Treatment of Meniscal Problems in Osteoarthritis: The TeMPO Trial

Statistical Analysis Plan

Principal Investigator: Jeffrey N. Katz, MD, MSc
Brigham and Women's Hospital

Version Date: March 18, 2024

This document discusses the statistical analyses that will be conducted to establish the comparative efficacy of two in-clinic physical therapy interventions (one focused on strengthening and one containing placebo) and two protocolized home exercise programs. This trial is a four-arm parallel RCT. This plan includes a description of the outcomes, hypotheses, approaches to testing the hypotheses and analyzing outcomes, sample size and power considerations, and safety data.

TRIAL FULL TITLE	Treatment of Meniscal Problems in Osteoarthritis: The TeMPO Trial
SAP VERSION	1.2
SAP VERSION DATE	March 18, 2024
TRIAL STATISTICIAN	Jamie E. Collins, PhD
Protocol Version (SAP associated with)	Treatment of Meniscal Problems in Osteoarthritis: The TeMPO Trial Detailed Protocol 2/11/2021
TRIAL PRINCIPAL INVESTIGATOR	Jeffrey N. Katz, MD, MSc
SAP AUTHOR(s)	Jamie E. Collins, PhD Jeffrey N. Katz, MD, MSc

Contents

1 SAP Signatures

I give my approval for the attached SAP for the Treatment of Meniscal Problems in Osteoarthritis: The TeMPO Trial dated March 18, 2024

Statistician (Author)

Name: Jamie E. Collins

Signature: Jamie E Collins

Date: March 18 2024

Principal Investigator

Name: Jeffrey N. Katz

Signature: Jeffrey N Katz

Date: March 20, 2024

2 Trial Registration

The TeMPO Trial was first registered at clinicaltrials.gov with registration No. NCT03059004. on February 14, 2017.

3 Protocol

A manuscript detailing the TeMPO rationale and design features was published as follows[1]:

Sullivan JK, Irrgang JJ, Losina E, Safran-Norton C, Collins J, Shrestha S, Selzer F, Bennell K, Bisson L, Chen AT, Dawson CK, Gil AB, Jones MH, Kluczynski MA, Lafferty K, Lange J, Lape EC, Leddy J, Mares AV, Spindler K, Turczyk J, Katz JN. The TeMPO trial (treatment of meniscal tears in osteoarthritis): rationale and design features for a four arm randomized controlled clinical trial. BMC Musculoskeletal Disorders. 2018 Dec;19(1):1-4.

4 Summary of Changes to the Protocol or SAP

January 2024:

The COVID-19 pandemic began in the middle of TeMPO enrollment. We have updated the SAP to include a covariate in the multivariable models indicating enrollment date prior to or after the start of the pandemic (enrollment date (prior to or after March 15, 2020)) and to include a sensitivity analysis excluding participants enrolled between January and March 2020.

We added a secondary analytic approach of the primary outcome (KOOS Pain) of baseline-adjusted ANCOVA.

Rules for missing data in computing instruments (KOOS, EQ-5D) were clarified in section 13.4.

March 2024:

A sensitivity analysis excluding participants with protocol violations that affect study eligibility or data integrity was added (section 9.4).

5 Abbreviations and Definitions

ADL	Activities of Daily Living
AE	Adverse Event
APM	Arthroscopic Partial Meniscectomy
KLG	Kellgren-Lawrence Grade
KOOS	Knee Injury and Osteoarthritis Outcome Score
OA	Osteoarthritis
PT	Physical Therapy
RCT	Randomized Controlled Trial

SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
TKR	Total Knee Replacement

6 Introduction

6.1 Background and Rationale

Symptomatic knee osteoarthritis (OA) is a disabling problem affecting over 15 million adults in the US[2], many of whom also have MRI-based evidence of meniscal tear.[3] More specifically, over 30% of persons ≥40 years old have evidence of meniscal tear on MRI[3],including 80% of those with symptomatic knee osteoarthritis (OA).[4] Relatively few tears seen on MRI cause symptoms, but those that are symptomatic are often quite disabling.[3,4] Clinicians typically rely upon a history of mechanical symptoms to identify patients whose meniscal tears appear to be symptomatic, although the diagnostic value of mechanical symptoms is modest.[5] In this trial we will only focus on symptomatic meniscal tear in persons with imaging evidence of OA. We refer to this phenotype as "*symptomatic meniscal tear and OA*" — with "OA" including both radiographic and pre-radiographic MRI evidence of cartilage damage. We focus on these patients rather than those with traumatic tear and no OA because there is more uncertainty about appropriate treatment in patients with OA.

