Statistical Analysis Plan

The effect of high intensity interval training and surgical weight loss on distal symmetric polyneuropathy outcomes.

Principal Investigator	Brian C. Callagahan, MD, MS
	Professor of Neurology
	University of Michigan
	Department of Neurology,
	Ann Arbor, MI, 48104, USA
ClinicalTrails.gov identifier	NCT03617185
Author	Evan L. Reynolds, Ph.D.
	Assistant Professor of Biostatistics
	Michigan State University
	Department of Epidemiology and Biostatistics
	East Lansing, MI, 48824, USA
	Lavanya Muthukumar, MS
	Senior Statistician
	University of Michigan
	Department of Neurology,
	Ann Arbor, MI, 48104, USA
	Fallon Koenig, MS
	Clinical Research Coord Inter
	University of Michigan

Department of Neurology,
Ann Arbor, MI, 48104, USA

Version	Draft 0.2
Date	August 14th, 2026

Table of contents

Background and rationale

Distal symmetric polyneuropathy (DSP) affects nearly 15% of Americans over the age of 40. The prevalence in diabetes, the most common cause, is ~30%, and in the morbidly obese is 23%. Patients with DSP have lower quality of life, more pain, and fall more frequently than those without this disease. In diabetes, lower extremity amputations are 2.6 times more likely in those with DSP. Despite the substantial morbidity associated with this highly prevalent condition, DSP patients have limited therapeutic options. Some medications can reduce pain, but they are not disease modifying treatments. For patients with diabetes, better glucose control results in a large risk reduction of DSP in type 1 diabetes, but large, randomized trials failed to show a statistically significant risk reduction in type 2 diabetes. For obese patients, no proven treatments are available. Therefore, there is a *critical need* to develop disease modifying treatments for DSP patients with diabetes and/or obesity.

Multiple studies reveal an association between the metabolic syndrome and DSP. We further defined the contributions of individual metabolic syndrome components in three separate epidemiologic studies, including our ongoing NIH K23 study. While diabetes is the best established risk factor for DSP, our data support obesity and pre-diabetes as important metabolic drivers of nerve injury. The most effective therapies for obesity and pre-diabetes are exercise and/or weight loss. While the combination of these two vastly different interventions is likely optimal, each requires a life-altering change in behavior. Therefore, the *primary objective of this application* is to determine the effect of high intensity interval training on DSP outcomes. The secondary objective is to determine the effect of sleeve gastrectomy on DSP outcomes. Our *central hypothesis* is that exercise and surgical weight loss will both improve DSP outcomes, but

the effect of exercise will be significantly larger. The *rationale* for this hypothesis is our preliminary data revealing stable DSP outcomes 2 years after diet-induced weight loss (the natural history is worse DSP outcomes over time), while other groups show a substantial improvement in DSP outcomes after exercise in subjects with minimal weight loss. Limitations of previous work include the lack of a comparison group for the intervention. Likewise, the effect of surgical weight loss on DSP has not been rigorously tested.

Study design

We performed a randomized controlled trial of participants attending bariatric surgery clinics Participants undergoing sleeve gastrectomy (SG) or choosing not to pursue SG were randomized 1:1 to receive a high intensity interval training (HIIT) intervention or routine exercise counselling (RE), using a block randomization technique stratified by type 2 diabetes (T2D) (normoglycemia, prediabetes, T2D) and surgery status (SG or no). Participants were recruited from Michigan Medicine, Henry Ford Health System, and Trinity Health.

Inclusion criteria:

Inclusion criteria were attendance at a surgical weight loss clinic, age ≥40 years, body mass index [BMI]>35 kg/m^2 with a weight-related comorbid condition such as hypertension, hypercholesterolemia, sleep apnea, asthma, or heart disease, or BMI>40 kg/m^2 without a comorbid condition and being willing and capable to provide informed consent and to accept treatment assignment.

