STATISTICAL ANALYSIS PLAN

Protocol Number: STUDY00002427
NCT04637802

Effectiveness of an mHealth psychosocial intervention to prevent transition from acute to chronic postsurgical pain in adolescents (SurgeryPal)

Trial Lead Statistician:
Chuan Zhou, PhD
CHBD, Seattle Children's Research Institute

Principal Investigator(s):
Jennifer Rabbitts, MD
Seattle Children's Hospital/University of Washington

Tonya Palermo, PhD
Seattle Children's Research Institute/University of Washington

Funded by: National Institute of Child Health and Human Development

Version Number: Final 3.0

Version Date: 6/28/2024

(Based on Protocol Dated: 17 October 2023)

Summary of Changes from Previous Version:

Statistical Analysis Plan	July 27, 2020
Revised Statistical Analysis Plan	June 28, 2024
Protocol: SurgeryPal	17 October 2023

Description	Changes to the Statistical Analysis Plan were made to align with the protocol version 17 October 2024 and reduction of target sample size to at least 407

Section	Protocol Summary
Original Text	500
Revised Text	At least 407
Description	In paragraphs 1 and 3

Section	Protocol Summary
Original Text	Aim 3. Determine the combined effects of pre- and post-operative cognitive behavioral therapy (CBT) on chronic pain outcomes
Revised Text	Aim 3. Determine the combined effects of pre- and post-operative psychosocial intervention on chronic pain outcomes
Description	Removal of CBT language

Section	Power/sample size calculations for Aims 1-3.
Original Text	N = 100, 200, or 300 removed Comparison Groups
Revised Text	None – text removed
Description	No longer applicable

Section	Power/sample size calculations for Aims 1-3 (text under table)
Original Text	Five-hundred dyads will be randomized to 4 treatment arms at a 1:1:1:1 ratio (see treatment arms table below), therefore we anticipate 125 youth per study group.
Revised Text	*The number of treatment arms in the smallest comparison group differ for each aim – text removed
Description	No longer applicable

Section	4.0. Sample Size Determination for the Primary Efficacy Analysis
Original Text	Although we have had >95% retention in our prior perioperative and mHealth intervention studies,

	we conservatively plan for 20% attrition, and therefore plan to enroll 500 youth for a complete sample of 400 youth, which will allow us to test the primary study hypotheses with optimal power.
Revised Text	Although we have had >95% retention in our prior perioperative and mHealth intervention studies, we conservatively plan for 20% attrition. Based on this, our original plan was to enroll 500 youth for a complete sample of 400 youth, to test the primary study hypotheses with optimal power. After further assessment of attrition rates throughout the study, we have revised our target sample size to at least 407 based on observed attrition rates of 4% at T2, 6% at T3, and 8% at T4. Enrollment will continue irrespective of sample size through June 30, 2024. This will provide sufficient power for all aims (including exploratory aims)

Section	4.0. Sample Size Determination for the Primary Efficacy Analysis
Original Text	N/A – new text
Revised Text	The efficacy primary outcome for Aim 1 relies on T2 data and requires a total of n=124 with 90% power. The efficacy primary outcome for Aim 2 and Aim 3 requires a total of n=184 with 90% power. Our exploratory efficacy aim requires a total of n=344-452 with 80-90% power. With an enrollment end date of 6/30/24 and using the same attrition rates as have been observed in the trial, we project ending the trial with n = 389 completed T2 (post surgery), n = 383 completed T3, which exceeds the n required for 90% power on the primary efficacy analyses. We project n = 374 for the exploratory aim, which may be analyzed separately after the NCE year.

Section	6.0 Randomization
Original Text	Five-hundred dyads will be randomized to 4 treatment arms at a 1:1:1:1 ratio (see treatment arms table below), therefore we anticipate 125 youth per study group.
Revised Text	At least 407 dyads (youth and parent/guardian) will be randomized to 4 treatment arms at a 1:1:1:1 ratio (see treatment arms table below), therefore we anticipate up to 125 youth per study group
Description	Clarification

Statistical Analysis Plan	July 27, 2020

Revised Statistical Analysis Plan	June 29, 2021
Protocol: SurgeryPal	20 January 2021

Description	Changes to the Statistical Analysis Plan were made to align with the protocol version 20 January 2021

Section	Protocol Summary
Original Text	(4 week duration between weeks 2 and 6 following surgery)
Revised Text	(4 week duration starting 3 weeks following surgery)
Description	Change to be consistent with protocol

Section	Protocol Summary
Original Text	Aim 3. Determine the combined effects of pre- and post-operative cognitive behavioral therapy (CBT) on chronic pain outcomes
Revised Text	Aim 3. Determine the combined effects of pre- and post-operative psychosocial intervention on chronic pain outcomes
Description	Removal of CBT language

Section	Introduction
Original Text	selected data fields off or primary and selected secondary outcome measures
Revised Text	Selected data fields for primary and selected secondary outcome measures
Description	Typo correction

Section	Study Endpoints
Original Text	For Aim 1 we will examine between group differences in: a) individual level time-averaged BPI scores and b) individual rate of recovery based on BPI trajectories.
Revised Text	and individual rate of recovery will be computed for each participant over 14 days to yield the primary acute pain outcome measure for Aim 1.
Description	Original text deleted and replaced

Section	Study Endpoint
Original Text	As well as non-opioid and opioid analgesic use
Revised Text	
Description	Text deleted

Section	Study Endpoint
Original Text	T4
Revised Text	
Description	Deleted from page 13 chart: determine effectiveness of perioperative psychosocial intervention to reduce pain catastrophizing

Section	Study Endpoint
Original Text	T2, T3
Revised Text	T2 (PT1), T3 (PT2)
Description	Page 15

Section	Study Endpoints
Original Text	Post-operative (week 2 to 6 post surgery)
Revised Text	Post-operative (starting week 3 post-surgery)
Description	Page 15, page 16 for protocol alignment

Section	Study Endpoint
Original Text	parent education, and youth’s pain history
Revised Text	parent education, youth’s pain history
Description	Page 16: grammar correction

Section	Exploratory (Tertiary Efficacy Endpoint)
Original Text	
Revised Text	Added text explaining efficacy endpoint
Description	Page 19 and 20

Section	Sample Size Determination for the Primary Efficacy Analysis
Original Text	
Revised Text	Power/Sample size calculations for Aims 1-3
Description	Language added page 21

Section	Sample Size Determination for the Primary Efficacy Analysis
Original Text	For Aim 1, the first primary comparison will be time-averaged difference (TAD) between control and intervention arms using individual daily BPI assessments at T2. The power and sample size calculation is based on GEE model on the daily repeated measures of BPI scores and the following assumptions. We assume 1:1 allocation of subjects into control and intervention arms, a total of 14 days assessments per individual, within-subject observations follow AR(1) correlation pattern with a correlation coefficient of 0.7, a missing completely at random rate of 5% each day, when standard deviation is 2.2, N=177 per arm will provide 80% power to detect a between group difference of 1.4; when SD is 2.0, N=199 per arm will provide 80% power to detect a between group difference of 1.2.
Revised Text	Deleted
Description	Protocol alignment

Section	Sample Size Determination for the Primary Efficacy Analysis
Original Text	For Aim 1, the second primary parameter of interest

Revised Text	For Aim 1, the primary parameter of interest
Description	Deleted second primary

Section	Sample Size Determination for the Primary Efficacy Analysis
Original Text	
Revised Text	Based on our power calculations (provided in above table), a final sample size of 400 youth (100 per arm) will allow sufficient sample size in each comparison group to provide ample power to test each hypothesis on primary outcomes (Aims 1-3). For exploratory outcomes (Aim 4), we anticipate a rate of CPSP of 31% (see preliminary studies). With a sample size of 400 (i.e. 100 in smallest comparison group), we will have >80% power to detect an absolute reduction of 15% in the rate of CPSP in the group receiving combined full pre- and post-operative intervention.
Description	Adding power calculations

Section	Study Population
Original Text	Teen:
Revised Text	Seeing counselor for 4+ sessions specifically to prepare for surgery
Description	Update to inclusion/exclusion criteria

Section	Demographic and Baseline Characteristics
Original Text	age, sex, race
Revised Text	Age, sex, gender, race
Description	Adding gender to the demographics

Section	Randomization
Original Text	selected from the list (2, 4, 6, or 8)
Revised Text	selected from the list (4 or 8)
Description	Protocol alignment

Section	Randomization
Original Text	
Revised Text	Intersex study participants will be randomized as males
Description	Definition of randomization

Section	Primary Efficacy Analyses
Original Text	Adding Primary and Secondary analytic plan for Hypothesis 1
Revised Text	Removing Primary Outcome
Description	Protocol alignment

SIGNATURE PAGE AND APPROVALS

I acknowledge that I have reviewed and approved the document referenced above.

