WEARABLE SENSOR-BASED OUTCOMES FOLLOWING PHYSICAL THERAPY IN KNEE OSTEOARTHRITIS: A FEASIBILITY STUDY (WESENS-OA)

Statistical Analysis Plan
(SAP)

Version: 1
Date: 4-01-2021
Authors: Deepak Kumar, Charmaine Demanuele

NCT# NCT04243096

TABLE OF CONTENTS

LIST OF TABLES

APPENDICES

1. INTRODUCTION

Text taken directly from the protocol is *italicized*.

The main focus of management of knee osteoarthritis (OA) symptoms is to reduce pain. However, there is a concern that the pain improvement may lead to increased use of the previously underutilized joint in ways that may be detrimental to joint health. Exercise-based physical therapy (PT) interventions can reduce pain and improve function in people with knee osteoarthritis, and may mitigate abnormal joint mechanics that can be detrimental with greater use of the joint. Thus PT interventions may limit the risk of joint injury caused by increased activity in a previously underutilized joint. The gold standard assessment of joint mechanics is in a gait lab, but this would be infeasible on a large-scale. Thus, we aim to assess how improvements in pain and function with a PT program are associated with metrics obtained from laboratory assessments and wearable sensors in the real world. Additionally, to plan for scenarios where in-lab assessments may not be feasible in current or future studies, a substudy will be undertaken to assess reproducibility of sensor-based measures during physical performance testing across at-home and in-lab implementation, as well as, reproducibility of these measures over repeated at-home implementation.

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in this study. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

The current protocol has contingencies for dealing with any restrictions on data capture that may arise due to ongoing COVID-19 pandemic. If these contingencies are deployed, the SAP will be modified to account for differences in mode of data capture across participants and/or visits.

1.1. Study Objectives

The overall study objectives are to measure the effect of PT on functional performance and pain reduction using both patient reported outcomes questionnaires (PROs) and digital metrics obtained from the laboratory assessments and wearable sensors worn in the real world, and understand factors that may influence response to PT treatment.

More specifically, the study aims are as follows:

Primary Aim 1: Measure the effect of PT on function using digital measures from wearable sensors

Secondary Aim 1: Compare in-lab and real world (at-home) sensor metrics with ePRO outcomes
Secondary Aim 2: Assess the effect of PT on sleep using ePRO (DSIS questionnaire) and wearable sensor metrics
Secondary Aim 3: Assess the persistence of the PT effect from week 12 to week 18 using in-lab and real world (at-home) sensor metrics, and ePROs

Exploratory Aim 1: Assess whether sensor-based baseline measures can stratify treatment response
Exploratory Aim 2: Assess whether participant phenotype; pain, physical performance, affective traits (catastrophizing), can stratify treatment response

Substudy Aim 1: Examine the test-retest reliability of at-home instrumented physical performance tests
Substudy Aim 2: Examine agreement between wearable sensor data collected during in-lab and at-home instrumented physical performance tests

Study Design

This will be a longitudinal, single arm, 19-week study to investigate the utility of digital assessments to measure the efficacy of physical therapy (PT) for reducing pain and improving function in people with knee osteoarthritis (OA). Sixty participants will receive a supervised PT program for 12 weeks with in-person assessment of strength, balance, gait, and joint movement, as well as at-home assessment using wearable sensors, followed by 6 weeks of monitoring while they continue to exercise at home. Participants will also complete questionnaires and have quantitative sensory testing during this study to evaluate pain phenotypes. Additionally, to plan for scenarios where in-lab assessments may not be feasible in current or future studies, a substudy will be undertaken to assess reproducibility of sensor-based measures during physical performance testing across at-home and in-lab implementation, as well as, reproducibility of these measures over repeated at-home implementation.

