Section 1: Administrative information

1.a. Title

Statistical analysis plan for Predicting, Understanding, and Speeding Recovery after Total Joint Replacement; a Randomized, Controlled Clinical Trial.

1.b. Trial registration

Trial Registration: NCT02685735

2. SAP version

Version: 2.1 Date: February 6, 2023

3. Protocol Version

This document has been written based on information contained in the study protocol version 7.0, dated March 20, 2020.

4. SAP Revisions

Protocol version	Updated SAP version no.	Section number changed	Description and reason for change	Date changed
7.0	2.0	Section 6	Details provided after data acquisition and before unblinding	May 21, 2022
7.0	2.1	Sections 2 and 6	Corrected primary analysis to utilize resting pupil diameter rather than its response to pain as proposed in the NIH funding and protocol documents.	Feb 6, 2023
7.0	2.2	Section 6	Corrected the primary analysis to remove adjustment for baseline pain and in secondary analyses to remove adjustment for baseline activity, disability score, and opioid self administration.	April 26, 2023

5. Roles and Responsibilities – non-signatory names and contribution

James C. Eisenach, M.D.
Professor of Anesthesiology
Wake Forest School of Medicine
Winston-Salem, NC
Role: Chief investigator

Timothy T. Houle, M.D.
Associate Professor of Anesthesiology
Massachusetts General Hospital
Boston, MA
Role: Co-investigator and Senior Statistician

6.a. Roles and Responsibility – signature of the person writing the SAP

James C. Eisenach, M.D.
Chief Investigator

April 26, 2013
Date

6.b. Roles and Responsibility – signature of senior statistician responsible

Timothy T. Houle, Ph.D.
Co-Investigator & Senior Statistician

April 26, 2013
Date

Section 2: Introduction

7. Background and Rationale

People recover from pain after major surgery with very different speeds. In rats, recovery from hypersensitivity after major surgery also exhibits large variability, some of which is accounted for by individual differences in pain-induced activation of the locus coeruleus (LC) resulting in central norepinephrine release that dampens sensitization of pain circuits. Gabapentin alters LC coeruleus tonic activity in animals. We designed a study to assess, via pupillometry, LC tonic activity and responsiveness to pain prior to surgery in patients scheduled for lower extremity total joint arthroplasty and to test whether tonic activity and responsiveness to pain modulate the speed of recovery and efficacy of gabapentin to alter recovery from postoperative pain.

8. Objectives

The objective of this research study is to better understand patterns of recovery after total hip arthroplasty (THA) and total knee arthroplasty (TKA). The study will evaluate how pain, pupil diameter as a measure of LC activity and cognitive (i.e., thinking style) determine patterns of recovery, and the study will evaluate the efficacy of gabapentin versus placebo for improving recovery after surgery.

Research hypothesis: The null hypothesis (H_0) is that modeled trajectory of change in pain intensity report after THA and TKA does not differ between oral gabapentin and placebo in a manner dependent on its interaction with preferred cognitive style and pre-surgery pupil resting diameter . The alternative hypothesis (H_1) is that there is a difference between the two groups which is dependent on these interactions. The null hypothesis will be tested through contrasting the fit of a null model (m0), with only change parameters and prognostic covariates, and the full model (m1), which includes all of the predictors in null model with the addition of randomized treatment group, cognitive style, resting pupil diameter, and all two and three-way interactions. The difference in fit will be evaluated using a likelihood ratio test, as further described below.

H_0: -2 * (loglik[m0] – loglik[m1]) = 0
H_1: -2 * (loglik[m0] – loglik[m1]) > 0

Study objective: The primary objective is to determine the effectiveness of oral gabapentin to speed recovery from pain, as measured by growth curve modeling, after THA or TKA and how this effectiveness is modified by preferred cognitive style and pre-surgery resting pupil diameter .