Chuan Zhou, PhD Trial Lead Statistician Seattle Children's Research Institute	Printed Name Signature and Date
Gayane Yenokyan, PhD TIN Lead Statistician Johns Hopkins-Tufts Trial Innovation Center HEAL Pain ERN Statistical Core	Printed Name Signature and Date
Larry Cook, PhD TIN Statistician Utah Trial Innovation Center HEAL Pain ERN Data Coordinating Center	Printed Name Signature and Date
Jennifer Rabbitts, MD Principal Investigator Seattle Children's Hospital	Printed Name Signature and Date

Table of Contents

1.0. Protocol Summary 1
2.0. Introduction 2
3.0. Study Endpoints 3
3.1. Efficacy - Primary Outcomes 3
3.2. Secondary Outcomes 5
3.3. Safety Measures 11
3.3.1. Safety Measures Summary 11
3.3.2. Safety Monitoring and Recruitment Suspension Rules 12
3.4. Exploratory (Tertiary) Efficacy Endpoint 13
4.0. Sample Size Determination for the Primary Efficacy Analysis 15
5.0. Study Population 16
5.1. Final Manuscript population 17
5.2. Safety Analysis Sample 17
5.3. Demographic and Baseline Characteristics 17
6.0. Randomization 18
7.0. Blinding 18
8.0. Multiplicity 18
9.0. Missing Data 18
10.0. Outlier Measures 20
11.0. Exploratory Analyses 20
12.0. Primary Efficacy Analyses 20
12.1. Primary and secondary analytic plan for Hypothesis 1 20
12.2. Primary analytic plan for Hypothesis 2 Primary Outcomes 21
12.3 Primary analytic plan for Hypothesis 3 Primary Outcomes 22
13.0. Secondary Efficacy Outcomes Analyses 22
13.1. Primary analytic plan for Hypothesis 1 Secondary Acute Outcomes 22
13.2. Primary analytic plan for Hypothesis 3 Secondary Chronic Outcomes. 22
14.0. Safety Analyses 23
15.0. Tertiary (Exploratory/Sensitivity) Analyses 23
15.1 Analytic plan for Hypothesis 4 (exploratory) 23
15.2 Analytic plan for Hypothesis 5 (exploratory) 24
15.3 Analytic plan for Hypothesis 6 (exploratory) 24
15.4 Analyses to test for differences in effect among sex, racial, and/or ethnic groups 25
15.5 Secondary data analysis 25
16.0. References 26
17.0. Appendix 1: Variable List 28

List of Abbreviations:

AE	Adverse Event
ANOVA	Analysis of Variance
AR	AutoRegressive
BPI	Brief Pain Inventory
CBT	Cognitive Behavioral Therapy
CONSORT	Consolidated Standards of Reporting Trials
CPSP	Chronic Post-Surgical Pain
CRF	Case Report Form
DCC	Data Coordinating Center
DSMB	Data and Safety Monitoring Board
DRE	Disease-Related Event
GAD-2	Generalized Anxiety Disorder 2-item
GEE	Generalized Estimating Equation
HRQL	Health-Related Quality of Life
ITT	Intention-To-Treat
MCAR	Missing Completely At Random
MCID	Minimal Clinically Important Difference
MICE	Multiple Imputations with Changed Equations
PCS	Pain Catastrophizing Scale – Child Version
PedsQL	Pediatric Quality of Life Inventory - Short form, acute version
PGI-S	Patient Global Impression of Severity Scale
PHQ	Patient Health Questionnaire
PI	Principal Investigator
PROMIS	Patient-Reported Outcomes Measurement Information System
RCT	Randomized Controlled Trial
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCH	Seattle Children’s Hospital
SD	Standard Deviation
SE	Standard Error
SMC	Safety Monitoring Committee
SOA	Schedule of Activities
TAD	Time-Averaged Difference
TEI-SF	Treatment Evaluation Inventory-Short Form
UP	Unanticipated Problem
US	United States

1.0. Protocol Summary

This is a randomized controlled trial to test effectiveness of the SurgeryPal intervention vs. education control to improve acute and chronic pain and health outcomes in at least 407 youth undergoing major musculoskeletal surgery. Youth will be randomized on an individual level using a factorial design to SurgeryPal or Education during 2 phases of intervention: 1) pre-operative phase (4 week duration delivered over the 4 weeks leading up to surgery), and 2) post-operative phase (4 week duration starting 3 weeks following surgery). Thus there will be 4 treatment arms. Participants will undergo 4 assessments, independent of their treatment assignment: T1: Baseline (pre-randomization); T2: acute post-surgery outcomes (daily assessment of acute outcomes beginning day 1 through day 14 after hospital discharge from surgery); T3: Post-surgery follow-up (assessment of chronic outcomes at 3-months post-surgery); T4: Final post-surgery follow-up (assessment of outcomes at 6-months post-surgery).

The SurgeryPal intervention is a cognitive-behavioral program for youth undergoing major surgery. SurgeryPal is delivered in two 4-week phases: the pre-operative phase, and the post-operative phase. Each phase consists of 3 separate treatment modules for youth and parents, which teach cognitive and behavioral coping skills to target known psychosocial and behavioral risk factors for poorer postsurgical outcomes (i.e. anxiety/distress and sleep), and provide training in pain self-management skills. Pre-operative modules include: 1) Preparing for surgery, 2) Coping with stress before surgery, and 3) Getting ready for the hospital. Postoperative modules include: 1) Coping at home after surgery, 2) Return to activities and school, and 3) Long-term recovery. Modules are self-guided and interactive, and focus on preparation for and recovery from surgery. The goal of the program is to improve psychosocial functioning, reduce pain, and enhance quality of life following surgery in youth.

The proposed study will enroll at least 407 adolescents ages 12-18 who are scheduled to undergo spinal fusion surgery for idiopathic/juvenile scoliosis, spondylolisthesis or kyphosis plus one parent/caregiver participant for each adolescent. Adolescents will be recruited from approximately 25 pediatric centers across the United States. Families who do not speak or understand English, or do not have access to a smart device will be excluded. Adolescents who meet any of the following criteria will be excluded: severe learning disability/cognitive impairment/intellectual delay, diagnosed unstable psychiatric condition, severe/systemic disease (e.g. neuromuscular scoliosis, cancer), taking medication daily for a medical condition (except allergies/asthma/anxiety/depression), have had a prior major surgery, or have a diagnosed chronic musculoskeletal pain condition (see 5.1 and 5.2 for more detail).

Aim 1. Determine effectiveness of a pre-operative mHealth psychosocial intervention to improve acute pain outcomes in adolescents undergoing major musculoskeletal surgery.

Aim 2. Determine effectiveness of a post-operative mHealth psychosocial intervention to improve chronic pain outcomes at 3 months in adolescents undergoing major musculoskeletal surgery.

Aim 3. Determine the combined effects of pre- and post-operative psychosocial intervention on chronic pain outcomes.

Aim 4. Explore mechanisms of an mHealth psychosocial intervention to prevent Chronic Post-Surgical Pain (CPSP).

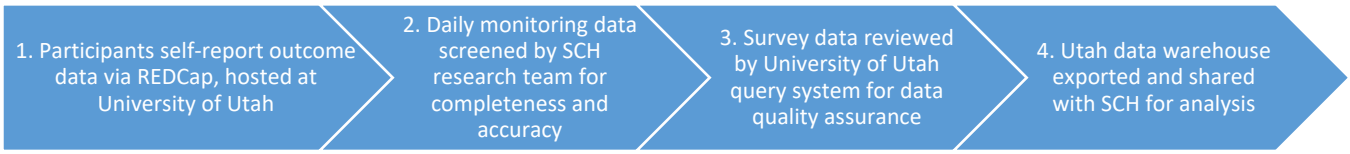
2.0. Introduction

This plan provides the details of statistical considerations, analyses, and reports planned for the SurgeryPal study, including proposed models, sample size estimates, power considerations, interim analyses, and safety analyses. In addition, this plan will define the analysis populations and specify which population will be used for the primary analyses. It will define subgroup analyses, define treatment arms, and address other statistical issues relevant to the analyses (e.g., sample data to be used, missing data, adjustments for multiplicity, etc.).

For the HEAL Pain Management ERN, the Data Coordinating Center (DCC) at the University of Utah has the primary responsibility for data quality. The JHU-Tufts Statistical Core, physically located at the Johns Hopkins University, will examine data quality, create analysis data sets and derived variables, perform appropriate exploratory and confirmatory analyses for interim reviews of the data, and perform open-session DSMB reporting. All variables in the analytic database will be evaluated to detect gaps, patterns, and inconsistencies in the data. These analyses will emphasize examination of the nature and extent of variability for all variables. Visual techniques will explore continuous variables. We will build data quality control mechanisms into our data collection tools implemented in REDCap which will include strategies like 1) branching logic to avoid skip pattern questions; 2) required completion of selected data fields (for primary and selected secondary outcome measures); 3) questions arranged in such a way so that it makes the responder less likely to provide incorrect responses. For example, the Brief Pain Inventory (BPI) asks questions about extreme values such as maximum and minimum pain levels first, then asks about average pain; 4) using consistent response categories; 5) being mindful of survey burden on respondents and avoiding unnecessary questions. Data will be pulled and examined on a regular basis with queries on suspected data errors generated by the DCC at the University of Utah. These queries will be shared with the study team so they can follow up with participants as needed. Automated Research Electronic Data Capture (REDCap) alerts will also be used to inform study coordinators about missing data so they can follow up with participants to obtain responses. At the data analysis stage we will only try to detect and fix error/discrepancy outliers, i.e., data points that are clearly out of range or inconsistent with their definitions. We will assume other extreme values are real data and will try to address their influence in the analysis. Summary statistics for continuous variables will include the number of patients, median, mean, standard deviation, interquartile range, and minimum and maximum values, while statistics for categorical variables will include the frequency and percentage of participants in each category.

Statistical analyses will be led by Trial lead statistician (Dr. Zhou) and will be performed on data that have been quality-assured through DCC and Statistical Core protocols and monitoring reviews, and that have been exported directly from the SurgeryPal data warehouse. The following analytical procedures may be applied to blinded (no treatment arm designation, for study investigators and staff), partially unblinded (treatment designation only as “A” and “B,” for DSMB and at interim analyses), and unblinded (full treatment designation, for interim or final analyses) data.

