Title: Pragmatic Cyclical Lower Extremity Exercise Trial for Parkinson's Disease
NCT#: NCT04000360
Document Date: January 29, 2024

The Cyclical Lower Extremity Exercise for Parkinson Disease II (CYCLE-II) Study
Principle Investigator: Jay Alberts, Ph.D.
Site Principal Investigator: Lee Dibble, Ph.D.

Statistical Analysis Plan: Assessment of Treatment Effect

Study Statistician: Peter B. Imrey, Ph.D.

Overview

CYCLE-II is a pragmatic two center, blinded examiner, 1:1 randomized clinical trial to assess whether one year of a home Peloton-cycle based aerobic exercise regimen alters Parkinson's disease progression from that of patients receiving usual care and usual physical activity. The trial is conducted at the Cleveland Clinic and the University of Utah, coordinated by the former. The CYCLE-II study design and basic analytic plan have been published (*Phys Ther* 2021 Nov; 101(11):pzab191, PMCID: PMC8632855). Multiple prespecified quantitative outcomes are hierarchically classified as primary inferential, secondary inferential, and descriptive exploratory outcomes, and assessed in randomly ordered on and off Parkinson's medication examinations at baseline, and again off medication at six and 12 months. Parkinson's disease progression is to be assessed in each group by the estimated annual rates of change of off-medication outcome measures, and the treatment groups are to be compared in intention-to-treat analyses of these estimates.

Study Outcomes

The primary outcome is the total score on the Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III: Motor Examination, an assessment of the motor signs of Parkinson's disease by an examiner masked to the participant's treatment group. Secondary inferential motor outcomes are:

i. Upper extremity dexterity: time to complete a iPad-based modification of the Nine Hole Peg Test;
ii. Lower extremity functional mobility
    a. time to complete an iPad-instrumented Time Up and Go Test;
    b. average turn velocity on an iPad-instrumented Time Up and Go Test;
iii. Ambulation gait speed
    a. average "comfortable speed" velocity on the Ten Minute Walk Test;
    b. average "fast speed" velocity on the Ten Minute Walk Test;

Secondary non-motor outcomes are:

i. Neural processing speed: number correct on the New Processing Speed Test;
ii. Episodic visual memory: Number of locations + number of symbols correctly recalled on the New Symbol Modalities Test;
iii. Working memory/set switching: the ratio of time to complete the Trail Making Test B to time to complete the Trail Making Test A.

Descriptive/exploratory motor outcomes are:

i. Cardiovascular fitness: Six Minute Walk Test distance;
ii. Postural stability:
    a. Balance Error Scoring System (BESS) double leg area
    b. Balance Error Scoring System (BESS) tandem area
    c. Balance Error Scoring System (BESS) double leg volume
    d. Balance Error Scoring System (BESS) tandem volume
iii. Total self-reported falls during the one-year observation period.

Descriptive/exploratory non-motor outcomes are:

i. Movement Disorder Society Unified Parkinson's Disease Rating Scale Part I: Non-Motor Aspects of
ii. Experiences of Daily Living (nM-EDL)
iii. Movement Disorder Society Unified Parkinson's Disease Rating Scale Part II: Motor Aspects of Experiences of Daily Living (M-EDL)

iv. Movement Disorder Society Unified Parkinson's Disease Rating Scale Part IV: Motor Complications.
The latter two scales are classified as non-motor rather than motor outcomes because they focus on patient-reported experience of functional and social impacts of the effects of motor symptoms rather than on the symptoms themselves.

Analysis Sets

The inferential and safety analysis set will consist of all randomized patients. The descriptive/exploratory analysis set will consist of all randomized patients who complete baseline procedures for initiating randomized treatment and follow-up, and for whom study follow-up is initiated. Descriptions of baseline variables will use all available data in the inferential and safety analysis set. Inferential and safety outcomes will be analyzed using the inferential and safety analysis set in accordance with the strict intention-to-treat principle, under which all randomized patients are included in analyses in accordance with their randomized treatment assignments, regardless of their post-randomization experience including treatment and dropout from treatment or study. Descriptive/exploratory outcomes will be analyzed using the descriptive/exploratory analysis set.

Missing Data

In inferential analyses we will presume that missing observations, for whatever reason they may be missing, are technically "missing at random," i.e., that the unobserved data come from the same distributions, conditional on other variables that have been ascertained in CYCLE II, as do the observed data. Under this assumption, we will multiply impute missing data using both baseline variables and outcome data and will include the multiply imputed outcome data in analyses. Specifically, missing data will be multiply imputed by sequential chained equations using fully conditional specification and the following sets of predictors:

i. baseline sociodemographic and other characteristics: age, sex, race, ethnicity, residential living situation (alone, with family member or friend, with significant other), employment status (full-time, part-time, retired due to Parkinson's disease related disability, otherwise retired), body mass index (BMI), hand dominance;
ii. baseline exercise habit: current regular exercise (yes/no), regular aerobic, resistance, and flexibility exercise (yes/no to each), regular home, gym, and outdoor exercise (yes/no to each), exercise intensity;
iii. Parkinson's disease-related factors: years since diagnosis, leva dopamine equivalent dose, primary Parkinson's disease symptom, patient-reported past year fall count (0, 1-2, 3-12, 13+);
iv. CYCLE-II characteristics: clinical site, randomized group;
v. concurrently measured primary and secondary (motor and non-motor) outcomes;
vi. the same outcome at other visits.