2. ABBREVIATIONS USED

ADL = Activities of Daily Living
COVID-19 = Coronavirus Disease 2019
CPM = Conditioned Pain Modulation
DSIS = Daily Sleep Interference Scale
EIH = Exercise-induced Hypoalgesia
ePRO = electronic Patient Reported Outcome
ICC = intra-class correlation coefficient
IMU = Inertial measurement unit
KOOS = Knee injury and Osteoarthritis Outcome Scale
MCID = Minimally Clinically Important Difference
MMRM = Mixed model repeated measures
MVPA = Time spent in moderate and vigorous physical activity
NRSna= Numeric Rating Scale for pain during nominated activity
OA = Osteoarthritis
PGA-OA = Patient Global Assessment of Osteoarthritis
PPT = Pressure Pain Threshold
PT = Physical Therapy
QOL = Quality of Life
SAP = Statistical Analysis Plan
SPARC = Spectral Arc Length
SPPB = Short Performance Physical Battery
TS = Temporal Summation
WASO = Wake after sleep onset
WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index

3. OUTCOMES AND ANALYSES

Terminology

Digital metrics = data from wearable sensors including lumbar and wrist-worn sensors

Physical Function = overall term that includes assessments of physical performance, functional biomechanical tests, and at home physical activities.

Physical performance tests = standardized tests including 5 times sit-to-stand test, 15 second step-up test, stair climbing test, 6 minute walk test, and SPPB

Functional biomechanical tests = activities assessed using optical motion capture including walking, fast walking, stair ascent and descent, sit to stand, and balance

3.1. Primary Endpoint(s)

Primary Aim 1: Measure the effect of PT on physical function using digital measures from wearable sensors
Time period: Change from baseline to week 12

Primary Endpoints

A. In-lab Data Gait Measures

Table 1. Gait measures obtained from Gaitpy algorithm applied on the lumbar sensor and from Optical Motion Capture System

	GaitPy applied on Lumbar Sensor[1]	**Optical Motion Capture System[2-4]**
a.	Cadence (steps/min)	Cadence (steps/min)
b.	Step length symmetry	Step length symmetry
c.	Gait speed (m/s)	Gait speed (m/s)
d.	Stride duration (s)	Stride duration (s)
e.	Stride duration asymmetry (s)	Stride duration asymmetry (s)
f.	Step duration (s)	Step duration (s)
g.	Step duration asymmetry (s)	Step duration asymmetry (s)
h.	Initial double support time (s)	Initial double support time (s)
i.	Terminal double support time (s)	Terminal double support time(s)
j.	Total double support time (s)	Total double support time (s)
k.	Single limb support time (s)	Single limb support time (s)

l.	Stance time (s)	Stance time (s)
m.	Swing time (s)	Swing time (s)
n.	Step length (m)	Step length (m)
o.		Peak knee index (%BodyWeight x Height
p.		Knee flexion excursion (deg)
q.		Knee extension excursion (deg)
r.		Peak medial quadriceps-medial hamstrings co-contraction during loading response during walking at self-selected pace
s.		Peak lateral quadriceps-lateral hamstrings co-contraction during loading response during walking at self-selected pace
t.		Peak sagittal trunk angle (deg)
u.		Peak knee adduction moment (%BodyWeight x Height)
v.		Peak knee flexion moment (%BodyWeight x Height)
w.		Knee adduction moment impulse (%BodyWeight x Height x Sec)
x.		Knee flexion moment impulse (%BodyWeight x Height x Sec)
y.		Peak sagittal total support moment (%BodyWeight x Height)

B. In-lab Data Sit-to-stand Measures

Table 2. Sit-to-stand measures derived from Sit2StandPy algorithm applied on the lumbar sensor and from Optical Motion Capture System

	Sit2StandPy applied on Lumbar Sensor[5]	**Optical Motion Capture System[6, 7]**
a.	Number of sit-to-stands	Number of sit-to-stands
b.	Duration (s)	Duration (s)
c.	Maximum acceleration (m/s^2)	Maximum acceleration (m/s^2)
d.	Minimum acceleration (m/s^2)	Minimum acceleration (m/s^2)
e.	SPARC: SPectral ARC length	
f.		Peak total sagittal lower extremity support moment (%Bodyweight*Height)

g.		Peak sagittal trunk angle (deg)
h.		Peak medial quadriceps-medial hamstrings co-contraction
i.		Peak knee adduction moment (%BodyWeight x Height)
j.		Peak knee flexion moment (%BodyWeight x Height)
k.		Peak knee flexion (deg)