Figure



All data collection procedures and statistical analyses in this Statistical Analysis Plan (SAP) are considered final and were finalized and approved by the SurgeryPal principal investigator (PI) and trial lead statistician prior to the unblinding of the data. In addition, the data collection procedures and statistical analyses described in this SAP may be modified and are given precedence over the analytical plans outlined in the clinical protocol; however, any modifications or changes in the primary endpoint and/or its analysis will also be reflected in a protocol amendment. Any updates to the SAP will be thoroughly detailed and documented at the beginning of this document.

3.0. Study Endpoints

3.1. Efficacy - Primary Outcomes

Primary Outcomes: Aim 1 (T2/ Acute Phase): Acute Pain Intensity and Interference: this will be measured using a patient-reported outcome measure, the 11 item Brief Pain Inventory (BPI). This measure has 2 scales: pain intensity (4 items) and pain interference (7 items). The 24 hour (short) version of this measure will be completed via REDCap daily for 14 days beginning on day 1 after hospital discharge, and individual rate of recovery will be computed for each participant over 14 days to yield the primary acute pain outcome measure for Aim 1.

Primary Outcomes: Aims 2 and 3 (T3/ Chronic Phase): Chronic Pain Intensity and Interference: this will be measured using a patient-reported outcome measure, the 11 item BPI. This measure has 2 scales: pain intensity (4 items) and pain interference (7 items). The 24 hour (short) version of this measure will be completed via REDCap daily for 7 days beginning 3 months after the participant's surgery date, and will be *averaged for each participant over 7 days* to yield the primary chronic pain outcome measures for Aims 2 and 3.

Missing data for multi-item instruments. If there is a published rule on treatment of missing data for a particular instrument, it will be followed. Otherwise, for less than 20% missing data, hot deck imputation will be used for individual items.

Objective	Brief Description of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
ASSESSMENT TIME POINT: **T1 = pre-randomization baseline, 4 weeks (window 3-8 weeks) before surgery** **T2 = acute outcomes, daily measures 1-14 days after hospital discharge;** **other questionnaire measures delivered 1 week after hospital discharge (window 1-3 weeks post hospital discharge)** **T3 = chronic outcomes, 3 months post-surgery (window 12-16 weeks)** **T4 = final follow up, 6 months post-surgery (window 6-8 months post-surgery)**				
Determine effectiveness of a pre-operative mHealth psychosocial intervention to improve acute pain outcomes in adolescents undergoing major musculoskeletal surgery	The Brief Pain Inventory Short Form assesses self-reported acute pain intensity, pain interference and medication use in a 24-hour timeframe. Questions evaluate worst, least, average, and current pain intensity, and pain interference with sleep, mood, walking ability, general physical activity, work/school, relationships, and enjoyment of life.	Brief Pain Inventory Short Form Pain Severity and Pain Interference Scales (Teen Report)	Daily for 14 days on days 1-14 post-hospital discharge (T2)	Pain intensity is continuous, reported on a scale of 0 (no pain) to 10 (worst pain imaginable). Each pain intensity item can be used on its own (worst, least, average, current). The mean severity score requires all 4 pain intensity items to be completed. Pain interference is reported on a scale of 0 (does not interfere) to 10 (completely interferes). It is scored as the mean of all interference items. This mean can be used if 4/7 or more items are completed.
Determine effectiveness of a post-operative mHealth psychosocial intervention to improve chronic pain outcomes at 3 months in adolescents undergoing major musculoskeletal surgery.	The Brief Pain Inventory Short Form assesses self-reported acute pain intensity, pain interference and medication use in a 24-hour timeframe. Questions evaluate worst, least, average, and current pain intensity, and pain interference with sleep, mood, walking ability, general physical activity, work/school, relationships, and enjoyment of life.	Brief Pain Inventory Short Form Pain Severity and Pain Interference Scales – (Teen Report)	Daily for 7 days at T3	Pain intensity is continuous, reported on a scale of 0 (no pain) to 10 (worst pain imaginable). Each pain intensity item can be used on its own (worst, least, average, current). The mean severity score requires all 4 pain intensity items to be completed. Pain interference is reported on a scale of 0 (does not interfere) to 10 (completely interferes). It is scored as the mean of all interference items. This mean can be used if 4/7 or more items are completed.
Determine the combined effects of pre- and post-	The Brief Pain Inventory Short Form assesses self-reported acute pain intensity, pain interference	Brief Pain Inventory Short Form Pain	Daily for 7 days at T3	Pain intensity is continuous, reported on a scale of 0 (no pain) to 10 (worst pain imaginable). The mean severity score requires all 4 pain intensity items to

Objective	Brief Description of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
operative CBT on chronic pain outcomes.	and medication use in a 24-hour timeframe. Questions evaluate worst, least, average, and current pain intensity, and pain interference with sleep, mood, walking ability, general physical activity, work/school, relationships, and enjoyment of life.	Severity and Pain Interference Scales (Teen Report)		be completed. Pain interference is reported on a scale of 0 (does not interfere) to 10 (completely interferes). It is scored as the mean of all interference items. This mean can be used if 4/7 or more items are completed.

3.2. Secondary Outcomes

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Determine effectiveness of psychosocial intervention to improve Health-Related Quality of Life (HRQL) following surgery (Aim 3).	15-item measure assessing self-reported physical, social, and HRQL over the last 7 days. Pain intensity and PedsQL will be used to define a binary variable for presence of chronic postsurgical pain.	Pediatric Quality of Life Inventory - Short form, acute version (PedsQL) (Teen Report)	T1, T3, T4	Items are scored on a 5-point Likert scale from 0 (Never) to 4 (Almost always). All items are reverse scored and linearly transformed on a 0-100 scale (0=100, 1=75, 2=50, 3=25, 4=0) so that higher score indicates better quality of life. If more than 50% of the items in a given scale are missing, the scale scores should not be computed. All subscale scores are a mean of all items within the scale. Yields the following subscales: physical health, psychosocial health, emotional functioning, social functioning, school functioning, and total. Both total score and individual subscale scores will be analyzed. This is analyzed as a continuous variable.

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Determine effectiveness of pre-operative psychosocial intervention to reduce opioid use.	Parents and youth will report medication use on the BPI, including selecting medication name from a dropdown list (opioid, acetaminophen, anti-inflammatory, other).	Brief Pain Inventory medication use items (Teen Report)	T1, T2, T3, T4	Participants will check boxes to indicate if they have used certain types of medication in the past 24 hours. At T1 only, the 7-day recall version of BPI will be used. We will analyze opioid use as days of use, as count. Days until opioid cessation will be used for descriptive purposes. They will also indicate on a continuous scale how much relief those medications have provided from 0% (no relief) to 100% (complete relief).
Determine effectiveness of psychosocial intervention to reduce opioid misuse following surgery (Aim 3).	This tool adapted for self-report in children age 11-17 by the American Psychiatric Association (40), and recommended by the National Institute on Drug Abuse for substance use screening in adolescents, will be used to assess alcohol, smoking, substance, and opioid use (43) (65). Additional questions on reasons for misuse (e.g. to treat pain, to get high, etc.) and source of opioids (e.g. prescription, from a friend or relative) will be collected using the corresponding questions and response options used in the National Survey on Drug Use and Health.	The American Psychiatric Association Adapted NIDA Modified ASSIST Tool 2 (Teen Report)	T1, T3, T4	Each item on the measure is rated on a 5-point scale (i.e., 0=not at all; 1=less than a day or two; 2=several days; 3=more than half the days; 4=nearly every day). Scores on the individual items should be interpreted independently because each item inquires about the use of a distinct substance. The rating of multiple items at scores greater than 0 indicates greater severity and complexity of substance use. To track change in the severity of the child's use of alcohol, tobacco/nicotine, prescription or illicit substance over time, the measure be may completed at regular intervals as clinically indicated, depending on the stability of the child's symptoms and treatment status.

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Determine effectiveness of perioperative psychosocial intervention to reduce psychosocial distress after surgery (Aim 3).	This measure assesses parent and youth self-reported anxiety (fear, anxiety, hyperarousal, and somatic symptoms) and depression (negative mood, negative views of self, and negative social cognition, decreased positive affect and engagement).	Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety 8a and Depression 8a Instruments (Teen Report)	T1, T3, T4	Each question has five response options ranging in value from one (Never) to five (Almost Always). A total raw score for anxiety or depression can be obtained for a short form with all questions answered by calculating the sum of the values of the response to each question. We will use the non-normalized score analyzed as a continuous variable. If 20% or fewer responses at items level are missing, we will impute the missing items using probabilistic hot-deck imputation as described in Andridge and Little (2010) then the total raw scores can be generated. If more than 20% items are missing we will not conduct imputation at item level for the participant. Instead the resultant raw total score will be considered missing and addressed by multiple imputation methods in the analysis stage.
Determine effectiveness of perioperative psychosocial intervention to improve mental health in parents after surgery (Aim 3).	A 4 item screening measure of general anxiety and depressive symptoms	PHQ-4 (GAD-2 and PHQ-2) (Parent Report)	T1, T3, T4	The PHQ-2 and GAD-2 each have four response options indicating the frequency of certain problems ranging from 0 (Not at all) to 3 (Nearly Every Day). Total scores for both the PHQ-2 and GAD-2 are calculated simply by adding the score of each item, with both items required for each measure. All 4 items are summed to yield the PHQ-4. These are analyzed as continuous variables.
Determine effectiveness of perioperative psychosocial intervention to reduce psychosocial distress in children after surgery.	A 4-item measure that screens for anxiety and depressive symptoms.	PHQ-4 (GAD-2 and PHQ-2) (Teen Report)	T1, T3, T4	The PHQ-2 and GAD-2 each have four response options indicating the frequency of certain problems ranging from 0 (Not at all) to 3 (Nearly Every Day). Total scores for both the PHQ-2 and GAD-2 are calculated simply by adding the score of each item, with both items required for each measure. All 4 items are summed to yield the PHQ-4. These are analyzed as continuous variables.