Exclusion criteria:

Exclusion criteria included having an abnormal cardiovascular stress test, alcoholism, anticoagulant therapy use, autoimmune disorders, active cancer within the last year, cardiac stent

within the last year, chemotherapy use, chronic kidney disease, cirrhosis of the liver, cognitive impairment, congestive heart failure class III, corneal transplant, current tobacco use, hepatitis B, hepatitis C, use of high dose steroids, history of organ transplant, history of open Nissen surgery or esophagectomy, human immunodeficiency virus, lupus, marijuana use, nicotine use, rheumatoid arthritis, sarcoid, Sjogren's, suicide attempt in the last year or multiple suicide attempts, thyroid-stimulating hormone, tuberculosis, use of a walking assistance device, vitamin B12 deficiency, vitamin B1 deficiency, weight>450 lbs, and others on a case by case basis.

HIIT intervention:

Two exercise facilities were used for supervised in-person training sessions. HIIT training was provided at no cost. There were multiple trained study personnel who offered exercise training, each certified with Basic Life Support. Exercise sessions were completed one-on-one, and each participant was assigned to a single trainer throughout the study. Trainers scheduled supervised visits based on the participants' availability, and participants completed their one weekly unsupervised session on their own time. Those that received sleeve gastrectomy did not begin sessions sooner than four weeks after the procedure to allow adequate time for healing and recovery. Heart rate was collected using a Polar device (Kempele, Finland).

Supervised tele-exercise training sessions conducted during the COVID-19 pandemic could take place at home or in any setting of the participant's choosing. Once COVID-19 restrictions were lifted participants were given the option to return to in person training, but many preferred to remain virtual for the remainder of the study.

Compensation increased during the study: $50 for the first 8 weeks of follow-up to $160 for the last 8 weeks of follow-up. If attendance was not perfect, compensation was given proportional to the number of sessions attended out of 24 (the maximum number per each 8-week interval).

RE:

At baseline, participants were counseled to participate in 60 minutes of aerobic exercise daily in addition to 2-3 non-consecutive days of strength training workouts every week and encouraged to obtain a gym membership, purchase exercise equipment, and/or join a walking group or fitness class. Participants were compensated for home exercise equipment or gym membership.

SG:

SG was completed within 6 weeks of the baseline assessment and randomization occurred 4 weeks after the completion of SG. Participants completing SG were supplemented with a multivitamin, sublingual B12, and calcium as part of their routine post-operative care. Furthermore, participants had laboratory monitoring including a comprehensive metabolic panel, complete blood count, folic acid, B12, TSH, ferritin, and vitamin D as part of their routine clinical care.

Timing of outcome assessments after SG and randomization:

Participants completed baseline assessments prior to SG and randomization. For those that did not undergo SG, to maintain consistency in time from baseline assessment to randomization, randomization occurred between 4-10 weeks after baseline assessments.

Metabolic risk factors:

During follow-up visits, we assessed height, weight, BMI, and nine anthropometric measurements. Blood pressure was assessed in a supine position. Laboratory values were taken after an overnight fast, and included total cholesterol, triglycerides, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, fasting glucose, fasting insulin, and hemoglobin A1c. T2D was determined based on the 2025 American Diabetes Association definition.[9] Participants with a previous diagnosis of T2D or taking T2D medications were considered to have T2D for stratification purposes. We recorded medications for T2D, hypertension, hypercholesterolemia, and obesity.

Randomization and blinding:

A study statistician (E.R.) performed the randomization using R statistical software and was the only individual to have access to the random allocation sequence. Outcome assessors were blinded to intervention, but the participants, data analysts, and the treatment team were not blinded. If outcome assessors became unblinded, then a new outcome assessor was assigned.

Follow-up details:

Participants were followed for two years with assessments completed at baseline, prior to SG and randomization, and 3-, 12-, and 24-months following randomization. Due to the COVID-19 pandemic, the study was temporarily paused from 3/13/2020-4/3/2020 with virtual study visits until 7/2020

Changes to the trial design after it commenced:

Of note, the initial protocol was amended to include Henry Ford Health System and Trinity Health as additional sites to aid recruitment and generalizability. To limit confounding, we

required participants to be eligible to undergo SG according to the bariatric surgery team and start the pre-surgical process. Initially we required full bariatric surgery committee review approval but changed this requirement to aid recruitment

Withdrawal:

140 participants were recruited and 109 participants completed the study. Reasons for withdrawal from the study included moving out of the area (n=3), changes in life circumstances (n=4), scheduling conflicts (n=2), dislike of HIIT sessions (n=2), being randomized to the routine exercise counseling arm but wanting HIIT training (n=2) or being lost to follow-up (n=18).