C. In-lab data obtained from the Optical Motion Capture System during standing balance tasks[8]

- Standing balance
  - Center of pressure displacement in A-P direction (mm)
  - Center of pressure displacement in M-L direction (mm)
  - Mean center of pressure velocity (mm/s)

D. In-lab Data: Physical Performance Outcomes

- 5 times sit to stand
  - Time taken to complete 5 chair stands
- 15 second step up test
  - Number of step ups with each foot in 15 seconds
- Stair climbing test
  - Time taken to complete (sec)
- 6 minute walk test
  - Distance covered (m)
- Short Performance Physical Battery
  - Total score
  - Gait speed score
  - Repeated chair stand score
  - Balance tests score

E. At-home Data: Gait, sit-to-stand and physical activity measures

- Gait measures derived from GaitPy algorithm applied on the lumbar sensor[1]
  - # of gait bouts
  - Cadence (steps/min)
  - Step length symmetry
  - Gait speed (m/s)
  - Stride duration (s)
  - Stride duration asymmetry (s)
  - Step duration (s)
  - Step duration asymmetry (s)
  - Initial double support time (s)
  - Terminal double support time(s)

  - Total double support time (s)
  - Single limb support time (s)
  - Stance time (s)
  - Swing time (s)
  - Step length (m)

- Sit-to-stand measures derived from Sit2StandPy algorithm applied on the lumbar sensor[5]
  - Number of sit-to-stands
  - Duration (s)
  - Maximum acceleration (m/s^2)
  - Minimum acceleration (m/s^2)
  - SPARC : SPectral ARC length

- Physical Activity measures based on Actigraph's Algotihms[9-12]
  - Steps/day
  - Time spent in moderate and vigorous physical activity (MVPA) (min)
  - % wear-time in MVPA
  - Sedentary time (min)
  - % wear-time sedentary

3.2. Secondary Endpoints

Secondary Aim 1: Compare in-lab and real world (at-home) sensor metrics with ePRO outcomes
Time period: Change from baseline to week 12.

Wearable sensor metrics from 3.1 A (Table 1 Column 1); 3.1 B (Table 2 Column 1) and 3.1 E. As described in the analysis Section 4 below, only a subset of these endpoints that show excellent agreement with the gold standard measures during in-lab assessments (ICC >= 0.75) will be investigated for this aim.

Electronic patient reported outcome (ePRO) endpoints

- KOOS-derived WOMAC 3.0 Pain score
- KOOS-derived WOMAC 3.0 Function score
- Numeric Rating Scale for pain during nominated activity (NRSna)
- Patient Global Assessment of Osteoarthritis (PGA-OA)
- KOOS-derived WOMAC 3.0 stiffness
- KOOS Pain
- KOOS Symptoms
- KOOS ADL
- KOOS Sports/Rec
- KOOS QOL

Secondary Aim 2: Assess the effect of PT on sleep using ePRO (DSIS questionnaire) and wearable sensor metrics
Time period: change from baseline to week 12

Sleep measures

- Daily Sleep Interference Scale (DSIS)
- Sleep measures from Actigraph's algorithm[13, 14]
  - Total sleep time from wrist-worn sensor (min)
  - Sleep efficiency from wrist-worn sensor (%)
  - Wake after sleep onset (WASO) from wrist-worn sensor (min)

Secondary Aim 3: *Assess the persistence of the PT effect from week 12 to week 18 using in-lab and real world (at-home) sensor metrics, and ePROs*
Time period: change from baseline to week 12 and from baseline to week 18

Wearable sensor metrics from 3.1 A (Table 1 Column 1); 3.1 B (Table 2 Column 1) and 3.1 E. As described in the analysis Section 4 below, only a subset of these endpoints that show excellent agreement with the gold standard measures during in-lab assessments (ICC >= 0.75) will be investigated for this aim.

ePRO endpoints listed above.

3.3. Exploratory Endpoints

Exploratory Aim 1: Assess whether sensor-based baseline measures can stratify treatment response

Wearable sensor metrics from 3.1 A (Table 1 Column 1); 3.1 B (Table 2 Column 1) and 3.1 E. As described in the analysis Section 4 below, only a subset of these endpoints that show excellent agreement with the gold standard measures during in-lab assessments (ICC >= 0.75) will be investigated for this aim.

ePRO endpoints listed in Section 3.2.