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Determine effectiveness of perioperative psychosocial intervention to improve sleep quality.	This 10–item measure assesses self-reported sleep quality. 4 additional items assess sleep duration on weekdays and weekends.	Adolescent Sleep Wake Scale 10-item version –(Teen Report)	T1, T3, T4	Continuous. Scores range from 1 (Never, 0% of the time) to 6 (Always, 100% of the time), with higher scores indicating better sleep quality. Some items are reverse scored. Mean subscale scores for each dimension (mean of items in subscale) and a full-scale sleep-quality score (ASWS total; mean of 5 subscales) are obtained. If 20% or fewer responses at items level are missing (of the original 10 ASWS items), we will impute the missing items using probabilistic hot-deck imputation as described in Andridge and Little (2010) then the total raw scores can be generated. If more than 20% items are missing we will not conduct imputation at item level. Instead the resultant raw total score will be considered missing and addressed by multiple imputation methods in the analysis stage.
Determine effectiveness of perioperative psychosocial intervention to reduce pain catastrophizing.	Self-report 13 item measures that assess child pain catastrophizing and parent's catastrophizing about child's pain, respectively.	Pain Catastrophizing Scale – Child Version (Teen Report) (PCS); Pain Catastrophizing Scale – Parent Version (Parent Report)	T1, T3	Each item uses a 5-point scale with the end points (0) not at all and (4) all the time. The PCS yields a total (raw sum) score and three subscale (sum) scores assessing rumination, magnification and helplessness. PCS total scores range from 0 – 52 analyzed as a continuous variable. If 20% or fewer responses at items level are missing, we will impute the missing items using probabilistic hot-deck imputation as described in Andridge and Little (2010) then the total raw scores can be generated. If more than 20% items are missing we will not conduct imputation at item level for the participant. Instead the resultant raw total score will be considered missing and addressed by multiple imputation methods in the analysis stage.

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Determine effectiveness of perioperative psychosocial intervention to reduce global pain severity.	A 1-item measure that assesses adolescent's self-reported global impression of pain severity.	Patient Global Impression of Severity Scale (Teen Report) (PGI-S)	T2, T3, T4	Categorical. One items, scores range from 0 (none) to 3 (severe) with higher scores indicating higher pain severity over the past week.
Assess baseline pain severity and interference.	The Brief Pain Inventory Short Form assesses self-reported acute pain intensity, pain interference and medication use in a 24-hour timeframe. Questions evaluate worst, least, average, and current pain intensity, pain interference with sleep, mood, walking ability, general physical activity, work/school, relationships, and enjoyment of life, as well as non-opioid and opioid analgesic use.	Brief Pain Inventory Short Form Pain Severity and Pain Interference Scales (Teen Report)	T1 (completed once)	Pain intensity is continuous, reported on a scale of 0 (no pain) to 10 (worst pain imaginable). Each pain intensity item can be used on its own (worst, least, average, current). The mean severity score requires all 4 pain intensity items to be completed. Pain interference is reported on a scale of 0 (does not interfere) to 10 (completely interferes). It is scored as the mean of all interference items. This mean can be used if 4/7 or more items are completed.
Assess treatment satisfaction and acceptability for a perioperative psychosocial intervention.	An adapted self-report measure specific to pediatric postsurgical pain that will collect parent and youth assessments of acceptability and satisfaction with the intervention's content and management approach	Treatment Evaluation Inventory, short form (TEI-SF) (Teen and Parent Report)	T2 (PT1), T3 (PT2)	Each item is reported on a scale from 1 (strongly disagree) to 5 (strongly agree). This variable will be used in descriptive analyses only. One item is reverse scored. A raw total score will be calculated by adding the sum of all items. If 20% or fewer responses at items level are missing, we will impute the missing items using probabilistic hot-deck imputation as described in Andridge and Little (2010) then the total raw scores can be generated. If more than 20% items are missing we will not conduct imputation at item level. Instead the resultant raw total score will be considered missing and addressed by multiple imputation methods in the analysis stage.

Objective	Brief Description/ Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Assess treatment engagement for a perioperative psychosocial intervention.	The mobile application and website will track each time a parent or youth uses the program. Treatment adherence will be collected by assessing number of completed modules/lessons.	Mobile application, website use	During assigned treatment phases: Pre-operative (4 weeks leading up to surgery); Post-operative (starting week 3 post-surgery)	Treatment engagement will be binary – if the participant has downloaded the app/logged into the website. Treatment adherence will be the # of completed modules ranging from 0-3 for both pre-op and post-op (categorical variable).
Collect information on concomitant therapy and other health services used.	We will collect information from parents at each assessment on use of concomitant therapy or additional health services (e.g. psychosocial interventions that the adolescent received during the pre- or post-operative period, visits to health care providers/specialists, not including standard surgery and anesthesia visits).	Concomitant Therapy Questionnaire (Parent Report)	T2 (PT1), T3 (PT2)	Categorical, descriptive data.
Report treatment experience.	A self-report scale with open ended questions and ratings of reactions to the treatments (e.g., stress, technical issues) from parents and youth.	Treatment Experience Questionnaire (Teen and Parent Report)	T2 (PT1), T3 (PT2), T4	Participants will indicate whether they experienced a range of potential negative effects (checkbox) and then report how much discomfort those experiences caused on a scale from 0 (did not affect me/my teen at all) to 3 (affected me/my teen very negatively).
Collect sociodemographic information.	Parents report on participants' sex, gender, race and ethnicity, family income, parent education, youth's pain history	Demographic Form (Parent Report)	T1	Categorical, descriptive data. Age (continuous), sex (categorical), race/ethnicity (categorical) will be analyzed as covariates.

3.3. Safety Measures

Clinically related adverse events and unanticipated problems (UP) cannot be collected systematically using this study design. This study does not have any study visits, and all outcomes are self-reported by the patient or by the parent. The study team will not have access to clinical data or medical records. Therefore, safety monitoring for this study will be based on three sources: 1) spontaneous reporting of unanticipated problems by parents or youth to study staff, 2) self-reported experiences with the treatments collected systematically using a questionnaire (Treatment Experience Questionnaire) completed by parents and youth at post-treatment time points, and 3) detection of any software or technology problems that compromise patient data.

An adverse event (AE) is defined as “Any unfavorable, unintended diagnosis, symptom, sign (including an abnormal laboratory finding), syndrome, or disease that occurs during the study, having been absent at baseline, or—if present at baseline—appears to worsen.”

A serious adverse event (SAE) is an adverse event that meets one or more of the following criteria:

- Results in death
- Is life threatening (places the subject at immediate risk of death from the event as it occurred)
- Results in inpatient hospitalization or prolongation of existing hospitalization
- Results in a persistent or significant disability or incapacity
- Results in a congenital anomaly or birth defect
- An important medical event that may not result in death, be life threatening, or require hospitalization may be considered an SAE when, based upon appropriate medical judgment, the event may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

Planned hospitalization scheduled before the enrollment of study participant is not an SAE.

An AE will be considered unexpected if it occurs in one or more subjects, the nature, severity, or frequency of which is not consistent with the expected natural progression of any underlying disease, disorder, or condition of the subject(s) experiencing the adverse event and the subject’s predisposing risk factor profile for the adverse event. AE or SAE expectedness is determined by the site PI. Additional adjudication with regard to expectedness, if necessary, will be provided by an independent Medical Monitor.

3.3.1. Safety Measures Summary

The study will use a self-report questionnaire (Treatment Experiences Questionnaire) developed to obtain systematic data on the experiences of youth and parents with the behavioral treatments being administered. The questionnaires will be administered at post-treatment time points (T2, T3 and T4). This includes the following items:

1. Did you experience any of the following negative effects due to your participation in the study (yes/no for each)?
    a. I couldn't stop thinking or worrying about my pain.
    b. I was overwhelmed by the treatment assignment.
    c. I felt there was more tension at home during treatment.

d. I was worried thinking if the treatment didn't work, nothing would work.
e. I was anxious because of the things I was asked to do.

2. How much discomfort did these experiences cause? 0 ('Did not affect me at all") to 3 ("Affected me very negatively").

3. During this study, did you experience any stressful life event unrelated to your participation in the study (for example: a car accident, an illness in the family, a break up, the loss of a pet)? (yes/no)
    a. If yes, tell us a little about what happened (open text field).

Study staff at SCH will follow up by phone with any participants who endorse experiencing a great deal of discomfort (i.e., ratings of 3, affected me very negatively) from study participation to find out more information and identify any adverse events.

At the time of DSMB evaluations, unanticipated problems from these three sources will be tabulated by intervention arm and included in the closed DSMB report.

3.3.2. Safety Monitoring and Recruitment Suspension Rules

Safety events greater than or equal to pre-specified thresholds will trigger a “suspend recruitment and review” by the DSMB to investigate the presumed cause and impact of these events. Such events are initially reviewed by the Statistical Core staff to determine if an event threshold has been reached, at which time the study investigators and the DSMB will be notified.

Circumstances that may warrant termination include, but are not limited to:

- Determination of unexpected, significant, or unacceptable risk to participants.
- Insufficient adherence to protocol requirements.
- Data that are not sufficiently complete and/or evaluable.
- Determination of futility.

