8	15
Overall Satisfaction	13, 14	10

The FSFI is a 19-item questionnaire which encompasses 6 domains of sexual function in females: Desire, Arousal, Lubrication, Orgasm, Satisfaction, and Pain. A full scale score is obtained by summing the scores for each domain. Each question is scored on a scale of 0 to 5 or 1 to 5. Missing responses are scored as 0, which also indicates no reported sexual activity. The score for each domain has a maximum of 6 and is obtained by summing the scores for each domain and multiplying by a factor as follows:

Domain	Items	Factor
Desire	1, 2	0.6
Arousal	3, 4, 5, 6	0.3
Lubrication	7, 8, 9, 10	0.3
Orgasm	11, 12, 13	0.4
Satisfaction	14, 15, 16	0.4
Pain	17, 18, 19	0.4

ANCOVA models for change from baseline in mean Average PI will be fit with terms for treatment and for each individual change from baseline endocrine endpoint along with interaction terms with treatment. Significance of terms will be reported.

For males and females separately, Pearson and Spearman correlations for changes in baseline to Week 12 and Week 24 in each endocrine function endpoint against each change from baseline to the corresponding time point in sexual function domain will be tabulated.

ANCOVA models for change from baseline to Week 12 and for change from baseline to Week 24 in mean Average PI will be fit with terms for treatment and each endocrine endpoint baseline as a covariate and its interaction with treatment in separate models. Significance of terms will be assessed.

Only observed data will be used in these analyses without imputation of missing values.

10. ANALYSIS OF PHARMACOKINETICS

Not applicable.

11. ANALYSIS OF PHARMACODYNAMICS

Not applicable.

12. SAFETY

Safety during the BSODP will be assessed using the Safety Set Post-Randomization by actual treatment received. Safety will be assessed based on occurrence of AEs, occurrence of abuse-related AEs, occurrence of withdrawal AEs, occurrence of opioid-specific side effects, changes in vital signs (heart rate, blood pressure, and respiratory rate only), changes in physical examination findings, changes in clinical laboratory parameters, occurrence of abnormal UDT findings, occurrence of C-SSRS findings, and severity of opioid withdrawal effects as measured by COWS and SOWS.

AEs, serious adverse events (SAEs), and vital signs will also be summarized for each of the Safety Sets Pre-Randomization for each Responder type.

12.1. Extent of Exposure

Extent of exposure to study medication will be evaluated as exposure duration and average daily dose for the index ER opioids and the IR opioids. Extent of exposure measures will only be assessed during the BSODP.

Exposure duration will be calculated in days as (last dose date – first dose date (i.e., randomization date) + 1), regardless of missed doses.

Average daily dose in mg will also be summed from dosing records as follows:

Total daily dose of index ER opioid at every dose level will be multiplied by the number of days on that dose during the BSODP and summed across all doses. Total daily dose of IR opioid during the BSODP will be summed for each day that such medication was taken. Those sums will then be divided by the total number of days of participation in the BSODP to calculate an average daily dose of index ER opioid and IR opioid.

Extent of exposure (duration) and average daily dose will be summarized with descriptive statistics for quantitative assessments by treatment group within each cohort (SR and OR) for each type of formulation (ER/CR and IR).

12.2. Treatment Compliance

Compliance with treatment will be calculated for the index ER opioids only as a percentage based on doses received and expected during the BSODP.

Compliance will be calculated from randomization up to the last dose date (early discontinuation or 24 weeks) in the BSODP by dividing the number of doses of ER opioids taken by the subject by the number of doses required by the protocol and multiplying by 100 to determine the percentage of compliance.

The compliance percentage will be summarized with descriptive statistics for quantitative assessments. The compliance percentage will also be summarized with frequencies and percentages of subjects with compliance in the following categories:

<80%
80-120%
>120%

12.3. Adverse Events

Adverse events (AEs) will be coded using MedDRA. Incidence of AEs will be summarized by primary system organ class (SOC) and by preferred term (PT) within primary SOC.

Treatment emergence within each Pre-Randomization period covered by the Safety Sets and during the BSODP is defined in Section 7.2.3. Each AE will be classified into the given period for which it was treatment emergent according to its start date and time relative to the start and end dates and times of each period. In general, if an AE has an unknown start time but the start date is the same as the start date of a Pre-Randomization period or the BSODP, the AE will be assumed to have started after the start time of the period and therefore will be considered a TEAE in that period. If the AE start time is present and the AE start date is the same as the start date of a period, then the AE will be considered as being a TEAE in the previous period only if the AE start time is before the start time of the period that starts on the same day.

AEs that start outside of these periods will be listed with all AEs. Only TEAEs will be summarized in tables.

For purposes of determining treatment emergence only, partial and missing AE start dates will be imputed according to the rules described here. The partial dates will be displayed in listings. Missing AE start times will not be imputed. But, as stated above, an AE with a missing start time will be considered to be a TEAE in a period which shares the same start date as the start date of the AE. This will also hold true for start dates imputed according to the rules below.

