Statistical Analysis Plan

Delaware Physical Exercise and Activity for Knee osteoarthritis

(Delaware PEAK)

IRB Protocol Number: 1730922

National Clinical Trial (NCT) Identified Number: NCT04980300

Principal Investigator (PI): Daniel K. White

Sponsor: University of Delaware

Funded by: Rheumatology Research Foundation

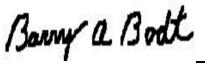
Date of SAP: 7/14/2023

Version	Date	Summary of changes
v0.01	**07/14/2023**	**Original**

1 SAP Signatures

I give my approval for the attached SAP entitled “Delaware Physical Exercise and Activity for Knee osteoarthritis (Delaware PEAK)” dated 06/14/2023

Statistician (Author)
Name: Barry A. Bodt

Signature: 

Date: July 14, 2023

Statistician Reviewer (As applicable)
Name:

Signature: _____

Date: _____

Principal Investigator
Name: Daniel K. White

Signature:
Date: 7/14/2023

2 Table of Contents

1 SAP Signatures 2
2 Table of Contents 3
3 Abbreviations and Definitions 5
4 Introduction 5
4.1 Preface 5
4.2 Scope of the analyses 6
5 Study Objectives and Endpoints 6
5.1 Study Objectives 6
5.2 Endpoints 6
6 Study Methods 6
6.1 General Study Design and Plan 6
6.2 Inclusion-Exclusion Criteria and General Study Population 8
6.3 Randomization and Blinding 8
6.4 Study Assessments 9
7 Sample Size 12
8 General Analysis Considerations 12
8.1 Timing of Analyses 12
8.2 Analysis Populations 12
8.2.1 Full Analysis Sample (Intention to Treat) 12
8.2.2 Per Protocol Sample, set 1 12
8.2.3 Per Protocol Sample, set 2 13
8.2.4 Secondary Data Analysis Sample 13
8.3 Covariates and Subgroups 13
8.4 Missing Data 13
9 Summary of Study Data 14
9.1 Subject Disposition 14
9.2 Protocol Deviations 16
9.3 Demographic and Baseline Variables 17
9.4 Treatment Compliance 17
10 Efficacy Analyses 17
10.1 Primary Efficacy Analysis 18
10.2 Secondary Efficacy Analyses 18
10.2.1 Secondary Analyses of Primary Efficacy Endpoint 18
10.2.2 Analyses of Secondary Endpoints 18
10.3 Exploratory Efficacy Analyses 18
11 Safety Analyses 19
11.1 Adverse Events 19

11.2 Deaths, Serious Adverse Events and other Significant Adverse Events 19
12 Reporting Conventions 19
13 Quality Assurance of Statistical Programming (As Applicable) 19
14 Summary of Changes to the Protocol and/or SAP 20
15 References 20
16 Listing of Tables, Listings and Figures 21

3 Abbreviations and Definitions

AE	Adverse Event
CRF	Case Report Form
SAP	Statistical Analysis Plan
PT	Physical Therapy
MVPA	Moderate-to-Vigorous intensity Physical Activity
SED	Sedentary Behaviors
LPA	Light Intensity Physical Activity
NICE	National Institute for Health and Care Excellence
IPAQ	International Physical Activity Questionnaire

4 Introduction

4.1 Preface

Knee Osteoarthritis (OA) is a serious disease. In 2010, OA moved up from 15th to 11th in rankings for global burden of disease,[1] and the most common weight-bearing joint affected by OA is the knee.[2] More than 15 million Americans currently have symptomatic knee OA.[3] The estimated annual cost to treat knee OA is $27 billion,[4] which is similar to the cost to treat stroke (~$34 billion[5]). The incidence of knee OA is expected to increase with the aging and obese US population.[6] Walking difficulty is the primary functional limitation in people with knee OA that, in turn, makes OA the leading cause of disability among older adults.[7] Not only is walking difficulty related to disability, but it is also associated with a 55% higher risk of premature death.[8,9] Without a cure, treatment for knee OA focuses primarily on managing pain and reducing disability.

Treatment guidelines for knee OA champion the use of supervised exercise. The recently released 2019 ACR/Arthritis Foundation guidelines for the treatment of OA strongly support the use of supervised exercise.[10] These recommendations are consistent with those of the Osteoarthritis Research Society International (OARSI), which promote exercise, self-management, and education.[11] Both guidelines (ACR and OARSI) come from the robust and pervasive research findings that exercise reduces pain and improves physical function in adults with knee OA.[8]

Few adults with knee OA try supervised exercise to manage their OA. For example, less than 10% of adults with knee OA exercise regularly,[12,13] and only one-in-ten patients are seen for supervised exercise 5 years before a knee replacement.[14] Health claims data show a similar trend, with a doubling in the prescription of opioids and a 50% reduction in referrals for Physical Therapy (PT) in the past decade.[15] Hence, few adults with knee OA are utilizing physical activity and/or supervised exercise to address their symptoms of knee OA.