3.4. Exploratory (Tertiary) Efficacy Endpoint

Objective	Brief Description/Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Explore mechanisms of an mHealth psychosocial intervention to prevent CPSP: a. compare rates of CPSP among youth receiving combined pre- and post-operative CBT (arm 4) to youth in arms 1, 2, and 3. b. examine mediation of treatment effects on CPSP via changes in acute *postsurgical pain.*	CPSP will be assessed as a binary outcome defined as pain that interferes with HRQL. Pain intensity will be assessed using a 0-10 NRS averaged over 7 days. HRQL will be assessed with the PedsQL – Short Form.	Pain intensity 0-10 numeric rating scale (NRS) for average pain in past 24 hours (from BPI), and PedsQL – Short Form (Teen Report)	T3 and T4; pain intensity assessed daily for 7 days at each timepoint.	Described in 15.1 below.
Characterize adolescent opioid use history.	Parents report on if their child has ever had a prescription for opioids, and if so, when and why their child was prescribed opioids.	Parent opioid question	T1	
Characterize adolescent lifetime misuse of opioid prescriptions, pain relief expectancy, and pain expectancy.	A single item from the Youth Risk Behavior Surveillance System (YRBSS) assesses how many times adolescents have misused opioids. Additional items assess expectations about the pain levels after surgery and pain medicine relief after surgery.	YRBSS question re: lifetime opioid misuse. Pain expectancy: 0-10 NRS. Pain medicine relief expectancy: 11-point scale from 0% - no relief to 100% - complete relief.	T1	
Explore adolescent pain relief preferences.	Teen participants endorse five risk-benefit items (e.g., “Having side effects from medicine is a bigger problem than having pain”) from strongly disagree (−2) to strongly agree (+2). This measure was adapted for ages 12-18.	Adolescent Pain Relief Preferences Survey	T1, T3	

Objective	Brief Description/Justification of Outcome Measure	Outcome Measured By	Time Frame	Scoring Summary
Explore adolescent perceptions regarding the riskiness of over the counter/opioid analgesics.	Teens rate the overall riskiness of taking over the counter and opioid analgesics across 6 items. Response options follow those of the NSDUH, with a 4-point Likert scale from 0 (No Risk) to 3 (Great Risk).	Risk Perceptions – Young Adults	T1, T3	

There are ADDITIONAL TERIARY/EXPLORATORY OUTCOMES (SEE PROTOCOL PAGE 20-21)

4.0. Sample Size Determination for the Primary Efficacy Analysis

Data available from our preliminary perioperative and chronic pain studies (Rabbitts, Groenewald, Tai, & Palermo, 2015; Rabbitts, Zhou, Groenewald, Durkin, & Palermo, 2015) and from prior perioperative intervention studies (Charette et al., 2015; Davidson, Snow, Hayden, & Chorney, 2016; LaMontagne, Hepworth, Cohen, & Salisbury, 2003) allowed us to conduct power calculations on primary pain outcomes for each of our aims (see below). For HRQL, we calculated sample size to provide adequate power to detect the Minimal Clinically Important Difference (MCID) score on the PedsQL of 4.5 (Varni, Burwinkle, Seid, & Skarr, 2003).

Power/sample size calculations for Aims 1-3.

Aim	Comparison Groups*		Primary Outcome variables	Mean difference (SD)	Power	N per smallest comparison group
	Control	Intervention				
1	Pre-operative education (Arms 1&3)	Pre-operative mHealth (Arms 2&4)	Brief Pain Inventory; T2	0.43 (0.38) 0.32 (0.38)	80% 90% 80% 90%	14 18 24 31
2	Education control alone (Arm 1)	Post-operative mHealth (Arm 3)	Brief Pain Inventory; T3	0.9 (1.7)	80% 90%	18 30
3	Single phase mHealth or education control alone (Arms 1,2,3)	Pre- & post-operative mHealth (Arm 4)	Brief Pain Inventory; T3	1.0 (1.8)	80% 90%	35 46
4	Single phase mHealth or education control alone (Arms 1,2,3)	Pre- & postoperative mHealth (Arm 4)	Rate of CPSP (defined by NRS and PedsQL; T3)	15%	80% 90%	86 113

*The number of treatment arms in the smallest comparison group differ for each aim

For Aim 1, the primary parameter of interest is the rate of recovery modeled using an exponential decay model. Based on our pilot data, we assume we will be seeing log(rate)=-1.21 (SE=0.38) in the intervention arm, which translates into a half-life (days taken to achieve 50% reduction in BPI outcome from time 0) of 4.13 days. If the control has a 50% increased half-life (then log(rate)=-1.64) then a sample size of 18 subjects per arm would provide 90% power to detect the difference (SD=0.38); N=14/arm for 80% power. If we assume the control arm will double the half-life (takes twice as long to achieve the same amount of reduction) which has log(rate)=-1.93, then sample size of 7/arm and 6/arm are needed for 90% and 80% power. Using a more conservative estimate of 35% increased half-life of recovery in the control arm, 31 subjects per arm would provide 90% power; N=24/arm for 80% power.

Based on our power calculations (provided in above table), a final sample size of 400 youth (100 per arm) will allow sufficient sample size in each comparison group to provide ample power to test each hypothesis on primary outcomes (Aims 1-3). For exploratory outcomes (Aim 4), we anticipate a rate of CPSP of 31% (see preliminary studies). With a sample size of 400 (i.e. 100

in smallest comparison group), we will have >80% power to detect an absolute reduction of 15% in the rate of CPSP in the group receiving combined full pre- and post-operative intervention.

Although we have had >95% retention in our prior perioperative and mHealth intervention studies, we conservatively plan for 20% attrition. Based on this, our original plan was to enroll 500 youth for a complete sample of 400 youth, to test the primary study hypotheses with optimal power. After further assessment of attrition rates throughout the study, we have revised our target sample size to at least 407 based on observed attrition rates of 4% at T2, 6% at T3, and 8% at T4. Enrollment will continue irrespective of sample size through June 30, 2024. This will provide sufficient power for all aims (including exploratory aims).

The efficacy primary outcome for Aim 1 relies on T2 data and requires a total of n=124 with 90% power. The efficacy primary outcome for Aim 2 and Aim 3 requires a total of n=184 with 90% power. Our exploratory efficacy aim requires a total of n=344-452 with 80-90% power. With an enrollment end date of 6/30/24 and using the same attrition rates as have been observed in the trial, we project ending the trial with n = 389 completed T2 (post surgery), n = 383 completed T3, which exceeds the n required for 90% power on the primary efficacy analyses. We project n = 374 for the exploratory aim, which may be analyzed separately after the NCE year.

5.0. Study Population

Our main analyses will be intent-to-treat; children are the unit of analysis. All randomized participants will be included in analyses. The inclusion and exclusion criteria are listed here:

Participant Inclusion Criteria

To be eligible to participate in this study, an individual must meet all of the following criteria:

- Teen
    - Willing to comply with all study procedures and be available for the duration of the study
    - 12 to 18 years old at the time of enrollment
    - Undergoing scheduled spinal fusion surgery (eligible indications: idiopathic scoliosis, juvenile scoliosis, spondylolisthesis or kyphosis)
- Parent/caregiver
    - Willing to comply with all study procedures and be available for the duration of the study
    - Parent or legal guardian of child who meets study criteria

Participant Exclusion Criteria

An individual who meets any of the following criteria will be excluded from participation in this study:

- Teen
    - Does not speak or understand English

  - Has severe learning disability, cognitive impairment or intellectual delay (i.e. unable to read at 5th grade level)
  - Does not have access to a smart device (smartphone, iPad or tablet)
  - Recent psychiatric admission (in the past 30 days)
  - Seeing counselor for 4+ sessions specifically to prepare for surgery
  - Severe systemic disease (neuromuscular scoliosis, cancer)
  - Takes medication daily for treatment of a chronic medical condition (except allergies, asthma, anxiety, depression)
  - Prior major surgery (open surgery such as heart, lung, brain, abdominal surgery, or prior spine surgery)
  - Diagnosed chronic musculoskeletal pain condition (complex regional pain syndrome, fibromyalgia or widespread musculoskeletal pain)
- Parent/caregiver
  - Does not speak or understand English
  - Does not have access to a smart device (smartphone, iPad or tablet)

5.1. Final Manuscript population

The primary efficacy analysis will be performed on the Intention-to-Treat (ITT) sample.
The ITT population includes all participants who were randomized and underwent surgery.
Patients whose surgery is cancelled or rescheduled outside the timeframe that allows study participation, regardless of whether they received the pre-surgery intervention, will be removed from the ITT population.
For the primary efficacy analysis, participants will be analyzed according to the treatment group to which they were assigned at the time of randomization.

5.2. Safety Analysis Sample

All randomized subjects will be included in the safety analysis sample. For safety endpoints, participants will be analyzed according to the assigned treatment.

5.3. Demographic and Baseline Characteristics

Demographics including age, sex, gender, race, and ethnicity will be summarized by treatment group using descriptive statistics for each of the defined samples. Baseline characteristics will be similarly summarized. Baseline variables are defined as pre-randomization variables such as comorbidities or any other potential confounders. If strong correlations exist within participants for repeated measures or among participants from the same study site, they will be accounted for using participant- and/or site- specific random effects in mixed effects models.
Randomization at the participant level with stratification by participant sex should balance many observed and unobserved confounders. Per the CONSORT 2010 Statement, we will not test for baseline differences between intervention and control groups as part of analysis (Schulz, Altman et al. 2010). ***We have a priori identified child/family socio-demographics (age, sex, and race/ethnicity), strong prognostic factors (pre-surgery pain severity), and study site as covariates*** for the proposed outcomes and will adjust for them in our analyses – this set of covariates will be referred to as the "prespecified adjustment".