Variables will be imputed by forward order in time with the primary outcome last within each time. At least 25 imputations will be obtained, and trace plots for the primary outcomes will be checked for evident departures for convergence, with burn-in iterations increased until stationarity appears to have been achieved.

Baseline description

The distributions of baseline variables will be summarized, overall and by treatment group, by frequency distribution for dichotomous and other categorical factors, mean and standard deviation for reasonably symmetrically distributed quantitative variables, and Tukey five-number summaries (median, quartiles, extrema) for substantially skewed variables. We will also examine boxplots, histograms and kernel density estimators for continuous variables, and use standardized mean differences for continuous variables and categories of nominal and ordinal variables to describe balance of the randomized treatment groups.

Outcome description and distributional assessment

All prespecified outcomes are quantitative or potentially treatable as such depending on the degree to which their full scales occur in our study population. We will thus describe each initially as indicated above for quantitative baseline variables, overall and by treatment × time combinations, including visually using grouped boxplots and kernel density estimators. To assess compatibility with the assumptions of the planned analytic approach, for each outcome we will use restricted (equivalently, residual) maximum likelihood (REML) estimation to fit preliminary mixed models with fixed effects of randomized treatment, visit, and their interactions, and with unrestricted covariance matrices -- thus allowing within treatment and visit variability to change over time and accounting for arbitrary correlations between measurements of the outcome in the same patient at different visits. These models will not be used to estimate treatment effects, but only to assess the compatibility of their residuals with the requisite distributional assumptions about random noise to support validity of fixed effect estimates and tests from the planned inferential analyses.

Estimated skewness, excess kurtosis, quantile-quantile (qq) plots, boxplots, and residual vs. predicted (for the six treatment × time combinations) plots, particularly of Studentized and standardized residuals, will be used for this purpose. We will not use formal hypothesis tests of normality of the residual distributions, as the relevant issue is not whether the data literally arise from a normal distribution, which is virtually never the case, but rather whether the extent of departure from normality is enough to meaningfully distort effect estimates and associated hypothesis tests. Models based on Gaussian assumptions are robust to modest to moderate violations of assumptions, particularly to the distributional assumption when samples are reasonably large, in which case formal tests of normality may have power against small and often inconsequential departures from it.

For outcomes with distributional characteristics reasonably compatible with standard linear mixed model assumptions, we will employ prespecified linear mixed model analyses for the observations at baseline, six months, and twelve months, i.e., one year, as described below. For outcomes whose distributions, conditional on treatment and time, are substantially skewed and/or lighter or heavier tailed than Gaussian distributions (leptokurtotic or platykurtotic), we will consider applying the planned analysis to a Box-Cox and other reasonably straightforward data transformation of the outcome and backtransforming the results, and/or use of a generalized linear mixed model with log or other link function and/or a non-Gaussian exponential family or a beta error distribution. In exploring Box-Cox transformations, we will be generally guided by the independence pseudologlikelihood of the fitted fixed effects model saturated for treatment × visit with clinical site as an additive covariate, as a function of the Box-Cox transformation power parameter. If and when transformations are used, we will report modeling results on the transformed scale and also report additive treatment effects, which for nonlinear transformations will vary by baseline value of the outcome, for selected percentiles of such baseline values. Nonparametric approaches such as quantile regression or product multinomial-based mixed categorical data models for marginal means will be considered as needed for bounded integer-scored outcomes with limited ranges and/or floor or ceiling effects, which may not admit of straightforwardly interpretable and communicable normalizing transformations.

Once the scale of analysis and analytic approach have been chosen for outcomes incompatible with the prespecified default model, a covariance structure will be selected for each non-multinomial analysis by comparing Schwarz' Bayesian Information Criterion (BIC) after fitting the unconstrained fixed effects model above using the following six choices: time series homogeneous and heterogeneous order one autoregressive AR(1), order one antedependence, homogeneous and heterogeneous compound symmetry, and unrestricted. The homogeneous AR(1) will be considered the default if it yields close to the minimum BIC; otherwise, the structure with minimum BIC will be used. Residuals from the above saturated fixed effects model with the selected covariance structure will then be checked as above to confirm adequacy of the transformation or generalized linear mixed model.

Treatment effect estimation and testing

Assuming required assumptions are reasonably met, missing data will be multiply-imputed with 50 imputations, as described above, after any transformations of all outcomes to the respective scales determined for analysis. Treatment effects will then be estimated and tested by fitting mixed models, also as described above, on the original or transformed scale, but assuming equally spaced visits and linear time effects, to each imputed full data set. Estimates of the annual rate of change of the outcome variable within each treatment group, and of the treatment effect as their difference, with respective standard errors, will be obtained from each imputation and combined over all imputations using Rubin's standard method, yielding overall point estimates of the rates in each group and the difference in rates between the home Peloton-based exercise and usual care groups (i.e., the treatment effect), with respective overall 95% confidence intervals. Statistical significance of the treatment effects and of temporal change within each group will be tested using Wald statistics. Point and interval estimates on a transformed scale will be back transformed and reported and interpreted on the scales of the original measurements as differences between model-fitted ("predicted") values of the outcomes for participants beginning with the same baseline value of the outcome measure but assigned to the two groups, for a selection of baseline percentiles of the outcome measure (e.g., 5, 10, 25, 50, 75, 90, 95)..