The objective of this study is to examine the efficacy of Delaware PEAK to increase physical activity in adults with knee OA compared to a control group receiving web-based resources about knee OA and exercise. The rationale for our study is that there is a need to examine whether Delaware PEAK can directly target the mismatch between OA recommendations and practice patterns. Our central hypothesis is that Delaware PEAK will increase physical activity and will increase the belief that exercise is helpful and not harmful, compared with a

control group receiving web-based OA treatment resources. Successful completion of this proposal will provide the evidence necessary to scale up this low-cost intervention, with the goal of increasing the number of adults who use exercise to manage their knee OA and thus reducing the burden of disease.

4.2 Scope of the analyses

These analyses will assess the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK) versus a control group receiving web-based resources about knee OA and exercise regarding change in Moderate-to-Vigorous intensity Physical Activity (MVPA) over 12 weeks in adults with knee OA. We call the intervention group an Expanded Intervention and the control group a Brief Intervention.

5 Study Objectives and Endpoints

5.1 Study Objectives

The purpose of this study is to examine the efficacy of Delaware PEAK to increase physical activity in adults with knee OA compared to a control group receiving web-based resources about knee OA and exercise.

5.2 Endpoints

The primary endpoint of our study is to examine the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK) to increase MVPA over 12 weeks compared to a control group receiving web-based resources about knee OA and exercise.

The secondary analyses endpoint of our study is to examine the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK) to increase health beliefs, light physical activity (LPA), and steps per day, over 24 weeks compared to a control group receiving web-based resources about knee OA and exercise. We also will assess change in MVPA over 24 weeks, the number of Adverse events over 12 weeks, and treatment adherence.

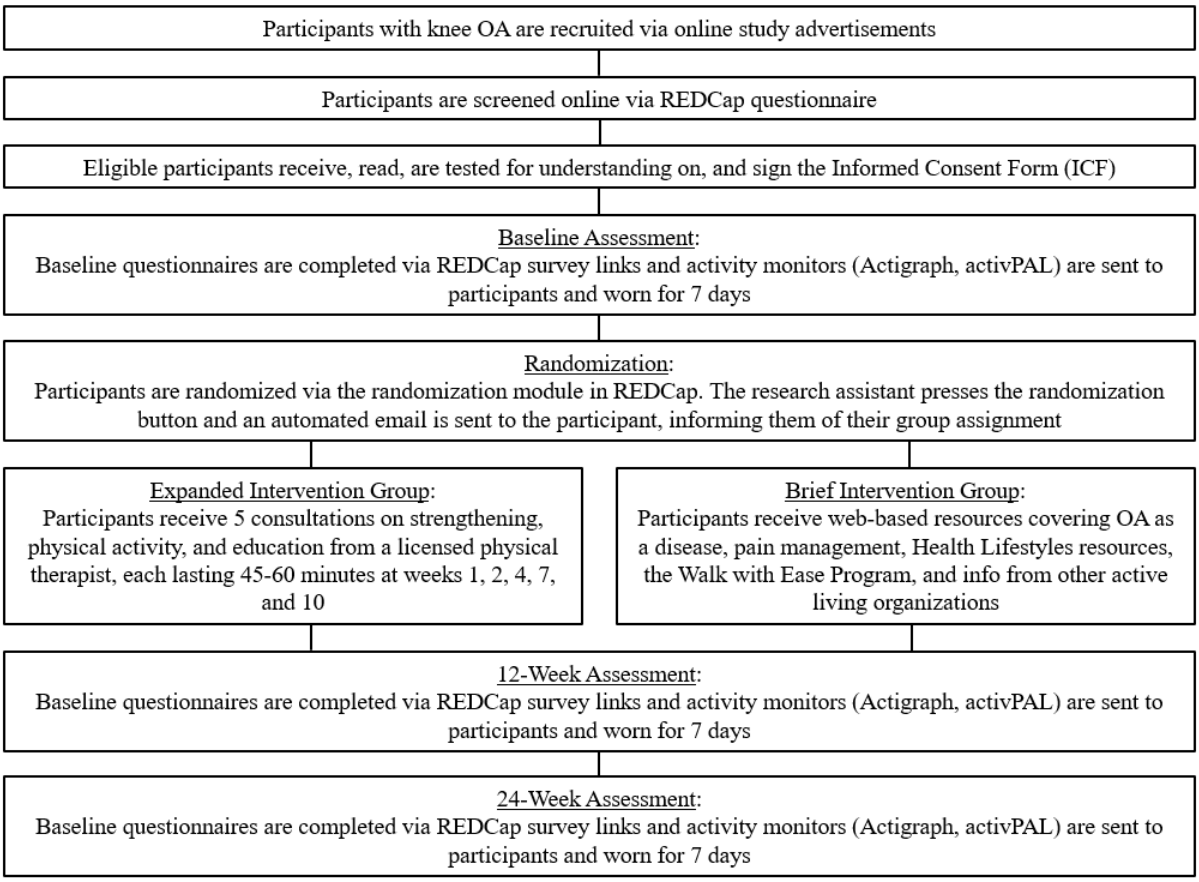
6 Study Methods

6.1 General Study Design and Plan

(ICH E3;9)

We will conduct a pragmatic randomized controlled trial with a 2-group, superiority, parallel-design. The intervention group will receive a physical therapist-delivered exercise intervention (Delaware PEAK), i.e., the Expanded Intervention, while the control group will receive web-based resources about knee OA and exercise, i.e., the Brief Intervention. The study was assessor-blinded, i.e., all members of the research team who were involved with the assessment of the primary and secondary outcomes were blinded. Participants will not be blinded to group assignment but will not be informed about study hypotheses until study completion. The physical therapist will not be blinded to group assignment.

