Improving Spine Surgical Care with Real-Time Objective Patient Tracking Using the Apple Watch

Statistical Analysis Plan

NCT04379921

November 26, 2025

Statistical Analysis Plan (SAP)

1. Study Overview

1.1 Background/Introduction

Spine surgery is among the most frequently performed surgical procedures in the United States, with significant implications in patient outcomes. A critical challenge in evaluating patient outcomes is highly depends on patient-reported outcome measures (PROMs), which can be subjective and influenced by psychiatric comorbidities.

The Emerging of consumer-grade wearable technology has the potential to support objectivity, real-time metrics of recovery in clinical research and patient management. Here we conducted the first prospective, non-blinded, randomized controlled trial (RCT) for patients with spine surgery.

Participants receive standard care, and an Apple Watch to record activity through the NeuroCoach App, recording patient's mobility information (e.g., step counts, distance traveled) as well as provide an additional platform for patients to complete questionnaires.

1.2 Study Aims
Aim 1: Correlation Between Objective Patient Measures (Steps, Distance Travelled, From Apple Watch) and Patient-Reported Outcome Measures (SF-36, EQ-5D, PROMIS, NDI, ODI)

Aim 2: Change in Objective Outcome Measures: Number of Steps After Surgery.

Aim 3: Patient Compliance with wearing Apple Watch - Wear Time.

Aim 4: Patient Satisfaction with their Spine Care.

2. Study Population

This prospective randomized controlled study will enroll 187 patients undergoing elective spine surgery at a single academic medical center, with participants randomized to intervention (Apple Watch monitoring, n=96) and control (standard care, n=91) groups. The intervention group will have a mean age of 61.8 years (SD=13.3) and body mass index of 29.3 kg/m² (SD=5.6), while the control group will have comparable demographics with mean age of 59.0 years (SD=13.2) and BMI of 28.4 kg/m² (SD=5.7). The cohort will include patients undergoing both cervical and lumbar spine procedures, with diagnoses primarily comprising degenerative spinal conditions including cervical and lumbar spinal stenosis, spondylolisthesis, radiculopathy, disc herniation, and myelopathy. Common surgical procedures will include anterior cervical discectomy and fusion (ACDF), posterior cervical fusion, lumbar laminectomy, transforaminal lumbar interbody fusion (TLIF), and multi-level posterior lumbar fusion with instrumentation. Patients will be monitored from 2-6 weeks pre-operatively through 12 months post-operatively, with standardized patient-reported outcome measures collected at each timepoint. Of the 96 patients assigned to the intervention group, we anticipate that 91 (94.8%) will generate usable Apple Watch activity data for analysis, reflecting high compliance with wearable device monitoring throughout the one-year post-operative follow-up period.

3. Outcomes, Exposures, and Additional Variables for Interest

3.1 Primary Outcome(s)

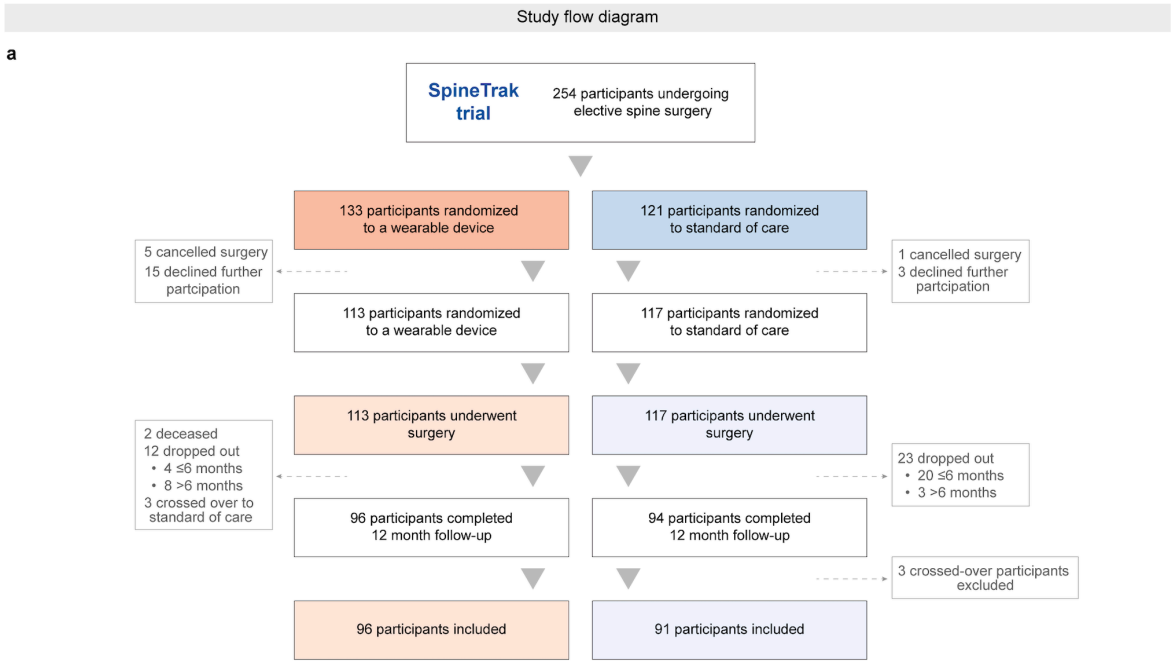
Outcome	Description	Variables and source	Specifications
Daily steps, daily distance, standing time	Objective Measure from Apple Watch.		

total SF-36 score, total EQ-5D score, total PROMIS score, ODI/NDI score (only one available per patient; ODI for back patients, NDI for neck patients)	Subjective Measure through Patient survey response.		