The general strategy for the imputation of dates will be to use the first of the month for a missing day and the first of the year for a missing day and month. If the month or year coincides with the start date of a period, the date will be set to the start date of the latest period the subject entered which shares the same date. If an AE start date is completely missing, the date will be set to the start date of the latest period which the subject entered. Treatment emergence will then be determined according to the rules already given.

Incomplete start dates for AEs will be estimated as follows:

- If the day is missing, then the first day of the month will be used. If the months and years are the same as one or more Pre-Randomization periods or the BSODP, then the start date of the latest period with the same month and year will be used.
- If the day and month are missing, then the first day of the year will be used. If the year is the same as one or more Pre-Randomization periods or the BSODP, then the start date of the latest period with the same year will be used.
- If the date is completely missing, it will be set to the first dose date of double-blind treatment. If a subject does not enter the Post-Randomization Period, then the date will be set to the date that starts the latest period into which the subject entered (i.e., for Suboptimal Responders, the first date of taking the standardized regimen Pre-Randomization; for Optimal Responders, the latest of the first date defining the Pre-Randomization Observation, Taper, and Titration Periods)
- If the estimated AE onset date is after the AE stop date, then the AE stop date will be used as the start date.

An overall summary of TEAEs within the BSODP will show the frequency and percentage of subjects with any TEAE of the specified type within the period:

- Any TEAE
- TEAEs related to study medication
- Fatal TEAEs
- Serious TEAEs
- Serious TEAEs related to study medication
- Any Abuse-related TEAEs
- TEAEs associated with opioid use
- TEAEs associated with withdrawal from opioids

Total counts of each type of TEAE will also be included in the summary without percentages.

Related TEAEs in the overall summary will be those that are assessed by the investigator to be either possibly or probably related. Any TEAE with missing relationship will be considered related.

Potential abuse-related events will be tracked using the Misuse Abuse Diversion Drug Event Reporting System (MADDERS®), which is a clinician-based assessment to identify

and classify such events based on selected AEs and drug accountability discrepancies. A listing of abuse-related TEAEs will include the MADDERS® classification.

TEAEs associated with opioid use are shown in the following table.

Adverse Events Associated with Opioid Use by MedDRA SOC and PT

SOC	PTs
Gastrointestinal Disorders	constipation, nausea, vomiting, dry mouth
Nervous System Disorders	somnolence, dizziness, memory impairment, disturbance in attention
General Disorders and Administrative Site Conditions	fatigue, decreased appetite
Investigations	weight increased
Psychiatric Disorders	sleep disorders
Skin and Subcutaneous Tissues Disorders	pruritus, hyperhydrosis

TEAEs associated with opioid withdrawal will be identified from the following table, where any 3 of the TEAEs occur with start dates within 3 days of each other and encompass at least 2 SOCs. A listing of subjects with TEAEs satisfying the condition will be produced including the details of the TEAEs.

Adverse Events Associated with Opioid Withdrawal by MedDRA SOC and PT

SOC	PTs
Psychiatric disorders	agitation, anxiety, insomnia, restlessness
Eye disorders	lacrimation increased
Respiratory, thoracic and mediastinal disorders	rhinorrhea, yawning
Skin and subcutaneous tissues disorders	piloerection, hyperhidrosis
Nervous system disorders	tremor
General disorders and administration site conditions	hot flush, chills
Musculoskeletal and connective tissue disorders	bone pain, myalgia, muscle twitching
Gastrointestinal disorders	nausea, vomiting, diarrhoea, abdominal pain upper

Summaries of TEAEs within the BSODP will be presented by SOC and by PT within SOC,

showing the frequency and percentage of subjects with TEAEs in each SOC and PT. Should a subject have more than one TEAE in a given SOC or PT, the subject will be counted only once. Tables will be ordered by SOC and PT within SOC according to descending frequencies in the discontinuation arm. The summaries to be produced are listed below:

- All TEAEs by primary SOC and PT
- All serious TEAEs by primary SOC and PT
- All TEAEs by maximum relationship to study medication, primary SOC and PT
- All serious TEAEs by maximum relationship to study medication by primary SOC and PT
- All TEAEs leading to discontinuation of study medication by primary SOC and PT
- All TEAEs by maximum severity, primary SOC, and PT
- Common TEAEs (≥5% in any treatment group) by PT
- Abuse-related TEAEs by primary SOC and PT
- TEAEs associated with opioid use by primary SOC and PT

TEAEs with missing relationship will be counted in the summary of TEAEs by maximum relationship as probably related to study medication. For the summary by maximum severity (mild, moderate, severe), TEAEs with missing severity will be tabulated as a severe TEAE.

Additional summaries of the potential abuse-related TEAEs identified by MADDERS® will be provided. The MADDERS® classifies such events as abuse, misuse, suicide-related, therapeutic error, none of the above, and unknown. The summaries will show the frequencies and percentages of subjects in each Safety Set Pre- and Post-Randomization with at least one TEAE in each classification followed by the TEAEs summarized with frequencies and percentages by MedDRA SOC and PT. The items on the MADDERS® questionnaire will be listed without summary.