6.0. Randomization

After baseline assessment is completed, parent/teen dyads will be block randomized with block sizes randomly selected from the list (4 or 8) to one of four arms in this 2x2 factorial design, stratified by teen participant sex (male, female). Intersex study participants will be randomized as males. At least 407 dyads (youth and parent/guardian) will be randomized to 4 treatment arms at a 1:1:1:1 ratio (see treatment arms table below), therefore we anticipate up to 125 youth per study group. Group 1 will receive education control only and will not receive access to the SurgeryPal intervention. Group 2 will receive the preoperative phase of the SurgeryPal intervention, and post-operative education. Group 3 will receive pre-operative education and the post-operative phase of the SurgeryPal intervention. Group 4 will receive both pre- and postoperative phases of the SurgeryPal intervention.

Group	Pre-operative	Post-operative
1	ED	ED
2	CBT	ED
3	ED	CBT
4	CBT	CBT

7.0. Blinding

Research staff will set up accounts and instruct the participants regarding how to access the assigned intervention. Investigators and healthcare providers will be blinded to treatment assignment. Participants will not be informed of their treatment assignment but may be able to deduce their condition based on the content they see in the app or website.

8.0. Multiplicity

Primary efficacy aims 1-3 involve a total of 6 planned models with primary BPI outcomes. The main hypothesis testing will be based on Wald tests or 95% confidence intervals on corresponding regression coefficients. Given the small number of tests, no adjustment for multiplicity of testing will be performed for primary aims.

Our secondary analyses will involve a large number of secondary outcomes (listed in Section 3.2). To account for multiple comparisons, all p-values for between arms comparisons of the differences in these outcomes will be adjusted to have a false discovery rate below 5% using the Benjamini-Hochberg procedure (Benjamini 1995, Benjamini 2001). An effect is considered significant only if the adjusted p-value<.05.

9.0. Missing Data

The clinical investigation team will make substantial efforts to ensure complete collection of data for all participants and to accrue minimal loss to follow-up to optimize evaluation of the primary efficacy outcome.

The primary intention-to-treat (ITT) analyses will be based on all available data. Given the RCT design, the predictor of interest, ie, the treatment group indicator will be complete for every subject. Baseline stratification variables (which will be adjusted in regression models) will be

complete as well. **We anticipate missing data will only occur in the outcome variables.** For T2 BPI measures, subjects may miss one or more of the 14 days assessments. Similarly, daily data may be missing for T3 and T4. But for T3 and T4 BPI, outcomes will be based on averages of 7 days of assessments – an average can be calculated unless subjects drop out and/or complete fewer than 3 daily surveys.

Amount of missing data and pattern of missing data will be examined. For multi-item instruments, if there is a published rule on treatment of missing data for a particular instrument, it will be followed. Otherwise, for less than 20% missing data, we will impute the missing items using probabilistic hot-deck imputation as described in Andridge and Little (2010) then the total raw scores can be generated. If more than 20% items are missing, we will not conduct imputation at item level for the participant. Instead the resultant raw total score will be considered missing and addressed by multiple imputation methods in the analysis stage.

At each time point (T2-T4), we will examine the extent of missing data for the study outcome variables, such as the number of days completed by each study participant and the distribution of the number of days of missing data as well as the pattern of missingness: whether the missing data only occurs in the early vs later portion of each data collection time point or whether it is random. Simple imputation using the “best” and “worst” scenarios for missing data will help identify the influence of missing data on the study results. If a substantial amount of missing data exists, in addition to our primary data analysis using all available data, we will conduct additional sensitivity analysis using Multiple Imputations with Changed Equations (MICE), assuming missing at random. Missing values are imputed based on the observed values for a given individual and the relations observed in the data for other participants, assuming the observed variables are included in the imputation model. Multiple imputation procedures, particularly MICE, are very flexible and can be used in a broad range of settings. Because multiple imputation involves creating multiple predictions for each missing value, the analyses of multiply imputed data take into account the uncertainty in the imputations and yield accurate standard errors. The MICE approach is a flexible yet powerful technique to address missing data issues and has demonstrated excellent performance in practice (Molenberghs and Kenward, 2007, Collins, Schafer et al. 2001, van Buuren 2007, Little, D'Agostino et al. 2012). If missing not at random is suspected based on the examination of the missing data, the pattern mixture model will be used to estimate model parameters for different patterns of missing data and then combine the estimates over strata (Little 2012).

In the event that a participant is lost to follow-up, the underlying reason(s) for dropping out will be tracked in order to help assess the mechanisms of missingness. However, if the amount of missing data is >5%, or if it is suspected that the missingness for the outcome measures does not follow MCAR upon detailed examination of reasons for missingness, then we will conduct a series of sensitivity analyses to determine how sensitive results are to extreme assumptions for missingness. For sensitivity analysis, we will apply the controlled multiple imputation (MI) techniques for sensitivity analysis of trials with missing continuous outcome data. These include δ- and reference-based MI procedures. In δ-based imputation, an offset term, δ, is typically added to the expected value of the missing data to assess the impact of unobserved participants having a worse or better response than those observed. Reference-based imputation draws imputed values with some reference to observed data in other groups of the trial, typically in other treatment arms (Cro et al. 2020).

10.0. Outlier Measures

An outlier for a particular measure is an observation that lies outside the expected distribution of observations for this measure. A common definition of an outlier is based on the interquartile range. Tukey (1977) defined an outlier as any observation that falls outside +/- 1.5 times the interquartile range. Outliers can occur due to measurement error; in which case it is justified to remove these outliers from the data set. Alternatively, outliers can also be "real" and "valid" data points, often resulting from chance variation in measurements, a skewed data distribution, or result from a variable comprised of multiple underlying distributions. The presence of outliers can give rise to biased statistics and treatment effect estimates. However, removing an outlier that reflects a valid data point is controversial. As an additional sensitivity analysis, we will explore any potential bias that may result from outlying observations by replacing these outlying observations with the first percentile or 99th percentile for observations on the lower and upper sections of the outcome distribution, respectively, for the primary and secondary aims, if outliers for the outcomes under these aims are present for any individual.

11.0. Exploratory Analyses

See section 15 below.

12.0. Primary Efficacy Analyses

12.1. Primary and secondary analytic plan for Hypothesis 1

Hypothesis 1: Youth receiving pre-operative CBT (arms 2&4) will have a faster rate of recovery (acute pain intensity and interference: primary outcomes), as well as lower opioid use during the 14 days following hospital discharge, compared to youth who receive preoperative education (arms 1&3). The primary outcome will be the daily BPI assessments treated as repeated measures. Separate models will be conducted for pain severity and pain interference.

The primary analytic plan for Hypothesis 1 will be examining the individual rate of recovery over 14 days in pain intensity and interference (**Brief Pain Inventory**) assessed immediately post-discharge. For this hypothesis, arms 1&3 will be combined into the control group (pre-operative education); arms 2&4 will be combined into the intervention group (receiving pre-operative mHealth intervention). Note these outcomes will be assessed before any post-operative mHealth intervention has taken place.

To test the hypothesis that intervention youth would have a more rapid rate of recovery in pain and interference than control youth, we will model a trajectory of pain symptoms using an exponential decay framework. Given that pain will be assessed 14 times, we believe the data will provide sufficient resolution to model time as a continuous variable when examining change in pain intensity and interference over time. Let Y_{ij} denotes the BPI measured for *i*th subject at time t_j, x_i denotes the group assignment for *i*th subject, $\boldsymbol{Z}_i$ denote the vector of covariates, we will consider the following general model specification:

$$(1)\ Y_{ij} = f\left(\boldsymbol{\beta}_i^{(x_i)},\ t_j\right) + \boldsymbol{\gamma Z}_i + \ \epsilon_{ij};\quad \boldsymbol{\beta}_i^{(x_i)} \sim N\left(\boldsymbol{b}^{(x_i)}, \boldsymbol{\Sigma}\right);\quad \epsilon_{ij} \sim N(0, \sigma^2).$$

Function $f(\boldsymbol{\beta}, t)$ can take flexible functional forms to capture different shapes of trajectories. For example, $f(\boldsymbol{\beta}, t) = \beta_0 + \beta_1 t$ represents a linear trend over time; $f(\boldsymbol{\beta}, t) = \beta_0 + \beta_1 t + \beta_2 t^2$

represents a quadratic trend over time; or more generally, we can assume a restricted cubic spline for $f(\boldsymbol{\beta}, t)$ to represent more flexible trajectory shapes. Based on our pilot study and prior studies in the literature, we have observed that pain symptoms and other health outcomes exhibit faster rates of improvement in the acute phase (<1 month), and then changes tend to slow down and level off. For these reasons we will use exponential decay; a more parsimonious and accurate function for time trend function in the form

$$(2)\ f(\boldsymbol{\beta}, t) = \beta_2 + (\beta_0 - \beta_2) \times \exp(-\beta_1 t),$$

where *t* is time, β_0 is the initial value of the BPI score at *t=0*, β_2 is the asymptote as *t* goes to infinity, β_1, ***constrained to be positive***, is the rate of change or how quickly the process decays/improves from the initial value to the asymptote.[147] Examples of trajectories based on exponential decay functions are shown in Figure 6. Parameters ${\beta_1}^{(1)}$ and ${\beta_1}^{(2)}$ are the group specific shape parameters, and equality of these parameters will be tested using multivariate Wald-test to assess the hypothesis that pain (BPI) declines more quickly in mHealth intervention subjects compared to control.