The efficacy of therapist-directed exercise for symptomatic meniscal tear in the setting of OA has not been evaluated with placebo-controlled studies. In fact, a recent literature synthesis by the American Physical Therapy Association identified *no* studies establishing the efficacy of conservative management of symptomatic meniscal tear.[6] This is a key research gap: there is currently no evidence as to whether any PT regimen is superior to placebo in persons with symptomatic meniscal tear and OA. This is a particularly pressing issue, as the lack of evidence supporting the superiority of surgery has more and more doctors turning to physical therapy as the first line of treatment.

In summary, symptomatic meniscal tear in the presence of OA is a prevalent, disabling, costly problem. Therapist-directed exercise and manual therapy have been associated with substantial pain relief in OA[7-12] but have not been studied in symptomatic meniscal tear with OA. Further, we do not know whether the pain relief observed in OA trials is due to the physiologic effects of strengthening and manual therapy, placebo effect due to attention from a therapist; the effect of enrolling in a trial, or natural history. The premise of the TeMPO Trial builds upon rich data in the chronic disease literature that suggests each of these factors may contribute to treatment efficacy. The different aspects of treatment must be distinguished to develop objective clinical and policy recommendations; these factors can only be disentangled with a carefully designed RCT, involving placebo elements.

6.2 Scope of Analysis

These analyses will assess the efficacy of home-based exercise compared to physical therapy among patients who have meniscal tear and osteoarthritis.

7 Study Objectives

7.1 Primary Objective

To conduct a multicenter, parallel, four-arm randomized controlled trial that will establish the efficacy of various components of the typical PT regimen for subjects 45-85 with symptomatic meniscal tear. The four arms will include:

Arm 1: Prescription of an information-based (exercise pamphlet, DVD) home exercise program

Arm 2: An information-based home exercise program (exercise pamphlet, DVD) with adherence optimization

Arm 3: Same as Arm 2 plus a clinic-based, therapist-directed sham intervention, with no known physiologic effect

Arm 4: Same as Arm 2 plus "true PT": clinic-based, therapist-directed exercise and manual therapy

The three primary hypotheses are:

Therapist-directed exercise and manual therapy plus optimized home exercise (Group 4) will lead to greater improvement in KOOS pain at three months than simple prescription of a home exercise program (Group 1).

Therapist-directed exercise and manual therapy plus optimized home exercise (Group 4) will lead to greater improvement in KOOS pain at three months than optimized home exercise and no interaction with a therapist (Group 2).

Optimized home exercises (Group 2) will have a greater effect on KOOS pain at three months than simple prescription of exercises without adherence optimization (Group 1).

7.2 Secondary Objectives

Secondary objectives include testing the additional hypotheses that:

In-clinic placebo PT plus optimized home exercise (Group 3) will lead to greater improvement in KOOS pain at three months than optimized home exercise and no interaction with a therapist (Group 2).

Therapist-directed exercise and manual therapy plus optimized home exercise (Group 4) will lead to greater improvement in KOOS pain at three months than in-clinic placebo PT plus optimized home exercise (Group 3).

Secondary objectives also include assessing secondary outcome measures and longer term durability of the treatment effect.

8 Study Methods

8.1 Trial Design

The TeMPO Trial is a four-arm single-blind multi-center randomized controlled clinical trial. The treatment allocation ratio is 1:1:1:1.

Table 1. Study Arms

Arm	Protocolized Home Exercise Program	Adherence Optimization	Placebo Therapy (14 sessions)	Supervised Exercise & Manual Therapy (14 sessions)
Arm 1	✓			
Arm 2	✓	✓		

Arm 3	✓	✓	✓	
Arm 4	✓	✓		✓

8.2 Randomization

A randomization schedule was created to randomize subjects to 1 of 4 treatment arms in a 1:1:1:1 fashion using permutated blocks with randomly varying block sizes 4 and 8. Randomization is stratified by study site and by baseline Kellgren-Lawrence (KL) grade (0-2 vs. 3). The randomization schedule is uploaded to the TeMPO REDCap data management system and research coordinators will obtain arm assignments through this system.