In addition to the data listing of AEs, the following subject listings will also be produced for Section 14 of the CSR:

- AE preferred terms and associated investigator's terms (not by subject)
- SAEs
- TEAEs causing discontinuation of study medication
- Fatal TEAEs
- Abuse-related TEAEs from MADDERS®

Listings of all TEAEs for subjects with any potentially clinically significant (PCS) abnormal laboratory value, as defined in Section 12.4, will be produced. Likewise, listings of all TEAEs for subjects with any PCS vital sign value, as defined in Section 12.5.

For the pre-randomization period(s), only a summary of TEAEs and of Serious TEAEs by primary SOC and PT will be produced.

12.4. Laboratory Evaluations

Laboratory evaluations include serum chemistry, hematology, and urinalysis. All chemistry and hematology assessments are numeric, as well as a few urinalysis assessments, and will be classified as low, normal, or high relative to reference ranges.

Numeric laboratory values and changes from baseline will be summarized with descriptive statistics for quantitative assessments. Before summary, the results will be standardized to the International System of Units (SI), or other appropriate common units. If values for a particular assessment cannot be standardized, then results will be summarized separately for each unit.

Qualitative urinalysis results will be summarized with frequencies and percentages of subjects with each recorded value at baseline and at each post baseline visit. Denominators for percentages will be the counts of subjects with a urinalysis value at the visit. Due to the variety of results reported from urinalysis, no shifts or changes from baseline will be summarized.

For quantitative laboratory assessments, shift tables for transitions from baseline to each post baseline visit will be produced for each lab test with baseline low, normal, high, missing, and total across the columns, and likewise down the rows for each visit. The total/total cell will contain the denominator for percentages in the shift tables.

Potentially clinically significant (PCS) high laboratory values will be those that are >2.0

times the upper limit of normal at a post-randomization visit. The frequencies and percentages of subjects with PCS high values will be summarized for subjects with a normal or low value at baseline for the laboratory test.

PCS low laboratory values will be those that are <0.5 times the lower limit of normal at a post-randomization visit. The frequencies and percentages of subjects with PCS low values will be summarized for subjects with a normal or high value at baseline for the laboratory test.

Adverse events associated with PCS high and low laboratory results will be listed as mentioned in Section 12.3.

12.5. Vital Signs

Vital signs collected at each visit in this study are systolic and diastolic blood pressures, heart rate, and respiratory rate.

Vital signs values and changes from baseline will be summarized by visit in the BSODP with descriptive statistics. Vital signs will be classified as PCS low and PCS high relative to the ranges given in the table below.

Measurement	PCS Low Reference Value	PCS High Reference Value
Systolic blood pressure	90 mmHg	180 mmHg
Diastolic blood pressure	50 mmHg	100 mmHg
Heart rate	50 beats per minute	110 beats per minute

Frequencies and percentages of subjects with PCS low values will be summarized relative to the frequency of subjects with a value above the PCS low reference value at baseline. Likewise, subjects with PCS high values will be summarized relative to the frequency of subjects with a value below the PCS high reference value at baseline.

Vital signs and changes from Screening will also be summarized with descriptive statistics for the pre-randomization periods using the Safety Sets Pre-Randomization described for each responder type.

For Suboptimal Responders, the vital signs at Screening, at the end of the Run-in Period (SR1), and at the Randomization Visit (SR2) will be summarized with changes from Screening to SR1 and to SR2 using the Safety Set Pre-Randomization.

For Optimal Responders, the vital signs at Screening and at the end of the Observation Period (OR1) will be summarized with changes from Screening to OR1 for the Safety Set Pre-Randomization (Observation Period). Vital signs at Screening, and visits OR2 1 week into the Taper Period and OR3 at the end of the Taper Period will be summarized with

changes from Screening to OR2 and OR3 using the Safety Set Pre-Randomization (Taper Period). Finally, the vital signs at Screening and at the Randomization Visit (OR4) will be summarized with changes from Screening to OR4 using the Safety Set Pre-Randomization (Titration Period).

No PCS tables for abnormal vital signs during the pre-randomization periods will be produced. Adverse events associated with PCS vital sign values will be listed as mentioned in Section 12.3.

12.6. Physical Examination

For each body system, PE results will be summarized with frequencies and percentages of subjects with a shift from normal or missing at baseline to abnormal at a post-randomization visit within the BSODP. The denominator for percentages will be the number of subjects with a normal or missing assessment at baseline for the body system and who have a post-baseline PE. A listing of PE data will also be provided.

12.7. Urine Drug Testing

UDT is done at Screening and at Weeks 4, 12, and 24 of the BSODP. Abnormal results will be listed and summarized by treatment group at each visit. The summary will show the frequencies and percentages of subjects with at least one abnormal result and the frequencies and percentages of subjects with each type of drug detected at the visit.

12.8. Subjective Opiate Withdrawal Scale

The SOWS consists of a series of 16 statements describing how a subject may feel. The subject is supposed to answer on a scale of 0 = not at all to 4 = extremely. The score is the sum of the items and ranges from 0 to 64.

Values and changes from baseline to each visit where collected in the BSODP will be summarized with descriptive statistics.