Exploratory Aim 2: Assess whether participant phenotype; pain, physical performance, affective traits (catastrophizing), can stratify treatment response

Physical Performance Endpoints listed in Section 3.1D and ePRO endpoints in Section 3.2.

Pain phenotyping endpoints

- PPT (kgf)
- Temporal summation (yes/no)
- CPM (yes/no)
- EIH (yes/no)
- Number of painful joints from joint homunculus
- Pain catastrophizing (modified)
- Widespread pain (single question)
- PainDETECT (modified)

Strength assessment endpoints:

- Isometric testing

- Peak extensor torque (Nm/kg)
- Peak flexor torque (Nm/kg)

▪ Isokinetic testing
- Peak extensor torque at 60 deg/s(Nm/kg)
- Peak flexor torque at 60 deg/s (Nm/kg)
- Peak extensor torque at 120 deg/s (Nm/kg)
- Peak flexor torque at 120 deg/s (Nm/kg)

3.4. Substudy Endpoints

Substudy Aim 1: Examine the test-retest reliability of at-home instrumented physical performance tests
Substudy Aim 2: Examine agreement between wearable sensor data collected during in-lab and at-home instrumented physical performance tests

Table 3. Gait measures obtained from Gaitpy algorithm applied on the lumbar sensor and from 3-sensor IMU System during the 7-meter walk tasks performed at home and in-lab

	GaitPy applied on Lumbar Sensor[1]	**3-sensor IMU system[15]**
a.	Cadence (steps/min)	Cadence (steps/min)
b.	Step length symmetry	
c.	Gait speed (m/s)	Gait speed (m/s)
d.	Stride duration (s)	Gait cycle duration (s)
e.	Stride duration asymmetry (s)	
f.	Step duration (s)	Step duration (s)
g.	Step duration asymmetry (s)	
h.	Initial double support time (s)	
i.	Terminal double support time (s)	Terminal double support (% gait cycle, s)
j.	Total double support time (s)	Double support (% gait cycle, s)
k.	Single limb support time (s)	
l.	Stance time (s)	Stance phase (% gait cycle, s)
m.	Swing time (s)	Swing (% gait cycle, s)
n.	Step length (m)	
o.		Foot clearance (m)
p.		Lateral step variability
q.		Foot strike angle (deg)

r.		Toe off angle (deg)
s.		Toe out angle (deg)
t.		Turning angle (deg)
u.		Turning duration (s)
v.		Turning velocity (deg/s)

Table 4. Sit-to-stand measures derived from Sit2StandPy algorithm applied on the lumbar sensor and from the 3-sensor IMU system during the 5 times sit-to-stand tasks performed at home

	Sit2StandPy applied on Lumbar Sensor[5]	**3-sensor IMU system[16]**
a.	Number of sit-to-stands	
b.	Sit-to-stand duration (s)	Sit-to-stand duration (s)
c.	Maximum acceleration (m/s^2)	
d.	Minimum acceleration (m/s^2)	
e.	SPARC: SPectral ARC length	
f.		Stand-to-sit duration (s)

3.5. Safety Endpoints

An AE is defined as any untoward medical occurrence and can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease, whether or not related to the participant's participation in the study.
Any AE that occurs from the time the participant consents to the clinical research through the completion of the qualifying procedure will be recorded.

4. ANALYSES AND SUMMARIES

4.1. General Methods

Data will be summarized using descriptive statistics (number of subjects (n), mean, median, standard deviation, minimum and maximum) for continuous (or near continuous) variables, and using frequency and percentages for discrete variables. Shapiro-Wilk test will be used to assess normality of variables. Non-parametric statistical methods will be used to analyze non-normally distributed variables.

P-values will be generated where appropriate and any p-value < 0.05 will be considered statistically significant. If multiple p-values are generated within an analyses, Benjamini-Hochberg false discovery rate (FDR) correction will be used to determine significance. P-values will be rounded to 3

decimal places and therefore presented as 0.xxx; P-values smaller than 0.001 will be reported as '<0.001'.

4.1.1. Baseline and Handling Missing Data

All summaries and analyses will be based on observed data and missing data imputation is not planned.