8.3 Blinding

This trial is a single-blind study. The Analysis Core, including the Statistician and Principal Investigator, will be blinded to the subject's treatment assignment to avoid bias. Research coordinators communicating with the TeMPO subjects about outcomes assessment and performing follow-up assessments (i.e., performance tests, strength testing) are blinded to the subject's arm assignment. All other individuals, including the study subject, are not blinded. All subjects will be told during the informed consent procedure that this study has a placebo element. However, we will not specify what the placebo element is. Subjects will also be informed that all groups will receive at least some active intervention. The analysis will be done with group status masked; identification of each arm will not be revealed until the analyses are completed.

8.4 Study Assessments

TeMPO Schedule of Events

	Study Timepoint															
Study Form	*Pre-Screening*	*Screening*	*Enrollment Visit*	*Imaging*	Randomization	2 weeks	4 weeks	6 weeks	8 weeks	10 weeks	12 weeks	*At weeks (follow-*	*4 month*	*6 month*	*12 month*	As Needed
Pre-screening	X															
Screening		X*														
Physician Diagnosis		X*														
Baseline Visit Tracking Form		X	X													
MSK Exam			X*									X				
Patient Questionnaire		X*	X*									X*		X*	X*	
Image Assessment				X												
Randomization Form					X											
Post-Randomization Tracking					X	X										
Health Pamphlet Mailing					X*	X*	X*	X*	X*	X*	X*					
Home exercise log						X	X	X	X	X	X					
PT form																X
6 week call								X								
PT Summary Form													X			
Medical Record Review							X		X		X			X	X	
Reumuneration Form					X		X		X		X		X	X	X	
AE Form																X
Protocol Deviation Form																X
Injection Form																X
Discontinuation Form																X

Key:
Filled out by Research Assistant
Filled out by Physician
Filled out by Subject
Filled out by Radiologist
Filled out by Therapist
X* = potentially hard copy
X' = performed by blinded staff member
X = done by BWH only
Italic study timepoint = seen by blinded assessors
seen by therapists

8.5 Sample Size

We power the trial to detect a modest effect of 0.33 SD, reasoning that patients and clinicians will adopt treatments with modest effects if the intensity of treatment and risk of harm is low. As the KOOS Pain scale had a baseline SD of 16 points in MeTeOR, an effect size of 0.33 SD equates to 5.3 points, less than the estimated minimally clinically important difference for the KOOS Pain scale of 8-10 points[9,13]. We envision that the difference between Arm 1 (least intensive) and Arm 4 (most intensive) will be 10-12 points and that the comparisons of Arms 1 vs. 2 (effect of adherence optimization), 2 vs. 3 (additional provider interaction), and 3 vs. 4 (additional therapist-directed exercise and manual therapy) will represent components of this overall difference. Thus, our plan to power for a difference of 5.3 points in pairwise comparisons should provide adequate power for all contrasts. We assume 80% power and Type I error of 0.0167 to reflect three comparisons of strategies that would be used in practice (Arms 4 vs. 1; Arms 4 vs. 2; Arms 2 vs. 1). Given these considerations and balanced allocation across arms, each Arm would need 194 subjects. Allowing for 10% dropout at one year we will require 194 x 1.1 = 214 subjects per Arm, 856 in total.

9 Analytic Considerations

9.1 Hypotheses

This study will address the following hypotheses, summarized in Table 2:

Table 2. Study Hypotheses

Arm	Arm 1 (home exercise)	Arm 2 (home exercise w/adherence optimization)	Arm 3 (placebo PT)	Arm 4 (therapist-delivered PT)
Arm 1				
Arm 2	H3			
Arm 3		H4		
Arm 4	H1	H2	H5	

Primary Hypotheses:

Hypothesis 1: Arm 4 (protocolized home program, adherence optimization, in-clinic physical therapist-delivered exercise and manual therapy) is more efficacious than Arm 1 (protocolized home program alone). This analysis contrasts the putatively most potent intervention against the least potent.

Hypothesis 2: Arm 4 (protocolized home program, adherence optimization, in-clinic physical therapist-delivered exercise and manual therapy) is more efficacious than Arm 2 (protocolized home program with adherence optimization). This contrast isolates the effect of the in-clinic physical therapist-delivered intervention of exercise and manual therapy.

Hypothesis 3: Arm 2 (protocolized home program with adherence optimization) is more efficacious than Arm 1 (protocolized home program without adherence optimization). This contrast isolates the effect of the adherence optimization program.