Aims and objectives

Primary aim: Determine the effect of high-intensity interval exercise training (HIIT) versus routine exercise counseling (RE) on change in peripheral neuropathy (PN) for persons with obesity.

Secondary aim: Determine if the effect of HIIT on PN varied by SG and separately determine the effect of SG on PN. Additionally, access the effect of HIIT and/or SG on other outcomes including other PN outcomes, cognition, retinopathy, chronic kidney disease, pain, and quality of life.

Hypothesis 1: We hypothesized that HIIT would improve PN outcomes compared with RE. We further hypothesized that SG would improve PN outcomes without a synergistic effect.

Outcomes

Primary outcome:

Our primary outcome was intraepidermal nerve fiber density (IENFD) measured at the proximal thigh. Thigh IENFD was assessed using a 3 mm skin punch biopsy 20 cm distal to the anterior superior iliac spine. Skin biopsies were fixed in Zamboni's fixative for 12-24 hours and then cryoprotected in phosphate buffered 20% glycerol prior to shipping to University of Rochester for processing. The skin biopsies were sectioned at 50 μm thickness using a sliding freezing microtome and 4 sections/biopsy (selected using systematic random sampling) were stained with monoclonal antibodies to human PGP 9.5 (BIO-RAD, Hercules, California) according to previously described brightfield techniques.[39–41] Individual nerve fibers that crossed into the epidermis in each of the 4 sections were counted by a blinded examiner using an established protocol, summed, and the divided by the length of the epidermis to determine the IENFD (fibers/mm) for each biopsy.[42]

Secondary outcome:

IENFD at the distal leg: In addition to the IENFD measure at the proximal thigh, we will also measure this entity at the distal leg. Previous data support that the proximal thigh measurement is more likely to show change over time after an exercise regimen in a population of patients with pre-diabetes.[30] However, IENFD at the distal leg will provide additional data as to the severity and length-dependence of the nerve injury. Furthermore, measurement at the distal leg may be more sensitive to change in those patients with obesity and normoglycemia, which may have milder nerve injury than those with diabetes.

Corneal confocal microscopy (CCM): While IENFD is the current gold-standard for detection of

small fiber nerve injury, CCM is an innovative approach that also attempts to measure small fiber nerve injury in the cornea. This procedure is less invasive than the skin biopsy required for IENFD measurement. The benefit of including this secondary outcome measure is that we will be able to compare and contrast the diagnostic characteristics of IENFD to CCM and also determine which measure is more sensitive to change over time. This data will help determine which of these two measurements should be utilized in future intervention studies. Specifically, we will measure nerve fiber density, nerve branch density, and nerve fiber length. A recent meta-analysis revealed that all three of these parameters are reduced in those with diabetic DSP compared to those with diabetes without DSP and to healthy controls.[92] Furthermore, studies have revealed similar diagnostic characteristics for CCM and IENFD for diabetic DSP.[93] CCM parameters also improve after simultaneous pancreas and kidney transplants, even when IENFD remains unchanged, indicating that CCM may be a useful early marker of nerve regeneration.[94] Consultant Roni Shtein, MD, Associate Professor of Ophthalmology and Visual Sciences, has extensive experience utilizing CCM as an outcome measure.

Electrophysiological measures: a) NCS of the sural, peroneal, and tibial nerves will be obtained following the DCCT/EDIC protocol.[95, 96] Briefly, measurements will be done at the dominant limb, using temperature control.[95, 96] NCS provide a quantitative, reproducible way to evaluate for large fiber function. b) Measures of CAN will comprise cardiovascular reflex testing using the DCCT/EDIC protocol (R-R interval ratio during deep breathing, postural change, and Valsalva) as described.[97, 98] These tests are recommended by the Toronto Consensus Panel on Diabetic Neuropathy[99] as gold-standard for CAN. We will also assess resting HR and HR variability. Consultant Dr. Busui has extensive expertise in performing and interpreting CAN

outcomes in clinical trials. c) Sudomotor function will be assessed by changes in the electrochemical skin conductance by Sudoscan™ (Impeto Medical, Paris, France) as described.[100, 101]