4.2. Analysis of Primary Endpoints

Below is a description of the analyses performed to address the study objectives and to analyze the corresponding endpoints.

Measure the effect of PT on physical function using digital measures from wearable sensors

The following analyses will be performed to address this primary aim.

A. Validate wearable sensor-based metrics of functional biomechanical tests in people with knee OA against optical motion capture

Wearable sensor metrics (Gait: Table 1 rows *a. – n.*; Sit-to-stand: Table 2 rows *a. – d.*) will be compared against the same measures obtained from the optical motion capture system (used as the gold-standard) using intra-class correlation coefficient (ICC) and its 95% lower and upper confidence limits. The ICC reflects both the degree of correlation and agreement between the measurements. We will interpret agreement between measurements according to the following benchmarks: ICC ≤ 0.4 indicates 'poor', 0.4 to 0.59 'moderate', 0.6 to 0.74 'good', and 0.75 to 1 'excellent' reliability[17]. In addition, we will also compute the mean, variance and percentage error between the standard and the digital measurements. Bland–Altman plots and 95% limits of agreement (average difference ± 1.96 standard deviation of the difference) will also be computed.

B. Compare wearable sensor-based metrics of physical performance in people with knee OA with standardized physical performance outcomes.

Wearable sensors metrics showing excellent reliability from analysis A will be examined (i.e. a subset of the metrics shown in Table 1 rows *a. – n.* for Gait; and Table 2 rows *a. – d.* for Sit-to-stand). Sensor-based metrics will be derived during physical performance tests conducted in-lab and will be correlated with the corresponding standardized physical performance outcomes using Pearson's product-moment correlations or Spearman's rank order correlations depending on their normality.

Specifically, the comparisons that will be computed are shown in Table 5.

Table 5. Comparisons of sensor-based metrics of physical performance with standardized physical performance outcomes

Task	Standardized Physical Performance Outcomes	Wearable Sensor Metrics
5 times sit-to-stand	Time taken to complete 5 chair stands	Subset of reliable metrics from Table 2 rows *a. – d.*

6 minute walk test	Distance covered (m)	Subset of reliable metrics from Table 1 rows *a.* – *n*.
SPPB	Gait speed score, Repeated chair test score	Subset of reliable metrics from Table 1 rows *a.* – *n*. for Gait; and Table 2 rows *a.* – *d*. for Sit-to-stand).

C. Assess the effects of exercise-based PT on outcomes from wearable sensors during in-lab physical performance tests

Wearable sensors with excellent reliability from analysis A will be examined (i.e. a subset of the metrics shown in Table 1 rows *a.* – *n*. for Gait; and Table 2 rows *a.* – *d*. for Sit-to-stand). Mixed model repeated measures (MMRM) analysis will be used to investigate the effect of PT on endpoints derived from the lumbar sensor during in-lab physical performance tests (5 times sit-to-stand, 6 minute walk test, SPPB). The outcome measures will be expressed as the change from baseline at each visit (i.e. 6 and 12 weeks). The model will include the baseline value of the outcome variable, week (as a categorical factor), and the interaction between the baseline value and week, as well as covariates such as age, sex, BMI, and presence of comorbidities. Unstructured covariance matrix will be assumed for the model errors, unless additional structure is required for convergence. If visual inspection of the model diagnostics (i.e. residuals) suggest that a transformation of the endpoint should be performed, this will be applied as appropriate prior to analysis (i.e. change from baseline would be calculated on the transformed scale) and documented in the final report as required. The Least Squares Means (LSMeans) together with 95% confidence intervals will be obtained for each visit and plots will be produced illustrating the trajectory of the LSMeans over time. The above outputs will be back-transformed to the original scale for transformed endpoints as required. Missing values will be accounted for by utilizing a maximum likelihood-based approach as part of the MMRM assumptions.

The primary analyses will be applied to each endpoint separately.

Sensitivity analyses will include (a) only in participants who attended at least 80% (14 visits) of the PT visits, and (b) in all participants stratified into < 6 PT visits attended, 6-12 PT visits attended, and >12 PT visits attended.