Figure 1: Study flow-chart



6.2 Inclusion-Exclusion Criteria and General Study Population

(ICH E3;9.3. ICH E9;2.2.1)

In order to be eligible to participate in this study, an individual must meet all of the following criteria:

1. Provision of an electronically signed and dated informed consent form
2. ≥45 years of age,
3. Resides in the contiguous United States
4. Available for the duration of the intervention portion of the study (12 weeks) and willing to wear physical activity monitors
5. Looking to move more, i.e., be more active.
6. Knee OA diagnosis by the NICE criteria
7. Comfortable participating in a program delivered in English
8. Is able to safely participate in moderate-intensity exercise as determined by a pre-exercise screen questionnaire,[29]
9. Has either a smartphone or a laptop/desktop computer with an internet connection.
10. Has a working email address.

An individual who meets any of the following criteria will be excluded from participation in this study:

1. Regularly exercise for more than 60 minutes/week.
2. Has a scheduled knee or hip joint replacement
3. Has had physical therapy for knee OA in the past 6 months
4. Participated in a strength training program for the lower extremities in the past 6 months

6.3 Randomization and Blinding

(ICH E3; 9.4.3, 9.4.6. ICH E9; 2.3.1, 2.3.2)

Participant-level randomization will be performed to assign each participant to either the Intervention group (referred to as Expanded Intervention) or the Control group (referred to as Brief Intervention) using block randomization schedule with block sizes of 4. Randomization was performed using REDCap's built-in randomization function and will occur after participants sign the informed consent and complete baseline data collection. Participants will be randomized into one of two groups: (1) the Expanded Intervention group, who will receive the Delaware PEAK program with a physical therapist, or (2) the Brief Intervention group, who will receive pre-recorded, web-based resources.

Participants will not be blinded to group allocation; however, they will not be informed about the study hypotheses until the study is completed, at which time they will be provided a lay summary of findings. The treating physical therapist will not be blinded to group assignment, as they are delivering the Expanded Intervention. Research team members responsible for outcome data management will be blinded to group assignment. As all outcomes are collected either through a physical activity monitor or online questionnaires, this study is thus considered assessor-blinded for our primary outcome, physical activity. Statistical analyses will be performed in a blinded manner.

6.4 Study Assessments

(ICH E3; 9.5.1. ICH E9; 2.2.2)

Table 1: Timeline of Assessments

	STUDY PERIOD								
	Enrollment/BL	**Randomization**	**Intervention**					**Follow-Up**	
Timepoint	*-t_1*	*0w*	*1w*	*2w*	*4w*	*7w*	*10w*	*12w*	*24w*
Enrollment:									
Screening Form	X								
Verbal Eligibility Form	X								
Informed Consent	X								
General Information (Demographics)	X								
Randomization		X							
Intervention:									
Expanded Intervention			X	X	X	X	X		
Brief Intervention			X	X	X	X	X		
Assessment:									
Primary outcome									
Physical activity (MVPA via Actigraph)	X							X	X
Secondary outcomes									
Treatment Beliefs of Physical Therapy	X							X	X
Treatment Beliefs of Physical Exercise	X							X	X
Physical activity (steps/day, LPA via Actigraph)	X							X	X
Other measures									
Charlson Comorbidity Index	X								
Knee Pain (Visual Analogue Scale)	X							X	X
Knee Injury and OA Outcome Score	X							X	X
Treatment Expectations Question	X	X							
Adverse Events (check, recording)		X	X	X	X	X	X	X	X
Exercise Adherence Rating Scale (EARS)*								X	X
*Collected in the those randomized to the expanded intervention group									

Table 2: Analysis Time Windows

Visit (target day)	**Lower bound (days)**	**Upper bound (days)**
Enrollment (0)	n/a	n/a
Baseline (14)	-7 days	+7 days
12-week Follow-up (112)	-7 days	+28 days
24-week Follow-up (206)	-7 days	+28 days

Data collected outside of the time windows listed above will be classified as the intended visit the data was to reflect and noted as being outside the time window. Such data will be included in the primary analyses, however we will also perform a secondary analyses excluding data collected outside the time window. While the spacing of follow-up assessments makes it unlikely that multiple data collections will occur during the same window, the first valid set of assessments will be used for the corresponding timepoint.

Descriptions of each outcome, timepoint of collection, and score range are described below. A table is also provided that summarizes this information.

MVPA: Objectively-measured physical activity will be collected using the Actigraph GT3x at the baseline, 12-week, and 24-week timepoints. The primary outcome is change in MVPA, in minutes/week, from baseline to 12 weeks. We also will assess change in MVPA from baseline to 24 weeks. Minutes/week of MVPA can range from 0 minutes/week to 10,080 minutes/week (7 days x 24 hours x 60 minutes).