12.9. Clinical Opiate Withdrawal Scale

The COWS presents a set of symptoms possibly associated with opiate withdrawal. The presence of or severity level of the symptoms are rated as categories which translate to scores for each symptom. The total score is the sum of the individual symptom scores and ranges from 0-44. Scores in the 0-4 range indicate no particular evidence of withdrawal effects. Otherwise, the scores can be interpreted as follows:

5-12 = mild
13-24 = moderate
25-36 = moderately severe

>36 = severe withdrawal

Values and changes from baseline to each visit where collected in the BSODP will be summarized with descriptive statistics.

Shift tables showing changes from baseline to each visit in the categories above also will be presented.

12.10. Columbia-Suicide Severity Rating Scale

Suicidality assessment by the C-SSRS is done at Screening and at Weeks 6 and 24 during the BSODP. The screening assessment period is defined as lifetime. The assessment period for the BSODP visits is the time since the last visit. Summarization will be done at each visit as well as over the entire BSODP.

12.10.1. Suicidal Behavior

Four types of suicidal behavior are defined in the C-SSRS (actual, interrupted, and aborted attempts, and preparatory acts or behavior), as well as an overall suicidal behavior assessment. For each behavior type, the frequency and percentage of subjects with a yes response to each behavior type will be reported at each visit. The denominator for percentages will be the number of subjects with a C-SSRS assessment at the visit. Subjects with a yes response at any BSODP visit will be given a yes response for the behavior type over the full treatment period, which will also be summarized.

The number of actual suicide attempts per subject will be summarized with frequencies and percentages of subjects. The highest lethality per subject will be similarly summarized. Lethalities for all suicide attempts will be summarized with frequencies and percentages of all attempts.

For subjects for whom no suicidal behavior of any kind was indicated at the screening assessments, the frequency and percentage of subjects who demonstrated any type of suicidal behavior as assessed during the treatment period will be reported by treatment group.

12.10.2. Suicidal Ideation

Five types of increasing suicidal ideation are defined in the C-SSRS: (1) death wish, (2) active thoughts, (3) active ideation without intention to act, (4) active ideation with some intent to act without a plan, and (5) active ideation with a specific plan and intent. The incidence of each type of suicidal ideation and of any type of suicidal ideation will be summarized by treatment group at each visit with frequencies and percentages of subjects with an assessment at the visit. The suicidal ideation severity rating will take the numeric value of the most intense ideation type observed (1-5) or

else 0 if no ideation is observed. Baseline suicidal ideation severity rating will be the rating observed at Screening. The rating at baseline and the rating with change from baseline at each visit will be summarized by treatment group with descriptive statistics. The worst rating observed during treatment and change from baseline will also be summarized.

A scale for the intensity of suicidal ideation will be calculated as a sum of the 5 intensity items. A subject with no suicidal ideation during a period will be assigned a score of 0, so an intensity scale of 0 to 25 results. Suicidal ideation intensity and changes from baseline will be summarized by treatment groups similarly to suicidal ideation severity.

For the subjects for whom no suicidal ideation of any kind was identified at the Screening assessment, the frequency and percentage of subjects who demonstrated any type of suicidal ideation and also any type of serious ideation as assessed during the BSODP will be reported by treatment group. Serious suicidal ideation will be defined as an ideation severity rating of 4 or 5 at any time post-randomization.

A subject will be defined to have worsening of suicidal ideation if the greatest severity rating at any post-randomization visit is more severe than the ideation severity rating taken at screening. The frequency and percentage of subjects with worsening ideation will be reported.

12.10.3. Suicidality

The frequency and percentage of subjects reporting any type of behavior or ideation at any time during the BSODP will be reported by treatment group.

13. INTERIM ANALYSES

Not applicable.

14. END OF STUDY STATISTICAL ANALYSIS PLAN MODIFICATIONS

The study was ended earlier than planned. As a result, some of the planned analyses could not be completed due to insufficient sample size. The following is a list of changes for the respective section in the analysis plan. Section 6.4, the analysis population of the per-protocol set will not be used.

Section 6.5, for protocol deviations, summaries will not be provided. Only a listing will be provided.

Section 7.5, for the primary efficacy analysis, analyses by the subgroups sex and age will not be provided.

Section 8.1, for disposition a flow chart showing the number of subjects entering each period and moving onto the next will not be provided. The early discontinuation assessment tool will be listed only. No summaries will be provided.

Section 8.2, the demographic summaries will be provided for the Safety Set Post-Randomization only. Summaries for other populations will not be provided.

Section 8.3, the summary of the baseline characteristics will be provided for the Safety Set Post-Randomization only. Summaries for other populations will not be provided.

Section 8.5, prior medication summaries will be provided for suboptimal responders at pre-randomization and for optimal responders at pre-randomization (observation period). Concomitant medications will be provided at post-randomization for suboptimal and optimal responders.

Section 9, the planned efficacy analyses will not be provided. Summary statistics will be provided for the average PI score and change from baseline at each visit for suboptimal and optimal responders. Summary statistics will also be provided for the average PID cumulative response for percent improvement in suboptimal responders and percent worsening in optimal responders at Weeks 12 and 24. Summary statistics will be provided for observed values and changes from baseline for the MOS Sleep Scale, PHQ-8, EuroQOL 5 Dimensions, Digit Symbol Substitution Test, and the sexual function for suboptimal and optimal responders at each visit. Frequencies and percentages will be provided for the Patient Global Impression of Change at each visit for suboptimal and optimal responders. A listing will be provided for the quantitative sensory testing.