Secondary objectives are planned to explore the role of these factors interact to predict other outcomes from surgery, including speed of gain in activity, assessed by actigraphy, speed recovery from self-assessed disability, and time to cessation of opioid use; to explore whether change in pupil diameter to acute experimental pain explains additional variance in a 4-way interaction compared to the 3-way interaction in the primary analysis.

Section 3: Trial Methods

9. Trial design

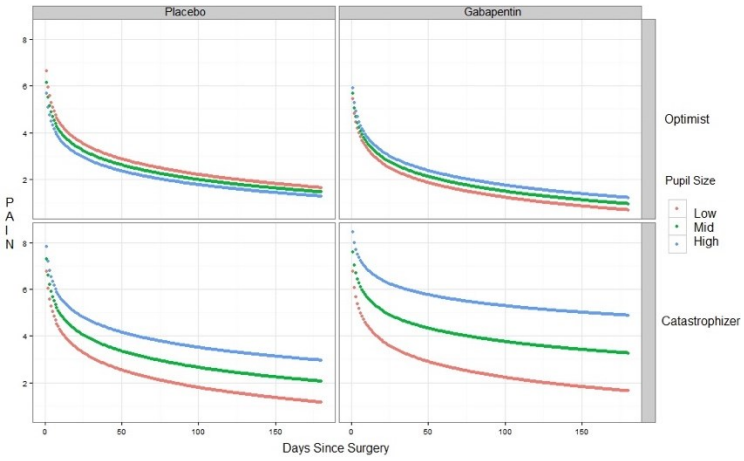
The trial is a two-center, randomized, double-blinded, parallel-group, placebo-controlled trial. Treatment allocation is a 1:1 ratio. Patients are randomized to either gabapentin or placebo control.

10. Randomization

A computer-generated randomization using a 1:1 allocation ratio in permuted blocks of 8 will be developed by the study statistician uninvolved with data collection and administered by the research pharmacist at Wake Forest School of Medicine. The 'blockrand' package in R will be used by the study statistician. The randomization will be stratified based on study recruitment site, concurrent use of a serotonin-norepinephrine reuptake inhibitor, and chronic low- or high-dose opioid therapy. Allocation will be conducted by the research pharmacist and will occur after consent and completion of the first visit. At all times, the allocation assignment will be concealed from the study coordinator and other investigators who manage participant recruitment and follow-up. Allocation concealment is maintained using a centralized pharmacy administration of study drug. Both study agents are designed to look identical to each other to maintain blinding.

11. Sample size

To calculate the statistical power for our analytical strategies, we simulated scenarios for which we instilled known relationships between the random effects of change models as well as a range of predictors of those parameters that were themselves correlated. A sample size of 250 was to determine to provide substantial power (> 90%) to detect clinically significant differences between a null model versus a demographic/perioperative + psychosocial predictors model using a likelihood ratio test. An example of the expected differences is displayed (left) that illustrates a three-way interaction between treatment (Placebo vs Gabapentin), cognitive-affective style (Optimist vs Catastrophizing), and pupil size (Low, Mid, High).



12. Framework

Superiority testing will be used to examine difference between the effect of treatment alone on the change process in pain after surgery and the effect of treatment as interacting with preferred cognitive style and pre-surgery pupil responses to painful stimuli.

13a. Information on interim analyses

There are no planned interim analyses.

13b. Any planned adjustment of the significance level due to interim analysis

None.

13c. Details of guidelines for stopping a trial early

There is no plan to stop the trial early aside from recommendations based on unanticipated serious adverse events by the DSMB.