LPA, Daily Walking: Data collected using the ActiGraph GT3x activity monitor will also be used to assess change in minutes/week of LPA and average steps/day between baseline and both follow-up timepoints. Minutes/week of LPA can range from 0 minutes/week to 10,080 minutes/week (7 days x 24 hours x 60 minutes). Steps/day can be any value greater than or equal to 0.

Health Beliefs: Health beliefs will be measured with the Treatment Beliefs of OA: Physical Therapy (TOA-PT) and Treatment Beliefs of OA: Physical Exercise (TOA-PE) questionnaires, collected at the baseline, 12-week, and 24-week timepoints. The TOA provides separate scores for positive and negative treatment beliefs. The primary outcome is change in TOA-PT and TOA-PE scores from baseline to 12 weeks. We also will assess change in TOA-PT and TOA-PE scores from baseline to 24 weeks. The TOA-PT negative beliefs scores range from 4 to 16, TOA-PT positive scores range from 7 to 28, TOA-PE negative beliefs scores range from 4 to 16, TOA-PE positive scores range from 7 to 28. For all scales, a higher score indicates a greater amount of positive or negative treatment beliefs.

Charlson Comorbidity Index (CCI): The Charlson Comorbidity Index (CCI) is a 17-item questionnaire that creates a weighted score based on a range of common, comorbid conditions. This questionnaire will be collected at the baseline timepoint and scores range from 0 to 29, with a higher CCI score indicating greater levels of comorbidity.

Visual Analog Scale (VAS) Pain: Change in knee pain will be measured using a single-item Visual Analog Scale (VAS) question in which the participant will indicate their pain in each knee. Scores range from 0-100 where 0 is no pain and 100 is the worst pain imaginable.

Knee Injury and Osteoarthritis Outcome Score (KOOS): The Knee Injury and Osteoarthritis Outcome Score (KOOS) is a 42-item questionnaire that assesses five domains, including pain, symptoms (other than pain), function in activities of daily living, function in sport and recreation, and quality of life. The KOOS has subscale scores for Pain, Symptoms, ADL Function, Sport & Recreation Function, and Quality of Life. Scores are transformed to a 0-100 scale with 0 indicating extreme knee problems and 100 indicating no knee problems.

Treatment Expectations Question: The participant will complete a Treatment Expectations Question, which asks the participant how effective they expect the intervention to be for their knee. This question will be collected at the baseline timepoint: right before and right after

randomization. Answers will be expressed as frequencies and percentages for each of the following categories: No effect at all, Minimal improvement, Moderate improvement, Large improvement, and Complete recovery, as well as reported as an increase, decrease, or no change in treatment expectations.

Exercise Adherence: For the Expanded Intervention, adherence to strengthening exercises and physical activity will be measured using the Exercise Adherence Rating Scale (EARS). The EARS includes 6 items whose total score ranges from 0 to 24, where a higher score indicates greater adherence. The EARS also includes 10 items on what helps or hinders exercise adherence. These scores range from 0 to 40, where a higher score indicates better outcome.

Brief Intervention Utilization: For the Brief Intervention, intervention utilization will be measured via a self-report questionnaire administered on REDCap. The participant will be asked to confirm that they viewed the informational videos, accessed the associated weblink, and note how many times they visited the site for each webpage section. Percentage of Brief Intervention participants that utilize the website will be reported. Utilization of the Brief Intervention will be measured at the 12-week follow-up timepoint.

Adverse Events (AE): Information on potential adverse events (AE) will be collected by the treating physical therapist at the beginning of each session in the Expanded Intervention, by a Research Assistant during follow-up reminder phone calls in the Brief Intervention, or by spontaneous report on the part of the participant (e.g., reaching out to the study team by phone or email). AEs will be reported as frequencies and percentages.

General Information Form: The participant will complete a General Information Form that collects height, weight, date of birth, gender, race, ethnicity, veteran status, highest level of education attained, employment status, and approximate household income. This form will be collected at the baseline timepoint. Continuous variables will be expressed as means and standard deviations while categorical variables will be expressed as frequencies and percentages.

Table 3: Outcome Timepoints and Scoring

Outcome	**Timepoints collected**	**Scoring**
MVPA	Baseline, 12-week, 24-week	0 to 10,080 minutes/week
TOA (PT and PE)	Baseline, 12-week, 24-week	4 to 16 for negative 7 to 28 for positive
EARS	12-week, 24-week	0 to 24
Telehealth Acceptability	12-week (Expanded only)	14 to 70
Participant Satisfaction	12-week	0 to 6
LPA	Baseline, 12-week, 24-week	0 to 10,080 minutes/week
Daily Walking	Baseline, 12-week, 24-week	At least 0 steps/day
VAS Pain	Baseline, 12-week, 24-week	0 to 100

KOOS	Baseline, 12-week, 24-week	0 to 100 for each subscale
General Information Form	Baseline	n/a
Treatment Expectations	Baseline (x2)	n/a
CCI	Baseline	0 to 29
Adverse Events	Throughout	n/a

7 Sample Size

(ICH E3; 9.7.2. ICH E9; 3.5)

The sample size is n = 100 randomized participants. In order to be randomized, participants must complete baseline assessments.