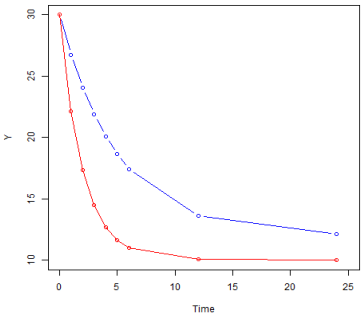


Figure 6: Examples of trajectories based on an exponential decay function

Secondary analysis of Hypothesis 1 will be comparing the time-average difference (TAD) between study arms. Suppose we have n_1 subjects in group 1 (treatment) and n_2 subjects in group 2 (control) for a total of N subjects, each measured on M=14 occasions at times t_j (j = 1, ..., M). The mean of continuous outcome y_{ij} is modeled by

$$(3)\ \mu_{ij} = \beta_1 + \beta_2 r_i + \beta_3 x_i$$

where
y_{ij} is the j^{th} response from subject i, with variance given by σ^2.
x_i is the vector of pre-specified baseline covariates
μ_{ij} is expectation of y_{ij},
r_i is the subject treatment indicator with 0 for control and 1 for treatment,
β_1 is the regression coefficient giving intercept of the control group,
β_2 is the regression coefficient giving the TAD between the two groups,
β_3 is the regression coefficient(s) of the baseline covariates

In this procedure, the primary interest is on β_2. GEE with assumed AR(1) working correlation structure will be used to estimate and test hypotheses about β_2 based on Wald test.

12.2. Primary analytic plan for Hypothesis 2 Primary Outcomes

Hypothesis 2: *Youth receiving post-operative CBT alone (arm 3) will have lower pain intensity and interference (primary outcomes), compared to youth receiving education control (arm 1).* The primary outcomes will be the average of 7 days of daily pain intensity and interference (**Brief Pain Inventory**) assessed at 3 months post-surgery. We will consider the education control arm (arm 1) as the control group and post-operative mHealth alone arm (arm 3) as the intervention group. Note both of these groups receive pre-operative education control. The main analysis will be based on covariate-adjusted linear regression models. We will consider pain outcomes as continuous outcomes and apply linear regression models. Specifically, let y_i denote the outcome for participant *i*, let U_i be the intervention indicator (0=control,

1=intervention), let $\boldsymbol{X}_i$ be the collection of covariates specified a priori, then the regression equations will take the following form:

$$(4) \qquad y_i = b_0 + U_i\alpha + \boldsymbol{X}_i\beta + \epsilon_i$$

The parameter of interest *α* captures the expected difference between intervention and control groups and will be tested for significance based on Wald t-test.

12.3 Primary analytic plan for Hypothesis 3 Primary Outcomes
Hypothesis 3: *Youth receiving both phases of peri-operative CBT (arm 4) will have lower pain intensity and interference compared to youth receiving single phase intervention (arms 2 and 3) or education control alone (arm 1) at 3 months.* The primary outcomes will be the average of 7 days of daily pain intensity and interference (**Brief Pain Inventory**) assessed at 3 months postsurgery. For this hypothesis comparing individual study arms, the predictor of interest U_i in the regression equation (4) will be treated as a 4-category factor (0=arm #4 both pre- and post-operative mHealth interventions, 1=arm #1 education control, 2= arm #2 pre-operative mHealth, 3= arm #3 post-operative mHealth). Primary outcomes will be the average pain intensity and interference over 7 days (**Brief Pain Inventory)** assessed at 3 months post-surgery. Using arm #4 as the reference group, the coefficients α estimate the expected differences between arms #1-3 versus arm #4 and will address our hypothesis. Note using proper contrast and linear combination of regression coefficients, Hypothesis 3 can be addressed by this regression model as well.

13.0. Secondary Efficacy Outcomes Analyses

13.1. Primary analytic plan for Hypothesis 1 Secondary Acute Outcomes
Youth receiving pre-operative CBT (arms 2 and 4) will have lower opioid use during the 14 days following hospital discharge, compared to youth who receive pre-operative education (arms 1 and 3). For this hypothesis, arms 1&3 will be combined into the control group (pre-operative education); arms 2&4 will be combined into the intervention group (receiving pre-operative mHealth intervention). Secondary outcomes at acute phase include opioid use (**Brief Pain Inventory Short Form**; # days of opioid use during 14 days daily monitoring). Note these outcomes will be assessed before any post-operative mHealth intervention has taken place. Our primary analysis will be based on covariates adjusted regression models. Opioid days will be treated as count variables and analyzed using Poisson regression or negative binomial regression if over-dispersion is detected. Specifically, let y_i denote the count for participant *i*, and μ_i is expectation of y_i, let U_i be the intervention indicator (0=control, 1=intervention), let $\boldsymbol{X}_i$ be the collection of covariates specified a priori (see section 5.3), then the regression equations will take the following form

$$(5) \qquad f(\mu_i) = b_0 + U_i\alpha + \boldsymbol{X}_i\beta \ .$$

where $f(\mu_i)$ is a natural log of μ_i. The parameter of interest is *α* which captures the expected difference between intervention and control groups and will be tested for significance based on Wald t-test.

13.2. Primary analytic plan for Hypothesis 3 Secondary Chronic Outcomes.
Youth receiving both phases of peri-operative CBT (arm 4) will have lower pain intensity and higher HRQL, as well as lower pain interference, psychosocial distress, sleep disturbance, and

opioid misuse compared to youth receiving single phase intervention (arms 2 and 3) or education control alone (arm 1) at 3 months. For this hypothesis comparing individual study arms, the predictor of interest U_i in the regression equation (4) will be treated as a 4-category factor (0=arm #4 both pre- and post-operative mHealth interventions, 1=arm #1 education control, 2= arm #2 pre-operative mHealth, 3= arm #3 post-operative mHealth). Secondary Outcomes include HRQL (**PedsQL**), psychosocial distress (**Pain catastrophizing- parent and child; PROMIS Anxiety and Depression Instruments- child**), sleep disturbance (**ASWS**), and opioid misuse (**NIDA Modified ASSIST Tool**) assessed at 3 months post-surgery. Using arm #4 as the reference group, the coefficients α estimate the expected differences between arms #1-3 versus arm #4 and will address our hypothesis. Note using proper contrast and linear combination of regression coefficients, Hypothesis 2 can be addressed by this regression model as well. We will perform regression diagnostics on the linear regression models which will include assessing outliers, assessing influential observations, non-normality of error distribution, heteroscedasticity, multi-collinearity, non-linearity and correlated errors.

14.0. Safety Analyses

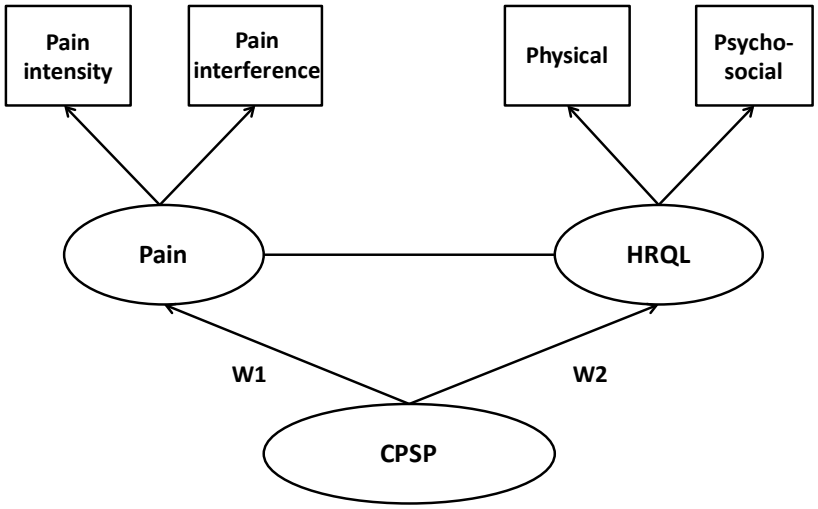
Clinically related adverse events and unanticipated problems cannot be collected systematically using this study design. This study does not have any study visits, and the study team will not have access to clinical data or medical records.

For the self-report questionnaire on the experiences of youth and parents with the study and treatments being administered (see Section 3.3.1), we will describe the frequency of negative responses to treatment and summarize the amount of discomfort caused by those experiences.

15.0. Tertiary (Exploratory/Sensitivity) Analyses

15.1 Analytic plan for Hypothesis 4 (exploratory)

Youth receiving combined pre- and post-operative CBT (arm 4) will have lower rates of CPSP (outcome: moderate-severe pain with impairment in HRQL) at 3- and 6-months compared to arms 1, 2, and 3. A binary outcome chronic post-surgical pain (CPSP) will be defined based on pain intensity (**Pain Intensity**), and HRQL (**PedsQL**) at 3-months post-surgery. We will develop and test cut points for chronic post-surgical pain (CPSP) by considering both pain severity and health related quality of life (both treated as latent constructs). Toward this we will conduct latent class analysis (LCA) following the proposed framework (see figure). This model will be implemented using multilevel structural equation models (SEM) (Bollen & Curran; Little; Preacher, Wichman, MacCallum, & Briggs). The parameters of interest are coefficients linking latent constructs “Pain” and “HRQL” (W1 and W2) to latent construct “CPSP” as the weighted sum will identify CPSP based on pooled information. We will conduct cross-validation using part of the data (70%) to develop the weights and part of the data (30%) to validate the predictive performance. Our evaluation of intervention effects on CPSP will then be based on the cut points derived from above analysis. It will be analyzed using a logistic regression model and the model will be specified similar to what is outlined for Hypothesis 3.