14. Timing of final analysis

Table 1. Timing of study measures

Study period	Screen	Baseline	In hospital	Daily diaries	Follow up 1	Follow up 2	Phone	Orthopaedic clinic visits
Visit number	0	Visit 1			Visit 2	Visit 3		
Week		-4 to -2*	0-3 days	0-8	8	26	12,16,20,52	2, 6, 12-16, 26, 52
Informed consent		X						
Demographic		X						
Medical/surgical history		X						
Questionnaires		X			X	X		
Pain testing & pupillometry		X						
Randomisation		X						
Study drug dispensing		X						
In hospital pain/medication data & disposition			X					
Daily diary & accelerometry		X		X				
Phone assessment of pain, medication use, function							X	
Joint function								X

*Study medication begins at -2 weeks and continues until 3 weeks after surgery

Final analysis of the gabapentin vs placebo comparison with and without inclusion of interactions with preferred cognitive style and pre-surgery pupil responses to pain will take place when all patients have completed the 12-month follow-up and data for the primary endpoint have been cleaned, which occurred in May 2022. Results were anticipated to be posted by Jan, 2023, but complexities associated with generation of a novel psychologic construct from the final dataset and defining parameters for pupil assessment prevented completion of the primary analysis until Mar, 2023.

15. Timing of outcome assessments

The schedule of study procedures is provided in the protocol and summarized in the table below.

Section 4: Statistical Principles

16. Level of statistical significance

All applicable statistical tests will be 2-sided and will be performed using a 5% significance level.

17. Adjusting for multiplicity

The primary analysis involves a single model-based inference that will be interpreted at the 5% threshold of statistical significance. If the primary analysis is significant, each main factor, all two way interactions, and the three way interactions will be evaluated and each will be interpreted at the 5% level of significance with no adjustments made for multiplicity.

18. Confidence intervals

All confidence intervals presented will be 95% and two-sided.

19a. Definition of adherence to the intervention and how it is assessed

Compliance is assessed based on the number of daily diary entries after hospital discharge. It is defined as:

% compliance = (number of entries/number of days entries were to be completed)*100%

19b. How adherence will be presented

The number and % of participants completing more than 80% of the daily diaries after surgery will be provided by treatment group in a table. All available data will be included in the analysis from the intention to treat sample.

19c. Definition of protocol deviation

The following are pre-defined major protocol violations with a direct bearing on the primary outcome:

1) Surgery during the primary outcome assessment period (first 8 weeks after surgery)

19d. Description of which protocol deviations will be summarized

Protocol deviations are classified prior to unblinding of treatment. The number (and percentage) of patients with major and minor protocol deviations will be summarized by treatment group with details of type of deviation provided. The patients that are included in the ITT analysis data set will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

20. Analysis population

The intention-to-treat population will include all randomized patients who received study drug and provided at least one pain score after the day of surgery.

Section 5. Trial Population

21. Screening data

Enrollment: The dates of recruitment, the number of patients screened, the number of patients recruited, the number of screened patients not recruited, and the reason for non-recruitment will be reported as part of the CONSORT diagram.

22. Eligibility criteria

Inclusion and exclusion criteria are provided in the protocol. The number of ineligible patients randomized, if any, will be reported, with reasons for ineligibility

23. Information to be reported in the CONSORT flow diagram

A CONSORT flow diagram (appendix A) will be used to summarize the number of patients who were:

- assessed for eligibility at screening
- eligible at screening
- ineligible at screening*
- eligible and randomized
- eligible but not randomized*
- received the randomized allocation
- did not receive the randomized allocation*
- lost to follow-up*
- discontinued the intervention*
- randomized and included in the primary analysis

*reasons will be provided for post randomization exclusions.

24a. Description of level of withdrawal

The level of consent withdrawal will be tabulated (classified as “consent to continue follow-up and data collection” “consent to continue data collection only”, “complete – no further follow-up or data collection”).

24b. Timing of withdrawal/lost to follow up data

This will be presented in CONSORT diagram format rather than as a table, with numbers and reasons for withdrawal and/or exclusion from analysis given at each stage

24c. Reasons for withdrawal/lost to follow up data

The numbers (with reasons, if available) of losses to follow-up (drop-outs and withdrawals) over the course of the trial will be summarized by treatment arm.