Section 12.3, the following adverse event summaries will be provided for suboptimal responders:

- Overall summary of TEAEs at pre-randomization

- Overall summary of TEAEs at post-randomization
- TEAEs by system organ class and preferred term at pre-randomization
- TEAEs by system organ class and preferred term at post-randomization
- TEAEs by maximum relationship, system organ class, and preferred term at post-randomization
- TEAEs by maximum severity, system organ class, and preferred term at post-randomization
- TEAEs associated with opioid withdrawal by preferred term at post-randomization
- Potential abuse-related TEAEs by MADDERS Classification at post-randomization

The following adverse event summaries will be provided for optimal responders:

- Overall summary of TEAEs at pre-randomization (observation period)
- Overall summary of TEAEs at pre-randomization (taper period)
- Overall summary of TEAEs at pre-randomization (titration period)
- Overall summary of TEAEs at post-randomization
- TEAEs by system organ class and preferred term at pre-randomization (observation period)
- TEAEs by system organ class and preferred term at pre-randomization (taper period)
- TEAEs by system organ class and preferred term at pre-randomization (titration period)
- TEAEs by system organ class and preferred term at post-randomization
- TEAEs by maximum relationship, system organ class, and preferred term at post-randomization
- TEAEs by maximum severity, system organ class, and preferred term at post-randomization
- TEAEs associated with opioid withdrawal by preferred term at post-randomization
- Potential abuse-related TEAEs by MADDERS Classification at post-randomization

AE listings will be provided for:

- All adverse events
- Abuse-related events
- Abuse-related adverse events from MADDERS
- Serious adverse events
- Adverse events leading to study discontinuation
- Adverse events with potentially clinically significant abnormal laboratory values

- Adverse events with potentially clinically significant abnormal vital sign values

Section 12.4, for laboratory summaries, summary and shift tables will be provided for hematology and chemistry results. No summaries will be provided for urinalysis results. In addition frequencies of potentially clinically significant results and a listing of treatment-emergent abnormal laboratory results will be presented.

Section 12.5, for vital signs actual results, changes from baseline, and potentially clinically significant abnormalities will be summarized at post-randomization only.

Section 12.6, abnormal physical examination results will be listed. No other summaries will be provided.

Section 12.9, for the COWS only summary tables for observed values and changes from baseline will be presented. Shift tables will not be provided.

15. REFERENCE LIST

Little RJA, Rubin DB. Statistical analysis with missing data (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc., 2002.

Spritzer KL, Hays RD. MOS Sleep Scale: A Manual for Use and Scoring, Version 1.0. Los Angeles, CA: November 2003.

16. PROGRAMMING CONSIDERATIONS

All tables and listings will be generated using SAS® for Windows, Release 9.3 or higher (SAS® Institute Inc., Cary, NC, USA). Computer-generated table and listing output will adhere to the following specifications.

16.1. General Considerations

- A separate SAS program will be created for each output.
- Each output will be stored in a separate file.
- Output files will be delivered in RTF format.
- Numbering of tables and listings will follow International Conference on Harmonization (ICH) E3 guidance.

16.2. Table and Listing Format

16.2.1. General

- All tables and listings will be produced in landscape format.
- All tables and listings will be produced using the Courier New font, size 9.
- The data displays for all tables and listings will have minimum 1-inch margins.
- Tables and listings will be in black and white (no color), unless otherwise specified.
- Specialized text styles, such as bolding, italics, borders, shading, superscripted and subscripted text, will not be used in the tables and listings unless otherwise specified. On some occasions, superscripts 1, 2, and 3 may be used (see below).
- Only standard keyboard characters will be used in the tables and listings. Special characters, such as non-printable control characters, printer-specific, or font-specific characters, will not be used. Hexadecimal-derived characters will be use if they are appropriate to help display math symbols. Certain superscripts and subscripts (e.g., kg/m^2) will be employed on a case-by-case basis.
- Mixed case will be used for all titles, footnotes, and column headers, as appropriate.

16.2.2. Headers

- All output should have the following header at the top left of each page:
  Opioid PMR Consortium
  Protocol No. 2065-5

- All output should have Page n of N at the top right corner of each page. Tables and listings should be internally paginated in relation to the total length (i.e., the page number should appear sequentially as page n of N, where N is the total number of pages in the table).
- The date (date output was generated) should appear along with the program name in the footer on each page.

16.2.3. Display Titles

- ICH E3 numbering will be used. A decimal system (x.y and x.y.z) will be used to identify tables and listings with related contents. The title will be centered. The analysis set will be identified on the line immediately following the title. The title and table designation are single spaced. A solid line spanning the margins will separate the display titles from the column headers. There will be 1 blank line between the last title and the solid line.

Table x.y.z
First Line of Title
Second Line of Title if needed
Analysis Set

16.2.4. Column Headers

- Column headings should be immediately below the solid line described above.
- For numeric variables, include “unit” in the column or row heading when appropriate.
- Analysis set sizes for each treatment will be presented in the column heading as N=xxx (or row heading where appropriate). This is distinct from the ‘n’ used for the descriptive statistics, where n represents the number of subjects with non-missing data.