Clinical DSP measures: a) Structured neurological examination following the DCCT/EDIC protocol[95, 96]; b) Michigan Neuropathy Screening Instrument (MNSI) questionnaire and examination, validated instruments to identify diabetic DSP[102, 103]; c) Utah Early Neuropathy Scale (UENS), a validated tool to identify small fiber predominant neuropathy such as that found in obese and pre-diabetic individuals[104]; d) Modified Toronto Neuropathy Score (mTNS), a validated instrument for the detection of mild to moderate diabetic DSP[105]; and e) Survey of Autonomic Symptoms (SAS), a validated measure of autonomic symptoms[106]

DSP-specific pain: Pain will be assessed using the Short Form McGill Pain Questionnaire. This questionnaire is a reliable and validated pain questionnaire that utilizes a visual analogue scale, a 6 point rating scale of pain intensity, and a 4 point rating scale of 15 different neuropathic pain descriptors.[107]

Quality of life measures: We will use the NeuroQOL, a validated measure of quality of life specific to DSP,[108] which comprises 6 domains (pain, lost/reduced feeling, diffuse sensory-motor symptoms, restrictions in activities of daily living, disruptions in social relationships, and emotional distress) along with an overall measure of DSP impact on quality of life and general quality of life.[108]

Functional scales: a) The Berg Balance Scale, measures balance in 14 separate activities of daily living (the unipedal stance portion has been shown to be particularly reflective of DSP-related mobility loss)[109]; b) The 8 Foot Up and Go Test, a test of functional mobility that assesses the time needed for a subject to arise from sitting position, walk 8 ft and turn 180 degrees around a cone, and return to sitting[110]; c) The Modified Falls Efficacy Scale, assessing patient's self-reported ability to perform activities of daily living (e.g. get dressed/ undressed, walking, shopping)[86, 111, 112]

Other secondary outcomes included NIH Toolbox Cognition Battery's composite score [NIHTB-CB], estimated Glomerular Filtration Rate [eGFR], 24-2 FDT [Frequency Doubling Technology] - Mean Deviation [MD], and the Expiration/Inspiration ratio [E/I ratio].

Safety:

Patients were instructed to notify study team of any safety concerns for the duration of the study including during each study visit. Safety outcomes were summarized as number of adverse events (AEs).

Statistical analysis

Our primary analysis assessed the effect of HIIT versus RE on PN outcomes. In secondary analysis, we assessed the effect of HIIT with and without SG on PN outcomes. Thus, in primary analysis, there were two study groups: RE (with and without SG) and HIIT (with and without SG). In secondary analysis, there were four study groups (1) RE without SG, (2) HIIT without SG, (3) RE with SG, and (4) HIIT with SG.

All models were adjusted for the stratification variables during randomization type 2 diabetes status (normoglycemia *[reference group],* pre-diabetes, type 2 diabetes) and recruitment sites (Michigan Medicine *[reference group]*, Henry Ford Health System, and Trinity Health).

Descriptive analysis:

Descriptive statistics were used to summarize participant demographic information, anthropometric measurements, metabolic risk factor information, primary outcomes, and secondary outcomes, at baseline for HIIT versus RE groups. We also calculated descriptive statistics to summarize study variables across HIIT and RE, stratified by SG: RE without SG, HIIT without SG, RE with SG and HIIT with SG.

For each group, within participant change was reported for metabolic phenotyping, primary outcome, and secondary outcomes among those who completed the study. For continuous variables, within participant change was determined by subtracting baseline values from those measured at the 2 years follow up visit and reported as mean (SD). For categorical variables, within participant change was determined as those that improved from baseline to 2 years follow up visit and reported as N (%).

Unadjusted analysis:

Paired t-tests were used to compare within-participant change for continuous variables for the HIIT vs.RE group. In addition, the Wilcoxon signed rank test was used to compare within-participant change for ordinal categorical variables for the HIIT vs.RE groups. Within-participant change was also reported based on the 4-group comparison (RE without SG, HIIT without SG, RE with SG, HIIT with SG).