25a. List of baseline characteristics to be summarized

Patients will be described with respect to age, sex, race, ethnicity, medications, number of painful conditions excepting the hip and their locations, cognitive – optimism cognitive construct generated from questionnaires, baseline pain from the daily diaries prior to surgery, and surgical approach (anterior vs posterior).

25b. How baseline characteristics will be described

Categorical data will be summarized by numbers and percentages. Continuous data will be summarized by mean, SD and range if data are normal and median, IQR and range if data are skewed. Minimum and maximum values will also be presented for continuous data. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

Section 6: Analysis

26. Outcomes and timings

Primary outcome: Worst pain intensity measured using a 0 to 10 numerical rating scale on a daily diary. Measurements will be obtained each evening from day 0 (surgery) to day 56 after surgery. The pain measurements will be evaluated in a statistical model using an intercept (i.e., immediate pain after surgery) and slope (i.e., rate of change of pain during the observation period).

Secondary outcomes: Daily measurements such as physical activity, as measured via actigraphy, disability, as measured by daily diaries and weekly WHODAS 2.0 scores, and opioid self-administration, as measured by daily diaries will be collected from day 0 (surgery) to day 56 after surgery. Questionnaire and cognitive game data will be assessed at baseline and follow-up. Validated scoring systems from the individual items will be used to score each questionnaire or measure. These outcomes will not be used in the primary analysis.

27a. Analysis methods to be used
and
27b. Adjustments for covariates

Primary analysis:

Overview

The primary analysis will model trajectory of recovery of daily worst pain ratings and examine several sources of influence on this recovery. Sometimes referred to as mixed-effects models, or hierarchical linear models, growth curve modeling allows specification of a change trajectory that is unique to each individual. The nature of the common form of changes in daily pain will be modeled using a curvilinear form (e.g., polynomial regression). Through the use of fixed and random effects, variations in intercepts (i.e., levels of pain on day of discharge) and slopes (i.e., change in daily pain over time) can be examined using the pre-specified influences outlined below. To model individual pain trajectories, several model form candidates will be examined with the best fit determined using a Bayesian Information Criterion. These model forms will include linear, quadratic, cubic, and log(time), and each of these model candidates will be employed that reflect a fundamental change in trajectory that occurs during the observation period. Based on preliminary data it is highly likely that a log(time) form will be chosen.[1,2]

To examine the proposed hypotheses, several key sets of predictors will be specified including patient characteristics, operative characteristics, cognitive style, randomized treatment assignment, and pre-operative pupil diameter response. Each of these predictors are described below:

Patient characteristics: To adjust the associations for individual age and sex will be entered into the model.

Operative characteristics: To adjust the associations for any subtle differences in medical care, study site, and surgery type (hip or knee arthroplasty) will be entered into the model.

Cognitive style: Multidimensional scaling of the pre-operative questionnaires will be conducted to create a unidimensional scale that represents a broader spectrum of cognitive style that ranges from optimism/acceptance of pain to catastrophizing. This novel predictor will be examined for reliability and correlations with the original scales. If insufficient reliability ($\alpha < 0.70$) or construct validity ($r < |0.40|$) with the original scales cannot be obtained, the analysis will proceed using the originally scaled catastrophizing predictor.

Pupil diameter: Pupil recordings will be conditioned to remove artifacts (e.g., blinks), saccades, and deviations from central gaze that might bias recordings. A response window with appropriate lag reflecting delayed onset of pupil response due to physiological delays will be used to define change in diameter to each stimulus as previously reported.[17] Two pupil measurements will be estimated from the session data, the mean pupil diameter during baseline (i.e., pre-stimulus) and the modeled response to pupil to 50-degree stimulus.

Gabapentin: Randomized treatment assignment of either gabapentin or placebo will be a fundamental predictor of trajectory.