D. Assess the effects of exercise-based PT on outcomes from wearable sensors collected at-home

The following definitions will be used for analysis of at home sensor data. Based on the literature[9-11] , a recording day will be considered valid and included in the analysis if it contains at least 10 waking hours (identified using the Cole-Kripke algorithm implemented in ActiLife software) of wear-time. At-home wrist and lumbar sensor data will be summarized on a weekly basis. A week of recording will be included in the analysis if there is at least 4 valid days (as described above) in a 7-day period[11].

A subset of wearable sensor metrics for gait and sit-to-stand from Section 3.1. E. that are shown to have excellent reliability from analysis A, and the physical activity measures from Actigraph (Section 3.1. E) will be examined. Since *number of gait bouts* derived from free living data is not assessed in

the lab, this will be automatically included in this analysis. The same MMRM approach and sensitivity analyses described above will be used to investigate change from baseline to 6- and 12-weeks with PT on at home sensor endpoints. Mean/median for each endpoint (depending on the distribution of the endpoint) across a valid period will be used in the analyses.

E. Assess the effects of exercise-based PT in-lab tests of physical performance.

The same MMRM approach and sensitivity analyses described above will be used to investigate change from baseline to 6- and 12-weeks with PT on all physical performance outcomes described in Section 3.1. D.

F. Assess the effects of exercise-based PT using joint mechanics during in-lab functional biomechanical tests

The same MMRM approach and sensitivity analyses described above will be used to investigate change from baseline to 6- and 12-weeks with PT on all endpoints derived from optical motion capture system (Gait: Table 1. Column 2; Sit-to-stand: Table 2. Column 2; Section 3.1. C) during in-lab visits.

4.3. Analysis of Secondary Endpoint(s)

Secondary Aim 1: Compare in-lab and real world (at-home) sensor metrics with ePRO outcomes

A. Compare change in in-lab wearable sensor metrics with ePRO outcomes

The change from baseline in ePRO measures from Section 3.2 will be calculated using the same MMRM approach detailed in Section 4.2 C. ePRO measures that show a significant change from baseline to 12-weeks will be examined for this analysis.

Wearable sensor metrics showing a significant change from baseline to 12-weeks during in-lab physical performance tests will be examined (this is the outcome of analysis performed in Section 4.2. C).

The change from baseline to week 12 in these metrics will be correlated with change from baseline in ePRO outcomes using Pearson's product-moment correlations or Spearman's rank order correlations depending on their normality.

Sensitivity analyses will be conducted to estimate the correlation between wearable sensor metrics and ePRO outcomes in (a) responders and (b) non-responders. Responders are defined as participants with ≥50% reduction in KOOS-derived WOMAC 3.0 pain score[18].

B. Compare change in at-home wearable sensor metrics with ePRO outcomes

ePRO measures from Section 3.2 that show a significant change from baseline to 12-weeks will be examined (as determined by analyses in Section 4.3 A). Wearable sensors metrics which show a significant change from baseline to week 12 during at-home wear will be examined (the outcome of analysis performed in Section 4.2. D).

Change from baseline to week 12 in these sensor metrics will be correlated with change in ePRO outcomes using Pearson's product-moment correlations or Spearman's rank order correlations depending on their normality.

Sensitivity analyses will be conducted to estimate the correlation between wearable sensor metrics and ePRO outcomes in (a) responders and (b) non-responders. Responders are defined as participants with ≥50% reduction in KOOS-derived WOMAC 3.0 pain score[18].

Secondary Aim 2: Assess the effect of PT on sleep using ePRO (DSIS questionnaire) and wearable sensor metrics

The same MMRM approach described above will be used to investigate change from baseline to 6- and 12-weeks with PT on sleep outcomes described in Section 3.2.

Secondary Aim 3: Assess the persistence of the PT effect from week 12 to week 18 using in-lab and real world (at-home) sensor metrics, and ePROs

In-clinic and at-home endpoints that show a significant change from baseline to week 12 will be examined (output of analyses 4.2. C. and D.). ePRO measures from Section 3.2 that show a significant change from baseline to 12-weeks will be examined (as determined by analyses in Section 4.3. A). Similar MMRM approach to that described above will be used to investigate change in these outcome measures from baseline to 18-week timepoints. The outcome measures will be expressed as the change from baseline at each visit (i.e. 6, 12 and 18 weeks). We will compare the effects at week 12 to week 18 by using linear contrasts.