Welch two sample t-tests (for continuous variables) and Pearsons's chi-squared test of independence (for categorical variables) were used to compare the between group differences in within-participant change between HIIT vs. RE. Additionally, we assessed differences in within-participant change for each study variable using the 4-group comparison. Specifically, we compared within-participant change between those in the RE without SG group to those in the HIIT without SG, RE with SG, and HIIT with SG groups, separately, using two sample t-tests (for continuous variables) and Pearsons's chi-squared test of independence (for categorical variables).

Adjusted analysis:

Primary analysis: Effect of HIIT on IENFD of the proximal thigh

Our primary analysis fit a linear mixed effects model to determine the association between HIIT and IENFD of the proximal thigh. Specifically, we fit repeated assessments of IENFD of the proximal thigh, measured at baseline prior to randomization, and 3, 12, and 24 months following randomization as a function of treatment group (HIIT vs. RE), follow-up time, and an interaction parameter between treatment group and follow-up time. The model also included a random participant intercept, allowing us to account for the repeated assessments for each participant. The mixed effects model was adjusted for baseline diabetes status (type 2 diabetes, pre-diabetes, normoglycemia), site of recruitment (Michigan Medicine, Henry Ford Health System, Trinity Health), and SG (yes versus no). Model formula:

$$Y_{ij} = \beta_0 + \beta_{1a}HFHS_i + \beta_{1b}TH_i + \beta_{2a}prediabetes_i + \beta_{2b}diabetes_i + \beta_3\, time_i + \beta_4 SG_i + \beta_5 HIIT_i + \beta_6 HIIT_I\ x\ time_i + b_{0i} + \varepsilon_{ij}$$

Where Y_{ij} represents the thigh IENFD measurement for the ith study participant at the jth timepoint, b_{0i} represents the random intercept for participant i, ε_{ij} represents the residual error for participant i at timepoint j, $HFHS_i$ represents recruitment from Henry Ford Health System (1 vs. 0), TH_i represents recruitment from Trinity Health (1 vs. 0), $prediabetes_i$ indicates the presence of prediabetes at baseline (1 vs. 0), $diabetes_i$ indicates the presence of diabetes at baseline (1 vs. 0), $time_i$ indicates the follow up time in months (0 , 3, 12, 24), SG_i indicates if SG was completed (1 vs. 0), $HIIT_i$ indicates if patients were enrolled in HIIT vs. RE (1 vs. 0). In addition, β_0 represents the mean thigh IENFD for RE participants at baseline, who did not have SG, had normoglycemia, and were recruited from Michigan Medicine. Further, β_{1a} represents the mean difference in thigh IENFD between those recruited in Henry Ford Health System , compared to those recruit in Michigan Medicine, β_{1b} represents the mean difference in thigh IENFD between those recruited in Trinity Health compared to Michigan Medicine, β_{2a} represents the mean difference in thigh IENFD between those who had prediabetes compared to those with normoglycemia, β_{2b} represents the difference between those who had diabetes compared to those with normoglycemia, β_3 represents the mean change in thigh IENFD per month, among RE participants, β_4 represents the mean difference in thigh IENFD for participants who had SG, compared to those who did not have SG, β_5 represents the mean difference in thigh IENFD at baseline for HIIT participants, compared to RE participants. The interaction parameter (β_6) is the primary parameter of interest and allows us to assess whether the change in IENFD of the proximal thigh over time differs between the HIIT and RE participants. A 95% confidence interval for the interaction parameter that excludes 0 will indicate a statistically significant association between HIIT and IENFD.

Secondary analysis 1: Effect of HIIT on secondary outcomes

We fit additional mixed effects models to determine the association between HIIT and all secondary outcomes using repeated assessments measured at baseline prior to randomization, and 3, 12, and 24 months following randomization. Similar to the primary analysis of our primary outcome, we fit each secondary outcome as a function of treatment group (HIIT vs. RE), follow-up time, and an interaction parameter between treatment group and follow-up time. Random participant intercepts allowed us to account for the repeated assessments for each participant. All mixed models were adjusted for baseline diabetes status, site of recruitment, and SG.