Secondary Hypotheses:

Hypothesis 4: Arm 3 (protocolized home program, adherence optimization, in-clinic placebo PT) is more efficacious than Arm 2 (protocolized home program with adherence optimization). This contrast isolates the effect of attending an in-clinic program that provides face to face contact and support by clinicians with essentially no physiological effect directed at musculoskeletal impairments.

Hypothesis 5: Arm 4 (protocolized home program, adherence optimization, in-clinic physical therapist-delivered exercise and manual therapy) is more efficacious than Arm 3 (protocolized home program with adherence optimization and placebo PT). This contrast compares the effects of "active" PT and placebo PT, both given along with protocolized home exercises.

9.2 *Primary outcome*

The primary outcome in this trial is change in the KOOS Pain score from baseline to 3 months.

9.3 *Secondary outcomes:*

- Baseline to 3 month change in KOOS ADL
- Baseline to 3 month change in EQ-5D
- Treatment failure at 3 months
    - Defined as failing to achieve an 8-point improvement in KOOS pain from baseline and/or receiving injection or undergoing knee surgery prior to 3 months
- Baseline to 3 month change in 40 meter face-paced walk
- Baseline to 3 month change in number of repetitions in sit-to-stand test
- Baseline to 3 month change in single leg balance
- Baseline to 3 month change in muscle strength
    - Quadriceps
    - Hamstrings
    - Gluteus medius
- Durability of pain relief
    - Treatment Durability: Defined only for those participants achieving ≥ 8 point improvement in KOOS pain from baseline to 3 months, and who did not have an injection or knee surgery during this interval. Treatment durability will be defined as maintaining at least an 8-point improvement at the 12-month follow-up. Participants who receive injection or had knee surgery between 3 and 12-months will be classified as failing to achieve treatment durability.
- Longitudinal Outcomes:
        - KOOS pain at 3, 6, and 12 months
        - KOOS ADL at 3, 6, and 12 months

9.4 *Analysis Populations*

The primary analysis will be intention-to-treat, where each subject will be analyzed according to the randomization assignment, regardless of treatment received.

A secondary adherers analysis will exclude participants randomized to arms 3 or 4 who do not complete at least 8 in-person PT sessions.

A sensitivity analysis will exclude participants randomized between January and March 2020 to account for the lack of in-person PT during the first months of the COVID-19 pandemic.

A sensitivity analysis will exclude participants with protocol violations that affect eligibility or data integrity, including participants that were randomized but later found to be ineligible, participants receiving the incorrect PT protocol, and participants with incorrect index knee specified on the physician diagnosis form (see section 10.5).

10 Summary of Study Data

Study data will be summarized overall and by treatment group. Continuous data will be summarized by mean (SD) or median (IQR) range, as appropriate. Categorical data will be summarized by frequency and percent.

10.1 Participant Disposition

Participant disposition will be summarized in a CONSORT diagram.[14] This will include the number of patients screened for eligibility, the number of patients eligible, the number of patients interested, the number of patients randomized, and the number of patients randomized to each treatment group. The reasons for ineligibility will also be summarized as part of the CONSORT diagram.

Visit completion status and the reason for early discontinuation of the study will be summarized by site and treatment arm with descriptive statistics. These include the number and percentage of each reason for discontinuation. We will also summarize the number and percent of participants undergoing APM by timepoint and undergoing injection by timepoint.

10.2 Demographic and Baseline Variables

Baseline characteristics will be summarized overall and by treatment group and presented in a table. These include:

- Site
- Sex
- Race
- Ethnicity
- Educational attainment
- K-L Grade
- Age
- Body Mass Index
- Enrollment date (prior to or after March 15, 2020)

We will also present the baseline value of the primary and secondary outcomes.

10.3 Adherence

Adherence will be described in arms 3 and 4 by the mean, median, and range of the number of in-person PT visits attended.
All subjects will be asked to complete a brief home exercise adherence log once every two weeks from randomization through 3 months We will also describe self-reported adherence to exercise for all arms.

10.4 Data collected outside of visit window

The windows for data collection are shown in the table below:

Form	*Timeframe (based off intervention start unless noted)*
Baseline Questionnaire and Assessment	Any time before randomization
3-Month Questionnaire and Assessment	12-16 weeks after randomization
6-Month Questionnaire	22-26 weeks after randomization
12-Month Questionnaire	46-50 weeks after randomization

We anticipate that some study assessments may be completed outside of the visit window (for example, a subject may have to postpone a scheduled visit for a week or two because of a death in the family). We will summarize the number and percentage of assessments completed outside of the ideal window separately for questionnaire completion and performance assessments. If necessary, sensitivity analyses will be undertaken to examine if the conclusions about the study outcomes are affected by the exclusion of out-of-window visits.