4. Statistical Analysis Plan

4.1 Study cohort:

Figure 1 Consort diagram



4.2 Sample Size & Power Calculation

The correlation analyses will include 91 patients from the intervention group who have both Apple Watch activity data and patient-reported outcome measures. PROM assessments will be collected at five timepoints: pre-operatively (2-6 weeks before surgery) and post-operatively at 6 weeks, 3 months, 6 months, and 12 months, with 96, 96, 94, 92, and 96 total assessments respectively. Due to variations in data availability across different Apple Watch metrics and PROM instruments at each timepoint, actual sample sizes used in individual correlation tests will range from 68 to 90 patients. The pre-operative period will have the highest average sample size (mean n=88.4 across metrics), while the 12-month post-operative timepoint will have the lowest (mean n=71.2). Power calculations for detecting medium effect sizes (Pearson's $r \geqslant 0.3$) at $\alpha = 0.05$ with 80% power indicate that a minimum sample size of approximately 85 participants is required. With our sample sizes ranging from 68 to 90 patients, we will have adequate power to detect correlations of $r \geqslant 0.30$-0.33 between Apple Watch metrics and patient-reported outcomes. The patient satisfaction analysis will include a larger cohort of 187 total patients comprising both intervention (n=96) and control (n=91) groups. Note that the intervention group for satisfaction surveys (n=96) will be slightly larger than the subset with usable Apple Watch data (n=91), as some patients enrolled in the Apple Watch arm may not generate sufficient wearable device data for inclusion in the objective activity analyses.

4.4 Methods used to analyze the outcome measures

We will examine the relationship between objective Apple Watch activity metrics and patient-reported outcomes by conducting Pearson correlation analyses at each of the five study timepoints (pre-operative, 6 weeks, 3 months, 6 months, and 12 months post-operative). For each PROM assessment, we will average corresponding Apple Watch metrics over the 28 days preceding the assessment date for post-operative timepoints, or over the entire pre-operative monitoring period (2-6 weeks before surgery) for baseline assessment. We will analyze Apple Watch metrics: steps, distance traveled (meters), all normalized per hour of device wear time to account for variations in compliance. We will calculate Pearson correlation coefficients with 95% confidence intervals between each normalized metric and five primary patient-reported outcome measures: SF-36 Physical Component Score (PCS), EQ-5D utility score, PROMIS Global Health Score, Neck Disability Index (NDI), and Oswestry Disability Index (ODI). For each metric-outcome pair at each timepoint, we will compute correlation coefficients along with linear regression models to assess the strength and significance of associations. To control for multiple comparisons across the 50 correlation tests (5 timepoints × 2 metrics × 5 outcomes), we will apply the Benjamini-Hochberg false discovery rate (FDR) correction method with $\alpha = 0.05$. Statistical significance will be determined based on FDR-adjusted p-values, and results will be visualized using correlation heatmaps with annotations indicating significant associations.

We will quantify changes in objective activity measures following spine surgery by calculating individual patient-level changes from pre-operative baseline for ten Apple Watch metrics: wear time, steps, distance (meters). For each patient, we will first average daily values within each study timepoint (pre-operative 2-6 weeks before surgery, and post-operative assessments at 6 weeks, 3 months, 6 months, and 12 months). We will then compute individual change scores by subtracting each patient's pre-operative baseline value from their values at subsequent post-operative timepoints through a left join operation followed by iterative subtraction for each metric. Group-level trajectories will be summarized by calculating the mean change and standard deviation at each timepoint across all patients. This approach will allow us to characterize both the average recovery trajectory and the inter-individual variability in objective activity measures throughout the one-year post-operative period, with positive changes indicating improvement (increased activity) relative to pre-operative baseline. Results will be presented in separate matrices for mean changes and standard deviations.

We will assess patient compliance with Apple Watch wearing by analyzing daily wear time across all study timepoints. For each patient, we will group data by patient and timepoint, then average daily wear time values within each timepoint period (pre-operative and post-operative at 6 weeks, 3 months, 6 months, and 12 months). Group-level compliance will be summarized by calculating the mean wear time and standard deviation at each timepoint across all patients with mean and sd calculations. This will provide both a measure of average adherence and variability in device usage throughout the study period. Results will be saved in separate CSV files for mean wear time and standard deviation values to facilitate interpretation of compliance patterns over time.

We will assess patient satisfaction with spine care using a single-item question asking patients to rate their satisfaction on a 5-point Likert scale ("Not at all satisfied," "Somewhat satisfied," "Neutral," "Very satisfied," "Extremely satisfied") at four post-operative timepoints: 6 weeks, 3 months, 6 months, and 12 months. We will extract data from the REDCap database for both intervention (Apple Watch) and control groups, consolidating responses across different survey versions administered during the study period using the coalesce function to handle variable naming changes. For each group and timepoint, we will calculate the frequency distribution of responses across all five satisfaction categories using table functions, then compute percentages by dividing individual response frequencies by the total number of responses at each timepoint. This descriptive analysis will allow comparison of satisfaction trajectories between intervention and control groups throughout the one-year follow-up period, providing insight into patient perceptions of their overall spine care experience and potential differences attributable to Apple Watch monitoring.