These individual predictors sets will then be used to test the primary hypothesis that gabapentin alters time course of recovery after arthroplasty in a manner dependent on its interaction with preferred cognitive style and pre-surgery pupil diameter. The primary statistical inference will be the difference in the likelihood ratio between two nested models using a chi-square test with K degrees of freedom (i.e., K being the difference in the number of parameters between the two models). The first model (i.e., null model) will consist of the pain trajectory model with the addition of patient characteristics and operative characteristics. The second nested model will add medication assignment (gabapentin versus placebo), cognitive style (catastrophizing ↔ optimism latent score), and resting pupil diameter, and all two-way interactions and three-way interaction between these 3 predictors (i.e., 3 two-way, and 1 three-way interaction for both intercept and slope parameters) from the first nested model. This approach allows a single likelihood ratio test of the primary hypothesis and will be interpreted using a two-tailed hypothesis test using $p < 0.05$. A statistically significant difference in the likelihood ratio test will be interpreted as evidence to support the primary hypothesis in that the interaction of these predictors impacts some aspect of the change process beyond that of the basic model. We have planned several sensitivity analyses using multiple imputation to account for missing data.

Several exploratory secondary analyses of the data set:

1. Trajectories for daily activity, as measured via actigraphy, disability, as measured by daily diaries and weekly WHODAS 2.0 scores, and opioid self-administration, as measured by daily diaries, will be analyzed using a similar approach. As noted, other questionnaire and cognitive game scores will be described for the two groups but not analyzed or entered as covariates for the purpose of the primary analysis and publication. Changes in questionnaire data will be evaluated using the generalized linear model regressing follow-up score on baseline score and medication group assignment.

2. The contribution of pupil diameter change in response to acute preoperative pain testing to prediction of trajectory of change in postoperative pain intensity rating will be assessed. This will be performed in the same manner as the primary analysis. In addition, a 4 way-interaction test using the same approach as in the primary analysis will be used by included both pupil diameter at rest and its response to preoperative acute pain stimuli.

3. Development and internally validation a predictive model for recovery from pain from the full battery of preoperative questionnaires and application of item response theory to create a much smaller set of questions with retained predictive value for subsequent testing as a practical tool.

4. Modeling recovery of function from the World Health Organization Disability Assessment Score 2.0 (WHODAS 2.0), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and daily activity level as assessed using actigraphy in order to assess their relationships and those between these functional measures and daily pain, including lead-lag and influence of days of physical therapy, during recovery from physical therapy after lower extremity major total joint arthroplasty.

5. Examining the influence of anesthetic technique (general anesthesia, regional anesthesia via single injection regional anesthesia with sustained catheter infusion) on perioperative acute pain and pain and functional recovery over the subsequent two months.

6. Exploring biomarkers for stress and noradrenergic functioning by exploring the relationship between preoperative patient characteristics, including psychophysical responses to pain processing and CSF concentrations of norepinephrine and markers associated with stress.

27c. Methods to be used for assumptions to be checked for statistical methods for primary and secondary analyses

The assumptions underlying parametric modeling will be evaluated using regression diagnostic procedures and histograms of residuals.

27d. Alternative methods to be used if distributional assumptions do not hold for primary and secondary analyses

If necessary, the primary outcome distribution will be log transformed to satisfy assumptions. For secondary outcomes, the generalized linear model allows the specification of an array of distributions (e.g., normal, binomial, gamma) with corresponding link functions (e.g., log).

27e. Planned sensitivity analyses

Four sensitivity analyses will be performed to evaluate the robustness of the trajectory analysis to the primary and secondary analyses. These are to include multiple imputation for missing data, to include baseline measures of primary and secondary outcomes as predictors in the trajectory models, to include monthly through 6 months and the 12-month daily diary data and to include patient race and surgeon as other causes of variability in trajectory in the primary analysis.

27f. Planned subgroup analyses

No subgroup analyses are planned.