10.5 *Protocol violations*

Violations will be summarized overall and for each arm of the study. The number of participants with violations and the frequencies and percentages of the types of protocol violations will be presented. The pre-specified list of protocol violations is presented below:

Protocol violation: *Noncompliance with or divergence from IRB-approved protocol that does negatively affect 1) the patient's safety, welfare or eligibility, or 2) the integrity of the data. In particular, this includes any instance in which exclusion criteria, randomization criteria, discontinuation criteria, or confidentiality procedures are applied in a way that diverges from study protocol.* Examples include:

- Incorrectly enrolling an ineligible patient
- Missing MRI or x-ray
- Sending a subject materials for an arm to which the subject was not randomized
- Breaches of HIPAA
- Inadequate/missing informed consent
- Unreported SAEs
- Unblinding a party that should be blinded
- Performing the incorrect PT protocol with a subject (e.g. an arm 4 subject receives ultrasound)
- First PT visit (arms 3 and 4) occurring more than 30 days after the scheduled intervention start or not at all.

11 Analytic Approach

11.1 Primary Analysis

In the primary analysis we will test the 3 hypotheses outlined in section 7.1. As placebo is not used in practice, we do not consider tests involving Arm 3 as one of our multiple comparisons. For the primary analysis the alpha level is adjusted to 0.0167 to account for 3 comparisons. We will adjust for randomization arm, the randomization stratification factors of study site and baseline KL grade (0-2 vs. 3) as well as baseline KOOS pain and enrollment date (prior to or after March 15, 2020).

We will use linear regression as our main analytic approach, as our primary outcome of change in KOOS pain is continuous. The primary analysis will be complete case. Data will be summarized by treatment group with the number, mean, and standard deviation. 98.3% Cis will be used to summarize the three between-group comparisons of interest.

11.2 Secondary Analyses

11.2.1 Secondary Hypotheses

In a similar fashion to the primary analysis, we will test the two secondary hypotheses outlined in section 7.2 for the outcome of 3 month KOOS pain.

11.2.2 Secondary Outcomes

All models will adjust for randomization arm, the randomization stratification factors of study side and baseline KL grade (0-2 vs. 3) as well as baseline value of the outcome and enrollment date (prior to or after March 15, 2020). We will adopt an intention-to-treat approach.

We analyze 3 month KOOS pain with an Analysis of Covariance adjusted for baseline KOOS pain. (This approach uses the 3 month value, adjusted for baseline, as the outcome, rather than the difference between three months and baseline).

11.2.2.1 Continuous secondary outcomes for baseline to 3 month change (KOOS ADL, EQ-5D, performance measures)

We will analyze continuous secondary outcomes with linear regression additionally adjusting for the baseline value of the outcome. In the case that some measures deviate from a normal distribution (e.g., number of repetitions in sit-to-stand test) we will consider data transformation or a generalized linear model (e.g., Poisson regression).

11.2.2.2 Treatment Durability

We will assess treatment durability descriptively as the number and percent of participants in each arm meeting the criteria specified in 9.3. We will report alongside these results the number and percent of participants eligible for the durability analyses (meaning that they achieved improvement of ≥8 points on KOOS Pain between baseline and 3 months, in the absence of surgical intervention or injection).

11.2.2.3 Longitudinal Outcomes

Longitudinal KOOS pain and ADL will be assessed with linear mixed-effects models including the baseline, 3 month, 6 month, and 12 month data[15]. These data will also be presented graphically by treatment group with means and 95% confidence intervals.

11.3 Interim Analysis

We have no planned interim analyses in this trail.

11.4 Subgroup analysis

We have no planned subgroup analyses in this trial.

11.5 Statistical analyses to handle missing data

We anticipate ~5% dropout at 3 months and ~10% at 12 months based on dropout rates from the Meniscal Tear in Osteoarthritis Research (MeTeOR) RCT[9]. We will address dropout and other forms of missing data first by reporting the number of subjects dropping out in each arm and the reasons for dropout. We will determine whether the frequency of these dropouts differs across study arms. We will examine the distribution of baseline covariates and baseline KOOS pain in subjects dropping out compared to those completing the study. Our primary assumptions

are that data will be missing at random (MAR), and that dropout will not depend on unobserved outcomes.