To control the overall false positive rate of claiming disease-modifying therapeutic benefit from the home Peloton-based aerobic exercise program at 5% in the context of CYCLE II's nine inferential and other descriptive/exploratory outcomes, interpretation of significance testing will be in accordance with a prespecified outcome hierarchy. Specifically,

i. overall benefit will only be claimed if the test of the primary UPDRS III outcome is statistically significant in the direction of exercise superiority at a two-sided α=5%,
ii. benefit will only be claimed for a secondary motor inferential outcome if i) is satisfied and if the treatment effect is also statistically significant for that secondary motor outcome after Holm-Bonferroni adjustment for multiple testing of the five secondary motor outcomes; and
iii. benefit will only be claimed for a secondary non-motor inferential outcome if i) is satisfied and if the treatment effect is also statistically significant for that secondary non-motor outcome after Holm-Bonferroni adjustment for multiple testing of the three secondary non-motor outcomes;
iv. benefit will not be claimed to have been demonstrated based on estimated treatment effects on any variable other than the nine prespecified primary and secondary endpoints, and any associated p-values and confidence intervals for such other variables will be presented as descriptive only and potentially for guidance by designers of subsequent studies, and specifically noted to be unadjusted for multiple comparisons of over a dozen outcomes.

For inferential outcomes for which statistically significant treatment effects are found, in either direction, we will further explore the nature of the treatment effect. For the primary outcome and any descriptive/ exploratory outcomes that are also indices of multiple Parkinson's disease manifestations and that exhibit clinically relevant differences between treatment group, we will summarize and compare subscores associated with different symptom groups, e.g., speech and facial expression, hand movements, lower body movements, tremor. For any outcome with statistically significant treatment effect, we will examine the effect in a model saturated for treatment and time, within which we will separately estimate and test equality of the changes within each six-month period, i.e., test for nonlinearity over the one-year duration of treatment and compare treatments with respect to the changes within each six-month period and over the full year of treatment.

Sensitivity analyses

We will employ several sensitivity analyses to test the dependence of substantive results on our modeling assumptions. Our planned analyses rely on randomization to protect against confounding by baseline differences between the treatment groups. We will compare treatment groups on the baseline variables mentioned above and perform covariate-adjusted sensitivity analyses based on material baseline

differences between them. For this purpose, we will consider variables with standardized differences exceeding 15% and/or variance ratios exceeding 2.0 for inclusion in these models. Any adjustments for continuous covariates will include linear as well as nonlinear representations of these variables using restricted cubic spline terms. To assess sensitivity to the imputation models, we will also compare results of analyses using multiply-imputed data with results based on complete cases and available cases and, for any outcomes with borderline results, to those from alternative imputation models. To assess sensitivity to the choice of data transformations and distributional assumptions, we will compare results to those of multinomial based categorical data mixed models for marginal mean scores.

Exploratory/descriptive outcomes

In general, these variables will be described and modeled similarly to secondary outcomes, to potentially support future research. Hypothesis testing results, if obtained, will be interpreted descriptively and not to claim benefit or harm on the basis of CYCLE II data. The distributions of patient-reported falls within the course of the study year across patients within each treatment group will be assessed for compatibility with Poisson distribution. If, as is anticipated, overdispersion is observed, then the negative binomial or other models for overdispersed count data will be entertained.

Adverse events

Adverse events will be descriptively tabulated and compared by Common Terminology Criteria for Adverse Events (CTCAE) system/organ class and preferred term. Serious adverse events will be similarly tabulated and described in individual listings including calendar date and interval from randomization, severity grade (1-5), current status, and assessed relatedness to the exercise intervention. Serious adverse events related to the intervention are expected to be few, and formal hypothesis testing will not be performed due to anticipated low statistical power and the multiple testing context.

Effect modification/subgroup analyses

For those outcomes for which a statistically significant effect of the intervention is found, we will examine statistical interactions of the randomized aerobic exercise regimen with study site, age, sex, BMI, employment status, patient-reported intensity of regular pre-study exercise, primary Parkinson's disease symptoms, primary Parkinson's disease symptom, years since Parkinson's disease diagnosis, whether or not the patient reports any falls during the pre-study year, and baseline MDS-UPDRS III value. For potential categorical effect modifiers, we will focus on an overall Wald test rather than p-values for comparisons of specific subgroups. For continuous variables, we will focus on linear interactions. If a linear interaction is detected, however, we will also assess possible nonlinearity in models using restricted cubic spline effects, and we may also choose to reexpress and communicate the effect in terms of categories defined by ranges to better communicate clinical relevance.