28. Missing data – reporting and assumptions/statistical methods to handle missing data

Multiple imputation will be used to account for embedded missing data in daily diary entries. The autocorrelation of each time-series for each individual along with baseline predictors will be used in conjunction with the MICE algorithm to estimate m = 20 imputations. These imputations will then be re-analyzed using the primary model with Rubin's rules used to pool the estimates of model parameters.

29. Details of any additional statistical analyses required e.g., complier-average causal effect analysis.

None

30. Harms

The number (and percentage) of patients experiencing each AE/SAE will be presented for each treatment arm categorized by severity. For each patient, only the maximum severity experienced of each type of AE will be displayed. The number (and percentage) of occurrences of each AE/SAE will also be presented for each treatment arm. No formal statistical testing will be undertaken

31. Statistical Software

The analysis will be carried out using R version 4.1.2

32. References

There are no non-standard statistical methods used, but references to methods used are listed below. As regards the Data Management Plan, data handling and cleaning were provided at Wake Forest School of Medicine and Massachusetts General Hospital, where data are secured. The Trial Master File is included in Investigator New Drug (IND) 107166, Food and Drug Administration. The Statistical Master File materials are housed at Wake Forest School of Medicine. Standard Operating Procedures followed when writing the SAP are from the guidelines published in JAMA in 2017.[3]

Cited References

1. Houle TT, Miller S, Lang JE, Booth JL, Curry RS, Harris L, Aschenbrenner CA, Eisenach JC: Day-to-day experience in resolution of pain after surgery. Pain 2017; 158: 2147-2154
2. Booth JL, Sharpe EE, Houle TT, Harris L, Curry RS, Aschenbrenner CA, Eisenach JC: Patterns of recovery from pain after cesarean delivery. Pain 2018; 159: 2088-2096
3. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E: Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. Jama 2017; 318: 2337-2343

Appendix

SAP 2.1

The SAP was amended on Feb 6, 2023 to correct an error in the primary analysis in SAP 2.0 which stated that pre-surgery pupil response to acute experimental pain would be entered into a three way interaction with study drug and preferred cognitive style to predict modeled recovery from pain after lower extremity total joint arthroplasty. As stated in the NIH application which funded the research, the IRB approved protocol for the study and its analysis, and the ClinicalTrials.gov registration, the primary analysis was stated to use pre-operative resting pupil diameter rather than its response to pain. The error in the SAP was observed prior to performing the primary analysis and is described in this amendment.

Addition of pre-surgery pupil response to acute experimental pain as a 4^{th} interacting factor was inferred in the grant application and protocol and the SAP is amended to include this as a pre-planned secondary analyses from this extensive dataset, in addition to creating a unique subset of questionnaire questions to predict postoperative pain intensity trajectory, assessing the day to day time relationship between pain and disability during recovery, the role of anesthetic technique in long term recovery, and interrogating various biomarkers for speed of trajectory of recovery after surgery.

SAP 2.2

The SAP was amended a second time on April 26, 2023 to correct an error in the primary and secondary analyses to remove the statements that modeled trajectory of change in outcome measures for each analysis (pain in the primary; daily steps, disability scores, and opioid use in the secondaries) would be adjusted for preoperative values of these measures. Such adjustments were not included in either the IRB approved protocol nor the NIH application which funded the study and its analysis.

There is a large literature demonstrating a positive correlation between the extent of pain catastrophizing with pain, disability, and opioid use and a negative correlation with activity. Since the cognitive construct to be used in both the primary analysis and the secondary analyses is derived from the pain catastrophizing instrument, adjustment of modeled trajectories by preoperative pain, activity disability, and opioid use would be expected to diminish the independent effect of the cognitive construct in these analyses. The adjustments in modeling using these preoperative variables were intended as sensitivity analyses rather than testing the primary and secondary hypotheses. This mistake was identified and this amendment was prepared and submitted prior to performing analyses using these adjustments.