For IENFD of the leg, log transformed McGill Pain Questionnaire (scale 0-10), NIHTB-CB, eGFR, 24 FDT MD and E/I ratio, we fit linear mixed models as these outcomes were continuous and met the necessary assumptions for linear mixed effects models. Given that responses to the MNSI questionnaire, MNSI examination, neuropathy-specific QOL, UENS and mTCNS, were ordinal counts, we instead fit Poisson mixed effect models. Zero-inflated regression was used for UENS and mTCNS due to the presence of a larger number of zeros. A 95% confidence interval for the interaction parameter that excludes 0 (for linear mixed effects models) and 1 (for Poisson mixed effects models) determined whether there was a statistically significant association between HIIT and secondary outcomes.

Secondary analysis 2: Assessing if the effect of HIIT on study outcomes differed based on completion of SG.

As a secondary analysis, to determine the effect of HIIT, SG, and HIIT with SG on study outcomes over time, we fit additional mixed effects models using quasi-randomization via propensity score adjustment. Propensity scores adjustment allowed us to account for baseline

differences in demographic information between those that received or did not receive SG. Propensity scores were calculated using logistic regression, and represented the likelihood of completing SG, based on age, sex, race, education, marital status, physical activity, smoking, alcohol consumption, and health insurance type.

We fit additional linear mixed effects models for thigh IENFD, and each secondary outcome, as a function of treatment groups (HIIT vs. RE, SG vs. no SG, and an interaction of HIIT and SG), follow-up time, and an interaction parameter between treatment groups and follow-up time. Random participant intercept included in the model allowed us to account for the repeated assessments for each participant. All models were adjusted for baseline diabetes status and recruitment site. Model formula: $Y_{ij} = \beta_0 + \beta_{1a}HHFS_i + \beta_{1b}TH_i + \beta_{2a}prediabetes_i + \beta_{2b}diabetes_i + \beta_3\, time_i + \beta_4\, propensity\ score\ for\ surgery\ risk_i + \beta_5 SG_i + \beta_6 HIIT_i + \beta_7 HIIT_i\ x\ time_i + \beta_8 SG_i\, x\, time_i + \beta_9 HIIT_i\, x\, SG_i\ x\ time_i\ \ b_{0i} + \varepsilon_{ij}$. The interaction parameters $(\beta_7 - \beta_9)$ are the primary parameters of interest and allow us to assess whether the change in outcome over time differs following HIIT without SG (compared to RE without SG), RE with SG (compared to RE without SG), and HIIT with SG (compared to HIIT without SG). A 95% confidence interval for a interaction parameter that excludes 0 (for linear mixed effects models) or 1 (for Poisson mixed effects models) will indicate a statistically significant association between HIIT/SG/HIIT with SG, and outcomes.

Sensitivity analysis 1: additional adjustment for circumferences at the skin biopsy site.

To account for whether changes in thigh IENFD resulted from the changing size of the thigh during follow-up, we re-fit our primary linear mixed effects regression models, that also adjusted for thigh circumference at the skin punch biopsy location.

Sensitivity analysis 2: accounting for exercise compliance:

To determine whether compliance to the HIIT regimen during the course of the study had differential effects on change in our primary outcome, we fit a sensitivity analysis also including compliance to the HIIT intervention as a time-varying covariate into our primary analysis linear mixed effects model. Exercise compliance was calculated by adding the total compliant visits within each follow-up period (baseline to 3 months, 3 months to 12 months, 12 months to 24 months) and dividing it by the total number of expected compliant visits in that period. We then categorized compliance into three groups: less than half of HIIT sessions per week, less than one HIIT session per week and one or more HIIT session per week. Compliance was assigned as zero for those not in HIIT intervention group and was used as the reference category.

We then fit repeated assessments of IENFD of the proximal thigh, as a function of HIIT compliance, follow-up time, and an interaction parameter between compliance proportion and follow-up time. The model also included a random participant intercept and was adjusted for baseline diabetes status, site of recruitment, and SG (yes versus no). Model formula: $Y_{ij} = \beta_0 +$

$\beta_{1a}HHFS_i + \beta_{1b}TH_i + \beta_{2a}prediabetes_i + \beta_{2b}diabetes_i + \beta_3\, time_i + \beta_4\, SG_i +$

$\beta_{5a}less\ than\ half\ a\ HIIT/week_i + \beta_{5b}less\ than\ one\ HIIT\ session/week_i +$