16.2.5. Body of Data Display

16.2.5.1. General Conventions

Data in columns of a table or listing should be formatted as follows:

- alphanumeric values are left justified;
- whole numbers (e.g., counts) are right justified; and
- numbers containing fractional portions are decimal aligned.

16.2.5.2. Table Conventions

- Units will be included where available.
- If the categories of a parameter are ordered, then all categories between the minimum and the maximum category should be presented in the table, even if n=0 for a given category. For example, the frequency distribution for symptom severity would appear as:

Severity Rating	n
severe	0
moderate	8
mild	3

- Unless otherwise specified, the mean, median, first quartile, and third quartile for a set of values will be printed out to 1 more significant digit than the original values, and standard deviations should be printed out to 2 more significant digits than the original values. The minimum and maximum should report the same significant digits as the original values. For example, for systolic blood pressure:

n	XX
Mean	XXX.X
Std Dev	X.XX
Q1	XXX.X
Median	XXX.X
Q3	XXX.X
Minimum	XXX
Maximum	XXX

- P-values should be output in the format: “0.xxx”, where xxx is the value rounded to 3 decimal places. Any p-value less than 0.001 will be presented as < 0.001. If the p-value is returned is greater than 0.999, then the value will be presented as >0.999.
- Percentage values will be printed to one decimal place in parentheses. Percentages for zero counts will not be displayed. Percentages equal to 100% will be displayed as (100).
- The percentage of subjects will normally be calculated as a proportion of the number of subjects in the analysis set, unless otherwise specified in the footnote.
- For categorical summaries (number and percentage of subjects) where a subject can be included in more than one category, a footnote or programming note will describe if the subject should be included in the summary statistics for all relevant categories or just 1 category and the criteria for selecting the category.
- Where a category with a subheading (such as system organ class) has to be split over more than one page, the output heading followed by “(cont)” will be displayed at the top of each subsequent page. The overall summary statistics for the subheading should only be output on the first relevant page.

16.2.5.3. Listing Conventions

- Listings will be sorted for presentation in order treatment, subject number, visit/collection date, and visit/collection time.
- Dates should be printed in SAS® DATE9 format (ddmonyyyy or 01JUL2000). Missing portions of dates will be represented on subject listings as dashes (--JUL2000),
- All observed time values will be presented using a 24-hour clock format (HH:MM or HH:MM:SS). Time will only be reported if it was measured as part of the study.
- Units will be included where available.

16.2.5.4. Footnotes

- A solid line spanning the margins will separate the body of the data display from the footnotes. All footnotes will be left justified with single-line spacing immediately below the solid line underneath the data display.
- Footnotes will begin with “Note:” if it is an informational footnote or a, b, c, etc. if it is a reference footnote. Each footnote will start on a new line, where possible.
- Footnotes will be used sparingly and must add value to the table or listing. If more than six lines of footnotes are planned, then a cover page may be used to display the footnotes, and only those essential to the comprehension of the data will be repeated on each page.
- The last lines of the footnote will be a standard source line that indicates the name of the program used to produce the data display, the date and time the program was run and the source (either listing or eCRF page).

17. QUALITY CONTROL

SAS programs are developed to produce clinical trial output such as analysis data sets, summary tables, data listings, figures or statistical analyses. INC Research SOP 03.010.00 and 03.013.00 provide an overview of the development of such SAS programs.

INC Research SOP 03.009.00 describes the quality control procedures that are performed for all SAS programs and output. Quality control is defined here as the operational techniques and activities undertaken to verify that the SAS programs produce the proper clinical trial output by checking for their logic, efficiency and commenting and by review of the produced output.

18. INDEX OF TABLES

Number	Name
Table 14.1.1.1	Subject Disposition for Suboptimal and Optimal Responders – All Subjects
Table 14.1.1.1.1	Subject Disposition by Period – Pre-randomization – Suboptimal Responders
Table 14.1.1.1.2	Subject Disposition by Period – Pre-randomization – Optimal Responders
Table 14.1.1.2.1	Subject Disposition – Post-randomization – Suboptimal Responders
Table 14.1.1.2.2	Subject Disposition – Follow-up Period – Suboptimal Responders
Table 14.1.1.3.1	Subject Disposition – Post-randomization – Optimal Responders
Table 14.1.1.3.2	Subject Disposition – Follow-up Period – Optimal Responders
Table 14.1.2.1	Demographics and Baseline Characteristics – Safety Set Post-Randomization – Suboptimal Responders
Table 14.1.2.2	Demographics and Baseline Characteristics – Safety Set Post-Randomization – Optimal Responders
Table 14.1.3.1	Baseline Disease Characteristics – Safety Set Post-Randomization – Suboptimal Responders
Table 14.1.3.2	Baseline Disease Characteristics – Safety Set Post-Randomization – Optimal Responders
Table 14.1.4.1	Summary of Medical History – Safety Set Post-Randomization – Suboptimal Responders
Table 14.1.4.2	Summary of Medical History – Safety Set Post-Randomization – Optimal Responders
Table 14.1.5.1.1	Summary of Prior Medications Prior to Enrollment – Safety Set Pre-Randomization – Suboptimal Responders
Table 14.1.5.1.2	Summary of Concomitant Medications Post-Randomization – Safety Set Post-Randomization – Suboptimal Responders
Table 14.1.5.2.1	Summary of Prior Medication Prior to Enrollment – Safety Set Pre-Randomization (Observation Period) – Optimal Responders
Table 14.1.5.2.2	Summary of Concomitant Medications Post-Randomization – Safety Set Post-Randomization – Optimal Responders
Table 14.1.6.1	Extent of Exposure and Treatment Compliance – Safety Set Post-Randomization – Suboptimal Responders
Table 14.1.6.2	Extent of Exposure and Treatment Compliance – Safety Set Post-Randomization – Optimal Responders
Table 14.2.1.1.1	Mean Average PI Score: Actual and Change from Baseline Descriptive Summary – Observed Data – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.1.1.2	Mean Average PID: Cumulative Response in Percent Improvement at Weeks 12 and 24 – Observed Data – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.1.2.1	Mean Average PI Score: Actual and Change from Baseline Descriptive Summary – Observed Data – Intent-to-Treat Set – Optimal Responders
Table 14.2.1.2.2	Mean Average PID: Cumulative Response in Percent Improvement at Weeks 12 and 24 – Observed Data – Intent-to-Treat Set – Optimal Responders