We will perform sensitivity analysis for missing data for our primary outcome of KOOS pain and our secondary outcomes of KOOS ADL, EQ-5D, strength measures, objective function measures, and treatment failure for our three primary hypotheses. We will use linear mixed effects models to analyze the continuous outcomes of KOOS pain, KOOS ADL, and EQ-5D. Models will be adjusted for timepoint, randomization arm, the randomization stratification factors of study site and baseline KL grade (0-2 vs. 3) as well as the baseline value of the outcome and enrollment date (prior to or after March 15, 2020). A timepoint by treatment arm interaction term will test whether change from baseline to 3 months differs between treatment arms. Linear mixed-effects models can handle imbalanced data and are appropriate for data that are missing at random.

For the categorical outcome of three month treatment failure we will use multiple imputation for missing data. To more accurately reflect the uncertainty in missing values, MI entails first creating multiple versions of the dataset, with missing values imputed under an assumed distribution, and then averaging results across these datasets. We will impute data using logistic regression by fully conditional specification, as recommended for dichotomous outcomes[16,17]. In the primary analysis, we will impute data under the missing at random assumption, assuming that missing data were associated with observed data (baseline KL grade, age, sex, BMI, baseline KOOS pain, baseline KOOS ADL, study site)[18,19]. Imputations will be done stratified by treatment arm. We will generate 20 imputed datasets and used the MIANALYZE procedure in SAS 9.4 to combine the results across the imputations. If the proportion of missing data for continuous outcomes exceeds 10%, an additional sensitivity analysis using multiple imputation will be conducted as outlined above.

11.6 Multiple testing

In the primary analysis we will test the 3 hypotheses outlined in section 7.1. As placebo is not used in practice, we do not consider tests involving Arm 3 as one of our multiple comparisons. For the primary analysis the alpha level is adjusted to 0.0167 to account for 3 comparisons. Secondary hypotheses and outcomes will not be adjusted for multiple testing.

12 Analysis of Safety Endpoints

12.1 Adverse events

The number and percentages of adverse events (Aes) will be summarized in total and for each treatment arm. The severity and relatedness of each adverse event to the study will be broken down by intervention. Adverse events include:

- any risk specific to physical therapy or to exercise including an episode or exacerbation of musculoskeletal pain (knee, hip, back, neck). Musculoskeletal discomfort is common after exercise and to be expected. Thus, we have defined a musculoskeletal adverse event as requiring a level of severity that precluded walking for one day or more or that that required the subject to rely upon an assistive device such as a cane or crutch for at least one day to ambulate.
- Any risk related to topical treatments (ultrasound, gel) such rash or pruritis (itching).

12.2 Serious adverse events

As with AEs, the number and percentages of serious adverse events (SAEs) will be summarized in total and for each treatment arm. The relatedness and type of each SAE will be broken down by treatment arm. Serious adverse events include:

- Death from any cause
- Overnight hospital admission as a result of medical problems such as myocardial infarction, stoke, pulmonary embolus, congestive heart failure decompensation, pneumonia, serious infection
- Overnight hospital admission as a result of surgical problems such as emergent surgery (e.g. cholecystectomy for biliary sepsis, acute appendicitis or diverticulitis)
- Arthroscopic partial meniscectomy (APM)
- Total knee replacement (TKR)

12.3 Analysis for AEs and SAEs

We anticipate few AEs and negligible SAEs. We will present these endpoints descriptively. We will aggregate AEs and compare the number of AEs across the four arms using a Kruskall-Wallis Test or Fisher's exact test if the numbers are small.

13 Derived Variable Definitions, Coding, and QC Plan

13.1 Derived Variable Definitions and Coding

KOOS Pain[20,21]: The KOOS is a knee-specific self-administrated instrument, to assess the patients' opinion about their knee and associated problems. The KOOS evaluates five separately scored outcomes: Pain, other Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec), and knee-related Quality of Life (QOL).

KOOS Pain has 9 items and asks the amount of knee pain experienced in the last week. A Likert scale is used, and all items have five possible answer options scored from 0 (No problems) to 4 (Extreme problems) and each of the five scores is calculated as the sum of the items included. We will normalize raw KOOS Pain score to 100 (with 100 being the worst), by applying the mean of the observed items within the KOOS Pain, dividing by 4, and multiplying by 100. Formular is provided below:

*KOOSPain= Mean (KoosP_1 -- KoosP_9)/4 * 100*

As long as at least 50% of the subscale items are answered for each subscale, a mean score can be calculated. If more than 50% of the subscale items are omitted, the response is considered invalid, and no subscale score should be calculated. For KOOS Pain, a valid score requires at least 5 non-missing items.