8 General Analysis Considerations

8.1 Timing of Analyses

Final analyses will be performed when data collection is completed. In particular, the final analyses will be performed when all enrolled study subjects have completed the 24-week follow up timepoint or dropped out prior to the 24-week timepoint. The final analysis will be performed on data transferred to a file from REDCap, having been documented as meeting the cleaning and approval requirements after the finalization and approval of this SAP document. Data cleaning will involve checking REDCap surveys for errors/discrepancies, ensuring all responses are correctly scored REDCap, and then performing a quality check to ensure answers were scored accurately. For the ActiGraph, we will ensure data was recorded at 60Hz, that data fell within the expected wear time range, and that included participants met our wear time criteria of at least 4 days and at least 10 hours on those days. As well, we will cross-reference the date the monitor was worn from the Actigraph is between the dates the monitor was send the received from REDCap. We will also check that the file names and dates of the Actigraph files correspond to names and dates from REDCap.

8.2 Analysis Populations

(ICH E3; 9.7.1, 11.4.2.5. ICH E9; 5.2)

8.2.1 Full Analysis Sample (Intention to Treat)

- *All subjects who were randomized*
- *The ActiGraph monitor was worn for ≥ 4 days of 10 hours at the baseline and 12-week follow-up timepoints*
- *This sample will be the primary efficacy sample*

8.2.2 Per Protocol Sample, set 1

- *All subjects who adhere to the major criteria in the protocol*
    - *Subjects who did not drop out*

        - *Completed their respective intervention, (60% of Intervention)*
        - *Did not drop out over the 12 weeks of follow up after the intervention*
    - *The ActiGraph monitor was worn ≥ 4 days at each time point (baseline and 12-week follow-up timepoints).*

8.2.3 Per Protocol Sample, set 2

- *All subjects who adhere to the major criteria in the protocol*
    - *Subjects who did not drop out*
        - *Completed their respective intervention (60% of Intervention)*
        - *Did not drop out over the 24 weeks of follow up after the intervention*
    - *The ActiGraph monitor was worn ≥ 4 days at each time point (baseline, 12-week, 24-week follow-up timepoints).*

8.2.4 Secondary Data Analysis Sample

- *We will repeat the Intention to Treat and Per Protocol analyses excluding participants who worn their activity monitors outside of our established visit windows.*

The exact process for assigning the statuses (Full Analysis Sample and Per Protocol Sample) will be defined and documented prior to breaking the blind along with any predefined reasons for eliminating a subject from a particular population.

8.3 Covariates and Subgroups

(ICH E3; 9.7.1, 11.4.2.1. ICH E9; 5.7)

Baseline demographics will be assessed for group assignment differences after randomization. Covariates will be determined based on the presence of group differences.

8.4 Missing Data

(ICH E3; 9.7.1, 11.4.2.2. ICH E9;5.3. EMA Guideline on Missing Data in Confirmatory Clinical Trials)

Missing data will be given careful attention in the analysis. In this RCT we have two treatment arms and three fixed times for assessment. It is anticipated that drop-outs from the study will be at random because study participants were considered treatment-seeking. However, it is possible that participants may not attend all follow-ups. Drastically fewer participants in the final analysis would affect study sensitivity and excessive missed follow-ups would potentially bias model parameter estimates. Under the assumption of limited data that is missing completely at random (MCAR) or missing at random (MAR), the mixed-effect model using restricted maximum likelihood estimation is a full information approach and leads to unbiased model parameter estimates. Selection between models with two common covariance structures, Toeplitz and independent errors will be accomplished using the Akaike Information Criterion (AIC).

In the event that assumptions are not met regarding missing value frequency, MAR, or MCAR and missing values are occurring with higher frequency or are truly missing not at random (MNAR), the analysis may be biased. To test the missingness assumptions, missing data will first be summarized by group, time, each design cell, and each of the three

covariates, age, sex and body mass index in an effort to graphically or tabularly detect patterns in the variation of missingness. More formally, missing/non-missing will be taken as a dichotomous outcome in a logistic regression for continuous covariates and will be compared with between and within factors and sex using chi-square tests. Significant logistic regression coefficients or significant chi-square tests indicate a pattern of missingness variation that will be identified and reported and whose impact on the analysis will be explained if possible. Model residuals will be evaluated for normality using the Shapiro-Wilk test, homogeneity of variance using Levine's test, and transformations, if necessary, will be chosen from among the Box-Cox transformations.

9 Summary of Study Data

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, and standard deviation. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by treatment and subject, and when appropriate by visit number within subject. All summary tables will be structured with a column for each treatment in the order (Brief Intervention, Expanded Intervention) and will be annotated with the total population size relevant to that table/treatment, including any missing observations.