4.4. Analysis of Exploratory Endpoint(s)

For all these analyses, participants with ≥50% reduction in KOOS-derived WOMAC 3.0 pain score will be classified as responders[18].

Exploratory Aim 1: Assess whether sensor-based baseline metrics can stratify treatment response

Wearable sensor metrics with excellent reliability from analysis in Section 4.2.A will be examined. Wearable sensor metrics from in-lab physical performance testing (Section 3.2.A and 3.2.B) and baseline at-home testing (3.2.4 gait and sit to stand metrics) will be modeled as continuous and categorical (tertiles) exposures. Multivariate logistic regression will be used to assess whether baseline wearable sensor metrics can predict treatment response while accounting for confounders such as age, sex, BMI, and baseline KOOS-derived WOMAC 3.0 pain score.

Exploratory Aim 2: Assess whether participant phenotype; pain, physical performance, affective traits (catastrophizing) can stratify treatment response.

The following exposure definitions will be used:

- Pain phenotypes
    - Tertiles of PPT
    - Presence/absence of TS
    - Inadequate/adequate CPM
    - Inadequate/adequate EIH
    - Presence/absence of pain catastrophizing from pain catastrophizing questions

  - Presence/absence of widespread pain from joint homunculus
  - PainDETECT score (continuous)
- Physical performance
  - Tertiles of physical performance tests
  - Tertiles of strength tests

Multivariate logistic regression will be used to assess whether baseline exposure variables can predict treatment response while accounting for confounders such as age, sex, BMI, and baseline KOOS-derived WOMAC 3.0 pain score.

4.5. Analysis of Substudy Endpoint(s)

Substudy Aim 1: Examine the test-retest reliability of at-home instrumented physical performance tests

Test retest reliability of sensor metrics derived during the two sets of at-home physical performance tests (7-meter gait task, 5 times sit to stand tasks) will be assessed using ICC and correlation analyses. Pearson's product-moment correlations or Spearman's rank order correlations will be used depending on their normality. These analyses will be conducted on all sensor metrics derived from the 3-sensor IMU system and GaitPy and Sit2StandPy applied to the lumbar sensor (Tables 3. and 4.).

Substudy Aim 2: Examine agreement between wearable sensor data collected during in-lab and at-home instrumented physical performance tests

We will assess agreement between wearable sensor metrics collected in-lab and those collected at-home during the same physical performance tests (7-meter gait, 5 times sit to stand) from the 3-sensor IMU system and the lumbar sensor separately. Agreement will be assessed using ICC analyses as well as Bland–Altman plots and 95% limits of agreement between at-home and in-clinic sensor metrics for both systems (using metrics listed in Tables 3. and 4.)

As stated in the protocol, these analyses will be performed on a subset of participants (n=12). Once 12 participants have had the opportunity to complete the study, an interim analysis may be conducted (while continuing to enroll) to determine if additional participants are required to address the aims of this substudy.

4.6. Safety Summary

Adverse events (AEs) will be summarized and presented in a table.