KOOS ADL[20,21]: KOOS Function in daily living (ADL) has 10 items and ask about physical function. A Likert scale is used, and all items have five possible answer options scored from 0 (No problems) to 4 (Extreme problems) and each of the five scores is calculated as the sum of the items included. We will normalize raw KOOS ADL score to 100 (with 100 being the worst), by applying the mean of the observed items within the KOOS ADL, dividing by 4, and multiplying by 100. Formular is provided below:

*KOOSADL= Mean (KoosDailyLV_1 -- KoosDailyLV_10)/4 * 100*

For KOOS ADL, a valid score requires at least 6 non-missing items.

EQ-5D[22]: EQ-5D measures health using five levels of severity in five dimensions, including mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Each dimension now has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The respondent is asked to indicate his/her health state by choosing the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number expressing the level selected for that dimension. The digits for 5 dimensions can be combined in a 5-digit number describing the respondent's health state. The numbers 1-5 have no arithmetic properties. The coding for EQ-5D-5L is publicly available on EuroQoL website. The final EQ5D score ranges from 0 to 1 with 1 being the best health. A valid EQ-5D-5L score requires non-missing for all five dimensions.

Treatment Durability: Treatment durability assessed whether the treatment effect lasts beyond 3 months. This is defined only for those participants achieving ≥ 8 point improvement in KOOS pain from baseline to 3 months. Treatment durability is not defined for participants failing to achieve at least an 8 point improvement in KOOS pain or who undergo APM or injection prior to the 3 month visit. Treatment durability will be defined as maintaining at least an 8-point improvement from baseline at the 12-month follow-up, without undergoing APM or injection. Participants failing to maintain an 8-point improvement from baseline or undergoing APM or injection between 3 months and 12 months will be coded as not having treatment durability.

13.2 QC Plan

A second review statistician will independently reproduce the primary analyses and summary statistics in table 2. All analytic output will include the date and time, source dataset, and author.

13.3 Code for Primary and Secondary Analyses

Analyses will be conducted in SAS v 9.4 (SAS Institute, Cary NC).

```
PROC GENMOD DATA=data;

CLASS trt site klg enroll;

MODEL koosp_chg = trt site KLG enroll koosp_bl;

CONTRAST  'Group 1 vs Group 4' trt -1 0 0 1;

CONTRAST  'Group 2 vs Group 4' trt 0 -1 0 1;

CONTRAST  'Group 1 vs Group 2' trt -1 1 0 0 ;

LSMEANS trt / PDIFF DIFF CL ALPHA=0.0167 ;

run;
```

14 Shell Tables

Table 1: Participant Features at Baseline

	Total cohort	Arm 1 Home Exercise	Arm 2 Home Ex + Adherence	Arm 3 Home Ex + Adherence + PT (placebo)	Arm 4 Home Ex+ Adherence + PT (strengthening)
Number of subjects					
Site [N, (%)] Brigham and Women's University at Buffalo University of Pittsburgh Cleveland Clinic					
Sex [N, (%)] Male Female					
Race White Black, African American Asian Native American, Pacific Islander Multiple races endorsed Not answered					
Ethnicity Hispanic Not Hispanic					
Educational attainment High school or less Associate degree or some college College degree and higher					
K-L Grade[N, (%)] 0 1 2 3					
Age (mean (SD))					
Body Mass Index (mean (SD)					
Enrollment date [N, (%)] Prior to March 15, 2020 After March 15, 2020					

Table 2: Primary and secondary outcomes

	Arm 1 Home Exercise	Arm 2 Home Ex + Adherence	Arm 3 Home Ex + Adherence + PT (placebo)	Arm 4 Home Ex+ Adherence + PT (strengthening)	Arm 4 vs. Arm 1	Arm 4 vs. Arm 2	Arm 2 vs Arm 1
KOOS Pain Baseline 3 months 3 mo – BL	Mean (SD) Mean (SD) Mean (SD)	Mean (SD) Mean (SD) Mean (SD)	Mean (SD) Mean (SD) Mean (SD)	Mean (SD) Mean (SD) Mean (SD)	Mean (98.3% CI)	Mean (98.3% CI)	Mean (98.3% CI)
KOOS ADL Baseline 3 months 3 mo - BL							
EQ-5D Baseline 3 months 3 mo - BL							
Quadriceps Strength Baseline 3 months 3 mo – BL							
Hamstring Strength Baseline 3 months 3 mo – BL							
Gluteus Medius Strength Baseline 3 months 3 mo – BL							
40 meter timed walk Baseline 3 months 3 mo – BL							
30 second sit to stand Baseline 3 months 3 mo – BL							
Single leg stand Baseline 3 months 3 mo – BL							