9.1 Subject Disposition

We establish how many subjects reached the various stages of the trial (Enrollment, Allocation, Follow-up, and Analysis) using the following list:

Enrollment:
- Regularly exercise > 60 min/wk
- Scheduled knee replacement
- Received physical therapy or prescribed an exercise program in past 6 months
- Did not receive medical clearance after failing the Adult Pre-Exercise Screening System questionnaire
- < 45 years of age
- No activity related knee pain
- Morning stiffness > 30 min
- Reside outside of the contiguous United States
- Not comfortable with a program delivered in English
- Not seeking to be more physically active
- Did not have a smartphone/ computer with internet access
- Did not complete the baseline data collection

Allocation:
- Did not receive allocated intervention

Follow-up:
- Lost to Follow-up
- Discontinued the intervention

Figure 2: CONSORT Diagram Template

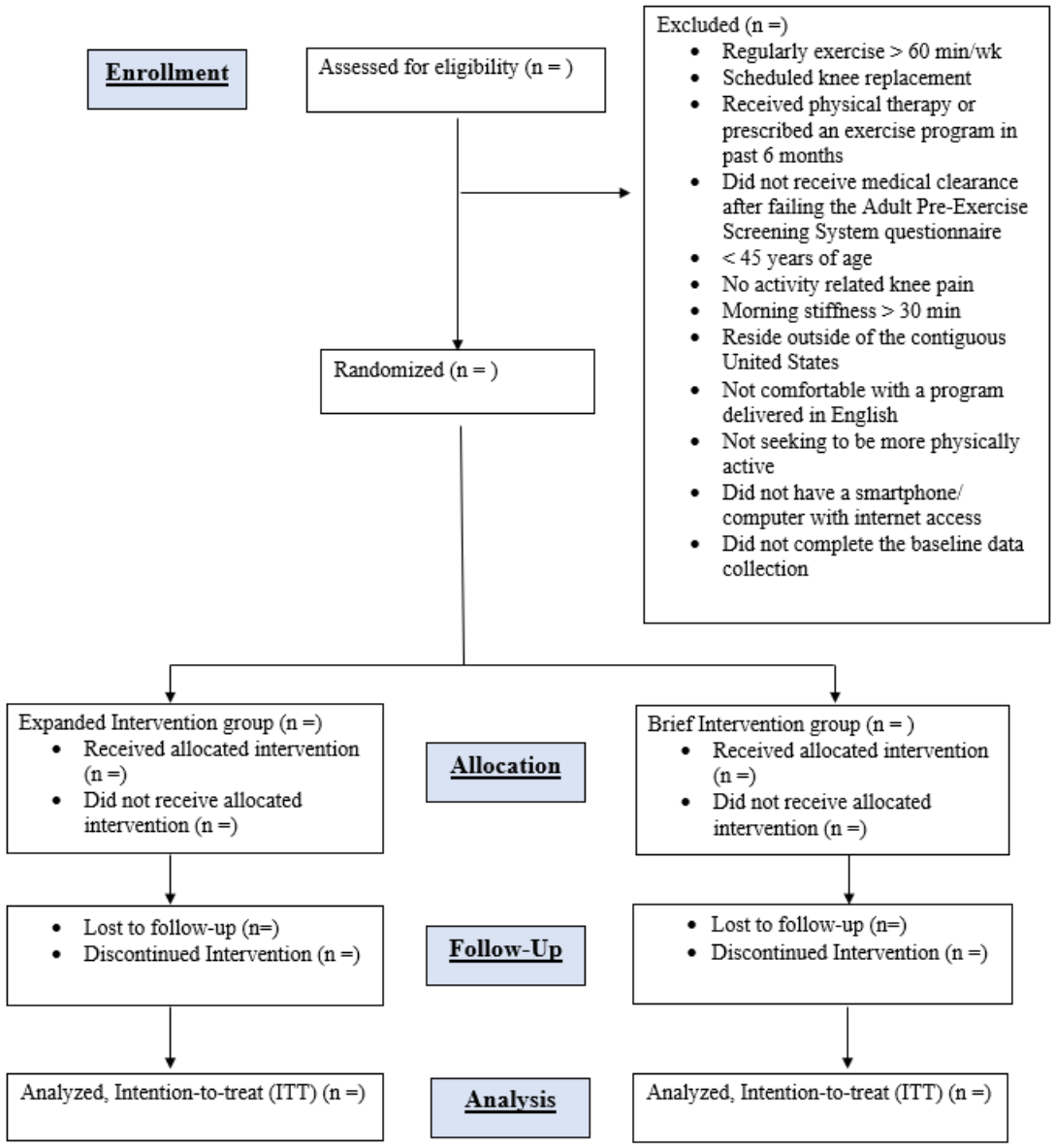


Table 4: Anticipated Recruitment Schedule

	ANTICIPATED	
Month	# to Randomize/month	Cumulative Total
July 2021	1	1
August 2021	4	5
September 2021	5	10
October 2021	6	16
November 2021	6	22
December 2021	6	28
January 2022	6	34
February 2022	6	40
March 2022	6	46
April 2022	6	52
May 2022	6	58
June 2022	6	64
July 2022	6	70
August 2022	6	76
September 2022	6	82
October 2022	6	88
November 2022	6	94
December 2022	6	100

9.2 Protocol Deviations

Potential protocol deviations include: follow-up assessments collected outside of preset time window, participants delaying assessment due to unforeseen circumstances, or the activity monitors not being sent/being lost in the mail. All would be considered missing data and would be handled in the manner described in the Missing Data section of this document.