Statistical Analysis Plan

Number	Name
Table 14.2.2.1	Medical Outcomes Study Sleep Scale: Change from Baseline Descriptive Summary – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.2.2	Medical Outcomes Study Sleep Scale: Change from Baseline Descriptive Summary – Intent-to-Treat Set – Optimal Responders
Table 14.2.3.1	Patient Health Questionnaire Depression Scale (PHQ-8): Actual and Change from Baseline Descriptive Summary - Intent-to-Treat Set – Suboptimal Responders
Table 14.2.3.2	Patient Health Questionnaire Depression Scale: Actual and Change from Baseline Descriptive Summary - Intent-to-Treat Set – Optimal Responders
Table 14.2.4.1	EuroQOL 5 Dimensions Instrument: Descriptive Summary – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.4.2	EuroQOL 5 Dimensions Instrument: Descriptive Summary – Intent-to-Treat Set – Optimal Responders
Table 14.2.5.1	Patient Global Impression of Change: Frequencies and Percentages of Subjects – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.5.2	Patient Global Impression of Change: Frequencies and Percentages of Subjects – Intent-to-Treat Set – Optimal Responders
Table 14.2.6.1	Digit Symbol Substitution Test: Actual and Change from Baseline Descriptive Summary – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.6.2	Digit Symbol Substitution Test: Actual and Change from Baseline Descriptive Summary – Intent-to-Treat Set – Optimal Responders
Table 14.2.7.1	Sexual Function: Actual and Change from Baseline Descriptive Summary – Intent-to-Treat Set – Suboptimal Responders
Table 14.2.7.2	Sexual Function: Actual and Change from Baseline Descriptive Summary – Intent-to-Treat Set – Optimal Responders
Table 14.3.1.1.1	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Pre-Randomization – Suboptimal Responders
Table 14.3.1.1.2	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.1.3	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Pre-Randomization – Suboptimal Responders
Table 14.3.1.1.4	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.1.5	Treatment-Emergent Adverse Events (TEAEs) by Maximum Relationship, System Organ Class and Preferred Term – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.1.6	Treatment-Emergent Adverse Events (TEAEs) by Maximum Severity, System Organ Class and Preferred Term – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.1.7	Treatment-Emergent Adverse Events (TEAEs) Associated with Opioid Withdrawal – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.1.8	Potential Abuse-related Treatment-Emergent Adverse Events (TEAEs) by MADDERS Classification – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.1.2.1	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Pre-Randomization (Observation Period) – Optimal Responders

Statistical Analysis Plan

Number	Name
Table 14.3.1.2.2	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Pre-Randomization (Taper Period) – Optimal Responders
Table 14.3.1.2.3	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Pre-Randomization (Titration Period) – Optimal Responders
Table 14.3.1.2.4	Overall Summary of Treatment-Emergent Adverse Events (TEAEs) – Safety Set Post-Randomization – Optimal Responders
Table 14.3.1.2.5	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Pre-Randomization (Observation Period) – Optimal Responders
Table 14.3.1.2.6	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Pre-Randomization (Taper Period) – Optimal Responders
Table 14.3.1.2.7	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Pre-Randomization (Titration Period) – Optimal Responders
Table 14.3.1.2.8	Treatment-Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term – Safety Set Post-Randomization – Optimal Responders
Table 14.3.1.2.9	Treatment-Emergent Adverse Events (TEAEs) by Maximum Relationship, System Organ Class and Preferred Term – Safety Set Post-Randomization – Optimal Responders
Table 14.3.1.2.10	Treatment-Emergent Adverse Events (TEAEs) by Maximum Severity, System Organ Class and Preferred Term – Safety Set Post-Randomization – Optimal Responders
Table 14.3.1.2.11	Treatment-Emergent Adverse Events (TEAEs) Associated with Opioid Withdrawal – Safety Set Post-Randomization – Optimal Responders
Table 14.3.1.2.12	Potential Abuse-related Treatment-Emergent Adverse Events (TEAEs) by MADDERS Classification – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.1.1	Laboratory Test: Chemistry – Actual and Change from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.1.2	Laboratory Test: Chemistry – Shift from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.1.3	Laboratory Test: Hematology – Actual and Change from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.1.4	Laboratory Test: Hematology – Shift from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.1.5	Laboratory Test: Frequencies and Percentages of Potentially Clinically Significant Results – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.1.6	Listing of Treatment-Emergent Abnormal Laboratory Results – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.2.1	Laboratory Test: Chemistry – Actual and Change from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.2.2	Laboratory Test: Chemistry – Shift from Baseline – Safety Set Post-Randomization – Optimal Responders