Table 3: Adverse Events

	Total cohort	Arm 1 Home Exercise	Arm 2 Home Ex + Adherence	Arm 3 Home Ex + Adherence + PT (placebo)	Arm 4 Home Ex+ Adherence + PT (strengthening)
Serious adverse events Deaths Hospitalizations (overnight admission) APM TKR Total					
Adverse events APM (index knee) Other MSK Cardiovascular Neurological Pulmonary Infectious Other Total					

Figure 1: CONSORT Diagarm

Figure 2: A figure showing the longitudinal results of KOOS Pain, KOOS ADL at baseline, 3, 6, 12 months (mean value for each group with 95% confidence intervals.

Additional tables to be presented in appendix:

- Analysis of secondary hypotheses for primary and secondary outcomes (arm 3 vs. arm 2, arm 3 vs arm 4).
- Descriptive statistics (n, mean, standard deviation, 95% confidence interval) of longitudinal outcomes (baseline, 3m, 6m, 12m) KOOS pain, KOOS ADL.
- Treatment durability overall and by treatment group
- Sensitivity analysis for missing data: Primary and secondary outcomes

15 Bibliography

1. Sullivan JK, Irrgang JJ, Losina E, et al. The TeMPO trial (treatment of meniscal tears in osteoarthritis): rationale and design features for a four arm randomized controlled clinical trial. BMC Musculoskelet Disord 2018;19:429.
2. Deshpande BR, Katz JN, Solomon DH, et al. Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity. Arthritis Care & Research 2016;68:1743-50.
3. Englund M, Guermazi A, Gale D, et al. Incidental Meniscal Findings on Knee MRI in Middle-Aged and Elderly Persons. New England Journal of Medicine 2008;359:1108-15.
4. Bhattacharyya T, Gale D, Dewire P, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. The Journal of bone and joint surgery American volume 2003;85:4-9.
5. Niu NN, Losina E, Martin SD, Wright J, Solomon DH, Katz JN. Development and preliminary validation of a meniscal symptom index. Arthritis Care & Research 2011;63:208-15.
6. Logerstedt DS, Snyder-Mackler L, Ritter RC, Axe MJ, Godges J. Knee Pain and Mobility Impairments: Meniscal and Articular Cartilage Lesions. Journal of Orthopaedic & Sports Physical Therapy 2010;40:A1-597.
7. Yim JH, Seon JK, Song EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. The American journal of sports medicine 2013;41:1565-70.
8. Gauffin H, Tagesson S, Meunier A, Magnusson H, Kvist J. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis and Cartilage 2014;22:1808-16.
9. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus Physical Therapy for a Meniscal Tear and Osteoarthritis. New England Journal of Medicine 2013;368:1675-84.
10. Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA 2013;21:358-64.
11. Buchbinder R. Meniscectomy in Patients with Knee Osteoarthritis and a Meniscal Tear? New England Journal of Medicine 2013;368:1740-1.
12. Stensrud S, Roos EM, Risberg MA. A 12-Week Exercise Therapy Program in Middle-Aged Patients With Degenerative Meniscus Tears: A Case Series With 1-Year Follow-up. Journal of Orthopaedic & Sports Physical Therapy 2012;42:919-31.
13. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health and Quality of Life Outcomes 2003;1:64.
14. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869-c.
15. Coffman CJ, Edelman D, Woolson RF. To condition or not condition? Analysing 'change' in longitudinal randomised controlled trials. BMJ Open 2016;6:e013096.
16. Liu Y, De A. Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. International Journal of Statistics in Medical Research 2015;4:287-95.
17. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. American Journal of Epidemiology 2010;171:624-32.
18. Rubin DB. Multiple Imputation for Nonresponse in Surveys. United States: John Wiley & Sons, Incorporated; 1987.
19. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Statistical Methods in Medical Research 2014;23:440-59.

20. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28:88-96.
21. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64.
22. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727-36.