9.3 Demographic and Baseline Variables

Table 5: Demographic and Baseline Characteristics

Characteristic [mean ± SD or n (%)]	**All *n*=**	**Expanded Intervention *n*=**	**Brief Intervention *n*=**
Age (years)			
Gender (female)			
BMI (kg/m^2)			
Race (Non-White)			
Comorbidity (CCI)			
Education (at least a college degree)			
Employment			
Income			
History of knee injury/surgery			
Duration of OA diagnosis			
MVPA at baseline (minutes/day)			
Steps/day at baseline			

9.4 Treatment Compliance

Table 6: Treatment Compliance Measures

Characteristic [mean ± SD or n (%)]	**All *n*=**	**Expanded Intervention *n*=**	**Brief Intervention *n*=**
Average number of Intervention sessions completed			
Percentage of participants completing 60% of Intervention sessions			

10 Efficacy Analyses

The primary and secondary outcomes will be summarized stratified by treatment group at each study timepoint (baseline, 12-week, 24-week). Our primary analysis will use an intention-to-treat (ITT) analysis that includes all participants who wore the monitor for ≥4 days for 10 hours/day at baseline. The statistical model underlying the analysis a mixed-effects model to examine the main and interaction effects of *group* (Expanded Intervention

vs. Brief Intervention) and *time* (BL, 12-week) on MVPA (minutes/day). The null hypothesis is that there is no difference in change in MVPA in the Expanded Intervention group compared to the Brief Intervention group. The alternative hypothesis is that there is a difference between groups. The nature of the hypothesis is confirmatory as this study has been powered to detect a difference should one exist. The mixed-effect model will be used to calculate the mean and 95% Confidence Interval time in MVPA at each study time point. We will also calculate the difference in means and 95% Confidence Intervals between the Expanded Intervention and Brief Intervention group at each study time point. We will also calculate 2-sided p-values to determine statistical significance of differences between groups by time in MVPA. Data will be summarized by intervention group. Means, Standard Deviations, Minimums and Maximums will be used to summarize continuous variables. Categorical variables will be summarized as frequencies and percentages.

Mixed-effect modeling will be used to estimate MVPA over time. The mixed-effect model using restricted maximum likelihood estimation is a full information approach and leads to unbiased model parameter estimates in the presence of MCAR or MAR missingness. Selection between models with two common covariance structures, Toeplitz and independent errors, will be accomplished using the Akaike Information Criterion (AIC).

10.1 Primary Efficacy Analysis

The primary endpoint of our study is to examine the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK, referred to as the Expanded Intervention) to increase MVPA over 12 weeks compared to a control group receiving web-based resources about knee OA and exercise (referred to as the Brief Intervention).

The statistical model underlying this analysis is a mixed-effects model to examine the main and interaction effects of *group* (Expanded Intervention vs. Brief Intervention) and *time* (baseline, 12-week) on MVPA (minutes/day).

10.2 Secondary Efficacy Analyses

10.2.1 Secondary Analyses of Primary Efficacy Endpoint

The per protocol samples, instead of the ITT sample will be used for the analysis of the primary efficacy endpoint.

10.2.2 Analyses of Secondary Endpoints

The secondary analyses endpoints are to examine the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK, referred to as the Expanded Intervention) to increase treatment beliefs, increase LPA, and increase steps/day over 12 weeks compared to a control group receiving web-based resources about knee OA and exercise (referred to as the Brief Intervention). Additionally, we will examine the efficacy of the intervention to increase MVPA over 24 weeks.

10.3 Exploratory Efficacy Analyses

Further, we will explore the efficacy of a physical therapist-delivered exercise intervention (Delaware PEAK, referred to as the Expanded Intervention) to decrease VAS pain and increase KOOS subscale scores over 12 weeks compared to a control group receiving web-based resources about knee OA and exercise (referred to as the Brief Intervention). We also will explore the change in exercise adherence from the 12-week to 24-week follow-up timepoints in the Expanded Intervention group only. Lastly, we will examine changes in treatment expectations between group from before to after randomization.

11 Safety Analyses

- All-Cause Mortality: A table of all anticipated and unanticipated deaths due to any cause, with frequencies and percentages of such events in each Intervention group will be reported.

- Serious Adverse Events: A table of all anticipated and unanticipated serious adverse events, grouped by organ system, with frequencies and percentages of such events in each Intervention group will be reported.

- Other (Not Including Serious) Adverse Events: A table of anticipated and unanticipated events (not included in the serious adverse event table) that exceed a frequency threshold of 5% within either Intervention group, grouped by organ system, with frequencies and percentages of such events in each Intervention group will be reported.

11.1 Adverse Events

We will report on the frequencies and percentages of adverse events that are possibly or definitely related to the study.