15.2 Analytic plan for Hypothesis 5 (exploratory)

Reductions in acute postoperative pain will mediate treatment effects on reducing incidence of CPSP, we will conduct exploratory mediation analysis. We will assess the extent through which acute post-operative pain intensity and interference exert their influence on (i.e., mediate) change in pain intensity, HRQL, and pain interference at 3 months follow up via the intervention. The mediational effects of acute pain reduction will be assessed using two comparison groups: education control only (Group 1) versus pre-operative mHealth intervention only group (Group 2), so that any mediation effect will not be confounded by second phase intervention. We will use a structural equation modeling (SEM) approach: specifically, we will evaluate mediation by decomposing a total effect of intervention on the pain (**Brief Pain Inventory; Pain Intensity**) and HRQL (**PedsQL**) outcomes into direct effects and indirect effects (David P. MacKinnon, 2008; D. P. Mackinnon & Fairchild, 2009; D. P. MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The indirect effects represent the influence of the intervention on change in chronic pain and HRQL via changes in acute post-op pain severity (**Brief Pain Inventory Short Form**). From a set of 5,000 bootstrap samples, 95% bias-corrected confidence intervals will be calculated for the indirect effects. A significant indirect effect will indicate that the pre-surgery intervention influences change in pain severity and interference at 3 months (short term chronic pain endpoint) by influencing acute post-op pain severity and interference. In addition, we will conduct exploratory mediation analysis by including all 4 study arms. Next we will extend this mediation analysis by considering “moderated mediation” through adding the interactions between second phase intervention and short term acute pain reduction in the mediation equations.

15.3 Analytic plan for Hypothesis 6 (exploratory)

We will assess the extent through which psychosocial factors exert their influence on (i.e., mediate) changes (T2-T4 relative to T1) in pain severity and pain interference during post-surgery recovery via the intervention. For this analysis, we will aggregate daily data to the timepoint level in order to achieve a single outcome value at each of the 4 timepoints (T1-T4). We will use a structural equation modeling (SEM) approach (Brown 1997, Little, Card, et al. 2007). Specifically, we will evaluate mediation by decomposing a total effect on the pain outcome slope into direct effects and indirect effects. The indirect effects represent the influence of the intervention on change in pain outcomes via psychosocial factors. From a set of 5,000 bootstrap samples, 95% bias-corrected confidence intervals will be calculated for the indirect

effects (Efron 1987, MacKinnon, Lockwood, et al. 2004, Shrout and Bogler 2002, Little 2013). A significant indirect effect will indicate that the intervention affects change in pain outcomes by influencing psychosocial factors. When appropriate, multi-level structural equation modeling (MSEM) will estimate predictive relationships while handling nested interdependence among observations (Little 2013). Using a SEM approach, we will conduct exploratory mediation analyses of sleep quality, anxiety and depression, and pain catastrophizing as mediators of the relationship between intervention and change in pain outcomes.

15.4 Analyses to test for differences in effect among sex, racial, and/or ethnic groups
We will conduct subgroup analysis to examine whether there are differential intervention effects across sex, racial, and ethnic groups on primary and secondary acute (Aim 1) and chronic (Aims 2 and 3) outcomes. These data are self-reported on the **Demographic Form.** The effect modifications will be examined by including group-by-moderator interaction terms in linear regression Equation (1). Moderation effects will then be tested by assessing the significance of the proper contrast among corresponding regression coefficients. Overall moderation will be assessed with likelihood ratio test (LRT) comparing models with and without interactions. We will also conduct additional subgroup analysis within groups defined by moderators to examine the differential effect sizes within individual subgroups.

15.5 Secondary data analysis
We will perform additional longitudinal data analysis to examine change in pain intensity and pain interference over time. Pain intensity and pain interference outcomes assessed at baseline, 2-weeks, 3-months and 6-months will be modelled using linear mixed effects regression models. Predictors will include time indicator, study arm indicator, arm-by-time interaction term and *a priori* identified covariates (see section 5.3). The arm-by-time interaction coefficients will assess differential changes in outcomes from baseline to post-op assessments between study arms thus will be parameters of interest. Participant-specific random effect will be included in the model to account for within participant correlations due to repeated measures. Estimation will be based on restricted maximum likelihood (REML) and significance will be assessed using likelihood ratio test (LRT) comparing models with and without interactions (Fitzmaurice & Laird; Pinheiro & Bates; Stroup).

Additional moderation analysis will also be conducted. Specifically, we will examine whether participant age, baseline adolescent sleep quality (**ASWS)**, and baseline psychosocial distress (**Pain catastrophizing- child and parent, PROMIS Anxiety and Depression Instruments- child, PHQ-4- parent**) strengthen or weaken the impact of the intervention on change in child pain intensity and pain interference. This will be achieved by including group-by-moderator interaction terms in Equation (1). Moderation effects will then be tested by assessing the significance of the contrast among corresponding regression coefficients. Overall moderation can be assessed with multivariate Wald statistic and F-test based on analysis of variance (ANOVA) for nested models. We will also conduct additional subgroup analysis within groups defined by moderators to examine effect modifications.

We will conduct additional sensitivity analyses to examine the influence of app engagement and adherence on treatment effect on primary outcomes.

16.0. References

Rebecca R. Andridge, RR and Little, RJA. A review of Hot Deck Imputation for Survey Non-response. International Statistical Review (2010), 78, 1, 40–64 doi:10.1111/j.1751-5823.2010.00103.x

Bollen, K. A., & Curran, P. J. Latent Curve Models: A Structural Equation Perspective. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

Charette, S., Fiola, J. L., Charest, M. C., Villeneuve, E., Theroux, J., Joncas, J., . . . Le May, S. (2015). Guided Imagery for Adolescent Post-spinal Fusion Pain Management: A Pilot Study. *Pain Manag Nurs, 16*(3), 211-220. doi:10.1016/j.pmn.2014.06.004

Davidson, F., Snow, S., Hayden, J. A., & Chorney, J. (2016). Psychological interventions in managing postoperative pain in children: a systematic review. *Pain, 157*(9), 1872-1886. doi:10.1097/j.pain.0000000000000636

Fitzmaurice, G. M., & Laird, N. M. W., J. Applied Longitudinal Analysis (2nd ed.), John Wiley & Sons, ISBN 0-471-21487-6, 2011.

LaMontagne, L. L., Hepworth, J. T., Cohen, F., & Salisbury, M. H. (2003). Cognitive-behavioral intervention effects on adolescents' anxiety and pain following spinal fusion surgery. *Nurs Res, 52*(3), 183-190.

Little, T. D. Longitudinal Structural Equation Modeling. New York, NY: The Guilford Press, 2013.

MacKinnon, D. P. (2008). Introduction to Statistical Mediation Analysis. Taylor & Fancis Group, LLC, New York.

Mackinnon, D. P., & Fairchild, A. J. (2009). Current Directions in Mediation Analysis. *Curr Dir Psychol Sci, 18*(1), 16. doi:10.1111/j.1467-8721.2009.01598.x

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods, 7*(1), 83-104.

Pinheiro, J. C., & Bates, D. M. Mixed-effects models in S and S-PLUS, Springer, New York, 2000.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. Latent growth curve modeling. Thousand Oaks, CA: Sage Publications, Inc., 2008.

Rabbitts, J. A., Groenewald, C. B., Tai, G. G., & Palermo, T. M. (2015). Presurgical psychosocial predictors of acute postsurgical pain and quality of life in children undergoing major surgery. *J Pain, 16*(3), 226-234. doi:10.1016/j.jpain.2014.11.015

Rabbitts, J. A., Zhou, C., Groenewald, C. B., Durkin, L., & Palermo, T. M. (2015). Trajectories of postsurgical pain in children: risk factors and impact of late pain recovery on long-term health outcomes after major surgery. *Pain, 156*(11), 2383-2389. doi:10.1097/j.pain.0000000000000281

Stroup, W. W. Generalized Linear Mixed Models, CRC Press, 2012.

Varni, J. W., Burwinkle, T. M., Seid, M., & Skarr, D. (2003). The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr, 3*(6), 329-341. doi:03-035R [pii]

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, *6*(4), 330.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., ... & Neaton, J. D. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, *367*(14), 1355-1360.

Molenberghs, G., & Kenward, M. (2007). *Missing data in clinical studies* (Vol. 61). John Wiley & Sons.

Tukey JW. *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley; 1977.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, *16*(3), 219-242.
Bollen and Curran, 2006, Little, 2013, Preacher, Wichman et al., 2008)
(MacKinnon, Lockwood et al. 2002, MacKinnon 2008, Mackinnon and Fairchild 2009
Efron 1987, MacKinnon, Lockwood, et al. 2004, Shrout and Bogler 2002, Little 2013)
(Fitzmaurice & Laird; Pinheiro & Bates; Stroup)
Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57, 289–300 http://www.jstor.org/stable/2346101.
Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics, 29, 1165–1188. doi: 10.1214/aos/1013699998
Cro, S, Morris, TP, Kenward, MG, Carpenter, JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine*. 2020; 1– 28. https://doi.org/10.1002/sim.8569

17.0. Appendix 1: Variable List

SAP Endpoint/Section	Analytic Variable in SAP	Variable Name in Dictionary
12.1-12.3, 13.1-13.2 Primary and Secondary Efficacy Endpoints	Acute pain intensity and interference	
	Chronic pain intensity and interference	
	Health-related quality of life	
	Medication use	
	Opioid misuse	
	Teen anxiety/depression	
	Parent anxiety/depression	
	Sleep quality, duration	
	Global impression of pain severity	
	Treatment acceptability and satisfaction	
	Application & website use	
	Concomitant therapy	
	Treatment experience	
	Chronic post-surgical pain – binary (CPSP)	
Baseline characteristics	Age	
	Sex	
	Ethnicity	
	Race	
	Sex	
	Gender	
	Income	
	Parent Education	
	Teen Surgery Indication (Diagnosis)	