$\beta_{5c}one\ or\ more\ HIIT\ session/week_i + + \beta_{6a}less\ than\ half\ a\ HIIT/week_i\ x\ time_i +$

$\beta_{6b}less\ than\ one\ HIIT\ session/week_i\ x\ time_i + \beta_{6c}one\ or\ more\ HIIT\ session/week_i\ x\ time_i +$

$b_{0i} + \varepsilon_{ij}$. The interaction parameters ($\beta_{6a} - \beta_{6c}$) are the primary parameters of interest and allow us to assess whether the change in IENFD of the proximal thigh over time differs by extent of compliance to the HIIT regimen. A 95% confidence interval for the interaction parameter that excludes 0 will indicate a statistically significant association between compliance and IENFD.

Other statistical considerations

Our primary analysis compared effects between HIIT and RE on change in IENFD of the proximal thigh during follow-up, and includes only one primary parameter of interest: the interaction effect between follow-up time and treatment (HIIT vs. RE). Therefore, we did not adjust for multiple comparisons on our primary outcome. In addition, secondary analyses, and analysis of the secondary outcomes, were considered to be hypothesis-generating and exploratory.

For all analyses, two-sided p-values were calculated, and a p-value threshold of 0.05 was used to determine statistical significance.

For each regression model, we performed residual diagnostics to confirm the necessary assumptions were met. Specifically, we used data visualization to confirm linear mixed effects models assumptions of normality of the residual distribution, homoscedasticity of the residual distribution, independence of the residuals, linearity, normality of the random participant intercepts, and homoscedasticity of the random participant intercepts were met. When assumptions were not met, we evaluated whether using simple transformations on the outcome (e.g., log, square root, etc.) were able to confirm model assumptions. We did find that our linear mixed effects model for a secondary outcome, the visual analogue pain scale, was in violation of normality and homoscedasticity of the residual distribution. Thus, we instead fit the model for log (visual analogue pain scale).

All analyses were completed using R version (4.4.3).

A priori power calculations: We determined the sample size needed to achieve 80% power to detect a difference in within-participant differences from baseline to 24-months, between the RE

without SG group and each other group: HIIT without SG, RE with SG, and HIIT with SG groups. With our expected sample size of 30 patients (allowing for ~15% attrition) in each of the four treatment groups, we estimated that we would have at least 80% power to detect differences of between 0.9 and 1.7 fibers/mm after 24 months as statistically significant based on a two-sided test with type 1 error rate of 0.05.

Deviation from the original trial protocol

Given that participants were not randomized to receive SG, upon completion of the study, we found that there were differences in demographic and metabolic risk factors between those that completed and did not complete SG. To account for this limitation, we updated our statistical analysis plan to allow us to focus on the effect of HIIT vs. RE, the randomization variable, on study outcomes. In addition, in our secondary analysis, we utilized a quasi-randomization approach to assess the effect of SG on PN outcomes. Specifically, we calculated propensity scores for receiving surgery, to account for differences in those that received vs. did not receive SG.

Updated power calculations: Given the change to our trial design to focus on the effect of HIIT vs. RE on IENFD thigh, we calculated the statistical power to detect a difference in the change in IENFD thigh after 24 months of follow-up between the HIIT and RE groups. Based on prior literature, we originally estimated that the mean change in IENFD thigh would be -3.0±5.0 for RE without SG, 0.2±7.0 for RE with SG, 1.4±2.3 for HIIT without SG, and 1.6±7.4 for HIIT with SG. We assumed that 60 patients in each group would complete follow-up (allowing for ~15% attrition). Based on the estimated effect sizes from our *a priori* power calculations, we estimated that the mean (SD) change in IENFD thigh is 1.5 (5.4) for HIIT and -1.4 (6.3) for RE.

We determined that we would have 70% power to detect this difference in IENFD between HIIT and RE as statistically significant, based on a two-sample t-test and a type I error rate of 0.05. We also determined statistical power to detect an interaction effect between treatment group and follow-up time as statistically significant, based on a linear mixed effects model. Assuming a sample size of 60 patients per group, a type I error rate of 0.05, and an intraclass correlation coefficient of 0.56, we determined we would have 81% power to detect the interaction effect of HIIT by follow-up time as statistically significant.