11.2 Deaths, Serious Adverse Events and other Significant Adverse Events

We will report on the frequencies and percentages of deaths, serious adverse events, and other significant adverse events that are possibly or definitely related to the study.

12 Reporting Conventions

P-values ≥0.01 will be reported to 3 decimal places; p-values less than 0.001 will be reported as “<0.001”. The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will be reported to one decimal place.

13 Quality Assurance of Statistical Programming (As Applicable)

At the time of this writing, data will be stored on REDCap, a password-protected 2-factor authentication, data management platform hosted by the Center for Human Research Coordination at the University of Delaware.

At the time of this writing, data analysis will be performed using JMP, Version 17.0.0. SAS Institute Inc., Cary, NC, USA.

For quality assurance, all summary score calculations will be performed by REDCap software and checked by a research assistant for accuracy.

14 Summary of Changes to the Protocol and/or SAP

Rationale for Adjustments of Statistical Analysis Plan from Protocol (March 12, 2023)
The changes from the protocol-specified definitions of aims, outcomes and statistical analytic approaches are outlined below. These changes reflect an error in data collection during the trial. Questionnaires are collected via automated survey in REDCap, and the automated survey did not include the SEE for the 12-week and 24-week follow-up timepoints. Therefore, SEE was only collected at the baseline timepoint and cannot be assessed for change after the intervention.

15 References

1. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I was really skeptical...But it worked really well": a qualitative study of patient perceptions of telephone-delivered exercise therapy by physiotherapists for people with knee osteoarthritis. Osteoarthritis Cartilage. 2018;26(6):741-750.
2. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I Was Really Pleasantly Surprised": Firsthand Experience and Shifts in Physical Therapist Perceptions of Telephone-Delivered Exercise Therapy for Knee Osteoarthritis-A Qualitative Study. Arthritis Care Res (Hoboken). 2019;71(4):545-557.
3. Hurley MV, Walsh N, Bhavnani V, Britten N, Stevenson F. Health beliefs before and after participation on an exercised-based rehabilitation programme for chronic knee pain: doing is believing. BMC Musculoskelet Disord. 2010;11:31.
4. Sanchez-Heran A, Agudo-Carmona D, Ferrer-Pena R, et al. Postural Stability in Osteoarthritis of the Knee and Hip: Analysis of Association With Pain Catastrophizing and Fear-Avoidance Beliefs. Pm r. 2016;8(7):618-628.
5. Lawford BJ, Bennell KL, Hinman RS. Consumer Perceptions of and Willingness to Use Remotely Delivered Service Models For Exercise Management of Knee and Hip Osteoarthritis: A Cross-Sectional Survey. Arthritis Care Res (Hoboken). 2017;69(5):667-676.
6. US Department of Health and Human Services. 2008 physical activity guidelines for Americans. 2008:1-10. www.health.gov/paguidelines/.
7. Dominick G, Zeni J, White D. The association of psychosocial factors with physical activity and function after total knee replacement: An exploratory study. Arch Phys Med Rehab. 2015:In Press.
8. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;1:Cd004376.
9. Christiansen MB, Thoma LM, Master H, et al. The feasibility and preliminary outcomes of a physical therapist-administered physical activity intervention after total knee replacement. Arthritis Care Res. 2019;25(10):23882.
10. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Care Res (Hoboken). 2020;72(2):149-162.

11. McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI Guidelines for the Non-Surgical Management of Knee Osteoarthritis. Osteoarthritis Cartilage. 2014(0).
12. White DK, Tudor-Locke C, Felson DT, et al. Do radiographic disease and pain account for why people with or at high risk of knee osteoarthritis do not meet physical activity guidelines? Arthritis Rheum. 2013;65(1):139-147.
13. Dunlop DD, Song J, Semanik PA, et al. Objective physical activity measurement in the osteoarthritis initiative: Are guidelines being met? Arthritis Rheum. 2011;63(11):3372-3382.
14. Dhawan A, Mather RC, 3rd, Karas V, et al. An epidemiologic analysis of clinical practice guidelines for non-arthroplasty treatment of osteoarthritis of the knee. Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association. 2014;30(1):65-71.
15. Khoja SS, Almeida GJ, Freburger JK. Recommendation Rates for Physical Therapy, Lifestyle Counseling and Pain Medications for Managing Knee Osteoarthritis in Ambulatory Care Settings. Cross-sectional Analysis of the National Ambulatory Care Survey (2007-2015). Arthritis Care Res (Hoboken). 2019:10.1002/acr.24064.

16 Listing of Tables, Listings and Figures

This section is to give precise details for each table, listing or figure to be produced.

Mock Table 1: MVPA at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 2: TOA-PT at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 3: TOA-PE at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 4: Steps/day at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 5: VAS Pain at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 6: KOOS at each study time point using the ITT sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 7: MVPA at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 8: TOA-PT at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 9: TOA-PE at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 10: Steps/day at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 11: VAS Pain at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			

Mock Table 12: KOOS at each study time point using the per-protocol sample

	Baseline	12-week	24-week
Brief Intervention			
Expanded Intervention			
Brief-Expanded			