Statistical Analysis Plan

Number	Name
Table 14.3.4.2.3	Laboratory Test: Hematology – Actual and Change from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.2.4	Laboratory Test: Hematology – Shift from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.2.5	Laboratory Test: Frequencies and Percentages of Potentially Clinically Significant Results – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.2.6	Listing of Treatment-Emergent Abnormal Laboratory Results – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.3.1	Urine Drug Testing – Summary of Abnormalities – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.3.2	Urine Drug Testing – Summary of Abnormalities – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.4.1.1	Vital Signs – Actual and Change from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.4.1.2	Vital Signs: Potentially Clinically Significant Abnormalities – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.4.2.1	Vital Signs – Actual and Change from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.4.2.2	Vital Signs: Potentially Clinically Significant Abnormalities – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.5.1	Subjective Opiate Withdrawal Scale (SOWS) – Actual and Change from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.5.2	Subjective Opiate Withdrawal Scale (SOWS) – Actual and Change from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.6.1	Clinical Opiate Withdrawal Scale (COWS): Actual and Change from Baseline – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.6.2	Clinical Opiate Withdrawal Scale (COWS): Actual and Change from Baseline – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.7.1.1	Columbia-Suicide Severity Rating Scale: Suicidal Behavior Descriptive Summary – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.7.1.2	Columbia-Suicide Severity Rating Scale: Suicidal Ideation and Suicidality Descriptive Summary – Safety Set Post-Randomization – Suboptimal Responders
Table 14.3.4.7.2.1	Columbia-Suicide Severity Rating Scale: Suicidal Behavior Descriptive Summary – Safety Set Post-Randomization – Optimal Responders
Table 14.3.4.7.2.2	Columbia-Suicide Severity Rating Scale: Suicidal Ideation and Suicidality Descriptive Summary – Safety Set Post-Randomization – Optimal Responders

19. INDEX OF LISTINGS

Number	Name
Listing 16.2.1	Subject Disposition
Listing 16.2.2	Protocol Deviations
Listing 16.2.3	Eligibility Criteria
Listing 16.2.4.1	Demographic and Baseline Characteristics
Listing 16.2.4.2	Medical History
Listing 16.2.4.3	Pre-Enrollment Prior Opioid Medications
Listing 16.2.4.4	Pre-Enrollment Prior Non-opioid Medications
Listing 16.2.4.5	Prior and Concomitant Medications
Listing 16.2.4.6	Prior and Concomitant Medications ATC Level and Associated Verbatim Terms
Listing 16.2.5.1	Study Drug Administration
Listing 16.2.5.2	MADDERS Supplemental Drug Accountability
Listing 16.2.6.1	Pain Intensity on NRS
Listing 16.2.6.2	Endocrine Laboratory Tests and Sexual Function Questionnaire
Listing 16.2.6.3	Quantitative Sensory Testing (QST)
Listing 16.2.6.4	Early Discontinuation Assessment Tool
Listing 16.2.7.1	Adverse Events
Listing 16.2.7.2	Abuse-related Events
Listing 16.2.7.3	Abuse-related Events from MADDERS Supplemental
Listing 16.2.7.4	Serious Adverse Events
Listing 16.2.7.5	Adverse Events Leading to Study Discontinuation
Listing 16.2.7.6	Adverse Events for Subjects with Potentially Clinically Significant (PCS) Abnormal Laboratory Values
Listing 16.2.7.7	Adverse Events for Subjects with Potentially Clinically Significant (PCS) Abnormal Vital Sign Values
Listing 16.2.7.8	Adverse Events Preferred Terms and Associated Verbatim Terms
Listing 16.2.8.1	Laboratory Test – Chemistry
Listing 16.2.8.2	Laboratory Test – Hematology
Listing 16.2.8.3	Laboratory Test – Urinalysis
Listing 16.2.8.4	Laboratory Test – Urine drug testing (UDT)
Listing 16.2.8.5	Laboratory Test – Pregnancy Test
Listing 16.2.8.6	Vital Signs
Listing 16.2.8.7	Subjective Opioid Withdrawal Scale (SOWS)
Listing 16.2.8.8	Clinical Opioid Withdrawal Scale (COWS)
Listing 16.2.8.9	Columbia-Suicide Severity Rating Scale (C-SSRS)
Listing 16.2.8.10	12-Lead ECG
Listing 16.2.8.11	Abnormal Physical Examination Results