APPENDICES

Appendix 1. References

1. Czech, M.D. and Patel, S., *GaitPy: An Open-Source Python Package for Gait Analysis Using an Accelerometer on the Lower Back.* J Open Source Software, 2019. **4**(43): p. 1778.
2. Winter, D.A., *Biomechanics and Motor Control of Human Movement*. Fourth ed. 2009: John Wiley & Sons, Inc.
3. Holsgaard-Larsen, A., Clausen, B., Søndergaard, J., Christensen, R., Andriacchi, T.P. and Roos, E.M., *The effect of instruction in analgesic use compared with neuromuscular exercise on knee-joint load in patients with knee osteoarthritis: a randomized, single-blind, controlled trial.* Osteoarthritis Cartilage, 2017. **25**(4): p. 470-480. DOI: 10.1016/j.joca.2016.10.022.
4. Kumar, D., Swanik, C.B., Reisman, D.S. and Rudolph, K.S., *Individuals with medial knee osteoarthritis show neuromuscular adaptation when perturbed during walking despite functional and structural impairments.* J Appl Physiol (1985), 2014. **116**(1): p. 13-23. DOI: 10.1152/japplphysiol.00244.2013.
5. Adamowicz, L., Karahanoglu, F.I., Cicalo, C., Zhang, H., Demanuele, C., Santamaria, M., Cai, X., and Patel, S., *Assessment of Sit-to-Stand Transfers during Daily Life Using an Accelerometer on the Lower Back.* Sensors (Basel), 2020. **20**(22). DOI: 10.3390/s20226618.
6. Petrella, M., Serrão, P., Selistre, L.F.A., Lessi, G.C., Gonçalves, G.H. and Mattiello, S.M., *Individual joint contributions to the total support moment during the sit-to-stand task differentiate mild and moderate knee osteoarthritis.* Clin Biomech (Bristol, Avon), 2019. **70**: p. 52-58. DOI: 10.1016/j.clinbiomech.2019.07.029.
7. Winter, D.A., *Biomechanics and motor control of human movement*. 1990, John Wiley & Sons, Inc.: New York, NY.
8. Lawson, T., Morrison, A., Blaxland, S., Wenman, M., Schmidt, C.G. and Hunt, M.A., *Laboratory-based measurement of standing balance in individuals with knee osteoarthritis: A systematic review.* Clinical Biomechanics, 2015. **30**(4): p. 330-342. DOI: https://doi.org/10.1016/j.clinbiomech.2015.02.011.
9. Choi, L., Ward, S.C., Schnelle, J.F. and Buchowski, M.S., *Assessment of wear/nonwear time classification algorithms for triaxial accelerometer.* Med Sci Sports Exerc, 2012. **44**(10): p. 2009-16. DOI: 10.1249/MSS.0b013e318258cb36.
10. Song, J., Semanik, P., Sharma, L., Chang, R.W., Hochberg, M.C., Mysiw, W.J., Bathon, J.M., Eaton, C.B., et al., *Assessing physical activity in persons with knee osteoarthritis using accelerometers: data from the osteoarthritis initiative.* Arthritis Care Res (Hoboken), 2010. **62**(12): p. 1724-32. DOI: 10.1002/acr.20305.
11. Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T. and McDowell, M., *Physical activity in the United States measured by accelerometer.* Med Sci Sports Exerc, 2008. **40**(1): p. 181-8. DOI: 10.1249/mss.0b013e31815a51b3.
12. Bagui, S., Fang, X., Bagui, S., Wyatt, J., Houghton, P., Nguyen, J., Schneider, J., and Guthrie, T., *An improved step counting algorithm using classification and double autocorrelation.* International Journal of Computers and Applications, 2020: p. 1-10. DOI: 10.1080/1206212X.2020.1726006.
13. Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J. and Gillin, J.C., *Automatic sleep/wake identification from wrist activity.* Sleep, 1992. **15**(5): p. 461-9. DOI: 10.1093/sleep/15.5.461.
14. Actigraph. *What does the "Detect Sleep Periods" button do and how does it work?* ; Available from: https://actigraphcorp.force.com/support/s/article/What-does-the-Detect-Sleep-Periods-button-do-and-how-does-it-work.

15. Washabaugh, E.P., Kalyanaraman, T., Adamczyk, P.G., Claflin, E.S. and Krishnan, C., *Validity and repeatability of inertial measurement units for measuring gait parameters.* Gait Posture, 2017. **55**: p. 87-93. DOI: 10.1016/j.gaitpost.2017.04.013.
16. Holstrom, L. *How are Mobility Lab's algorithms validated?* 2020 [cited 2022; Available from: https://support.apdm.com/hc/en-us/articles/360000177066-How-are-Mobility-Lab-s-algorithms-validated-
17. Cicchetti, D.V., *Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.* Psychological Assessment, 1994. **6**(4): p. 284-290. DOI: 10.1037/1040-3590.6.4.284.
18. Schnitzer, T.J., Easton, R., Pang, S., Levinson, D.J., Pixton, G., Viktrup, L., Davignon, I., Brown, M.T., et al., *Effect of Tanezumab on Joint Pain, Physical Function, and Patient Global Assessment of Osteoarthritis Among Patients With Osteoarthritis of the Hip or Knee: A Randomized Clinical Trial.* Jama, 2019. **322**(1): p. 37-48. DOI: 10.1001/jama.2019.8044.