Protocol Title: Comparative Effectiveness of an Exercise Intervention Delivered via Telerehabilitation and Conventional Mode of Delivery

Statistical Analysis Plan

PI: Dr. Deborah Backus, Ph.D.

Version: 6/17/2022

Sponsor: Patient Centered Outcomes Research Institute (PCORI)

NCT Number: NCT03468868

Data analysis plan

SA1: Compare the extent of change in walking and mobility outcomes, social and vocational participation, and quality of life. The primary hypothesis is a non-inferiority test of the changes in T25FWT between TET and FET. This will be tested using a linear model with site as the stratification factor included using an intention to treat approach (imputation for missing values are discussed below). The primary analysis will adjust for covariates age, gender and baseline EDSS. Further addition of covariates may be examined in sensitivity analyses based on any differences seen on examination of baseline variables between groups. Secondary endpoints include change in 6MWT and EDSS. In all of these endpoints we are looking at improvement over baseline values whereas many MS studies focus on worsening. We will examine the frequency of worsening as a safety comparison. We will also examine the primary endpoint of a 20% improvement in the T25FWT (yes or no) using logistic regression. Further, to characterize the changes over time, we will use linear mixed models to assess the slopes of change for FET versus TET groups. We also will examine the relationship between the dose of exercise measured by compliance and the change in the primary and secondary outcomes. These are exploratory outcomes and will be assessed with no corrections for multiple testing and are important mostly for descriptive analyses to see if the compliance measures indeed predict the benefits on the outcomes as this information is important for translation and information to people with MS.

SA2: Evaluate and compare the effectiveness between participants randomized to their preferred delivery mode (TET(Choice) or FET (Choice)) and those who are not randomized to their preferred delivery mode (TET or FET). For this aim we will use the same models as the primary but examining and interaction term for Choice and treatment and contrasts will be conducted comparing the four cells. We will assess if the effect of Choice is the same for both treatments. Logistic regression will be used to assess the interaction between choice and adherence and attendance or dropouts. We also will examine time to dropout and/or time until dropout or poor compliance (which will be defined prior to final analyses). We also are interested in examining the intensity or compliance of the intervention to the choice as well as impact on the primary endpoint as noted above.

SA3: Evaluate changes in self-efficacy and adoption of exercise between these groups (TET, FET, TET(Choice), FET(Choice)). Self-efficacy and adoption of exercise are discussed above and use the same models and approach. We will also utilize mediation analyses to understand if the changes induced by exercise are mediated via the changes in self-efficacy. Here we will use the approach by MacKinnon built on Judd and Kenny.[79-84] Analyses of the MSWS-12, MFIS, and the NeuroQOL will all be analyzed using regression models similar to the timed walk models above and will utilize the same contrasts for examining choice as well as treatment effects.

Data-source adequacy. All data will be collected specifically for this prospective research; thus, all needed covariates will be collected by the trained research staff. There will be a standard measurement manual and quality control procedures will be implemented. Each evaluator will be trained in data collection for the various outcome measures; new research personnel will also be trained as they join the study team. Data collection training will also be repeated annually. Demographic and disease related data will be collected and will allow us to examine cofounding variables, such as age (i.e. people with MS who are older may walk more slowly than those who are younger); type of assistive device (i.e. people with MS who use a walker may walk more slowly than those who do not use an assistive device); type of MS (i.e. people with MS who have RRMS may perform better than those with progressive forms of MS). Data for the study will be derived from several validated instruments and through participant assessment at enrollment and follow up PROs via iConquerMS. The use of these different data sources and modes of data collection provides some modest redundancy, while minimizing the amount of missing data, to assure that all outcomes are appropriately ascertained. Data will be entered in the HIPAA- compliant database maintained by iConquerMS with oversight by Dr. Backus. Sites will be responsible for maintaining all aspects of the database entered from the clinic and staff support from iConquerMS will be available to trouble shoot technical questions. Access to this data will be restricted to study personnel only. Individual site coordinators will each have a unique username and password to enter the participant information. Research data will be entered online through the secure system and source documents will be kept in a secure fashion. Staff from iConquerMS will monitor and track errors, completeness with routine reports and feedback to the sites. Data will be shared with the Statistical Center at UAB, who will provide analyses using SAS, SPSS, R and other programs, as necessary.

Data linkage plan. Each participant will be assigned a study ID number when randomized to the intervention group by the biostatistician. This will be the only identifier in all databases and systems used in this study. The Shepherd Center will maintain the master file. None of the investigators or staff involved in the assessments will have access to the identifier data, and thus will remain blinded to the assignment for the entire intervention phase.

Sensitivity analyses. Randomization should balance covariates across treatment groups, but this will be checked with frequency and means tables, and chi-square/Fisher exact and F-tests/t-tests respectively. If treatment groups show important imbalances among covariates, then they can be adjusted for using covariates in various types of regression models as we noted above under the primary analyses. As noted briefly in the primary analysis section, per protocol analyses will be used to support the conclusions of the intention to treat analyses. In the intention to treat analyses we must take into account dropouts. For anyone with at least one post randomization timed walk test, we will impute a model based final endpoint to assess change. Key sensitivity analyses will include adjustment for covariates: age, gender, sex and treatment preference at entry. Interaction terms will be used to identify potential impacts of key covariates. In addition, imputation for missing values will be conducted to assess the impact of these lost observations. We note again that rules are in place to deal with the time variables. For other variables (MSWS, MFIS and NeuroQOL, EDSS) will require various forms of imputation for sensitivity analyses. We will use a general approach discussed below for imputation of these secondary endpoints and use per protocol versus ITT for sensitivity analysis. For the major analyses, best case and worst case analyses will be done. Subgroup analyses will attempt to quantify the consistency of the findings amongst subgroups.

Management of missing data

a. **Methods to prevent and monitor missing data.**
   - To prevent the loss of CRO data by missed evaluation visits: 1) participants will receive a stipend to assist with travel for assessments; 2) participants will receive a written and electronic schedule of all visits; and 3) participants will receive a reminder from both the site RC and through automated messaging from the Study Portal. If a participant misses their clinic assessment visit, the visit will be rescheduled as soon as possible. All attempts will be made to capture the data within a specified time frame. iConquerMs will have automated processes in place and monitor for drop and completion rates that will be sent to Shepherd monthly.
   - CROs will be collected by the trained evaluator at each site and entered directly into an electronic database through the Study Portal. The site research coordinator (RC) will check the database for accuracy and completeness of data on the day completed to provide as close to real time entry as feasible. These strategies will allow more immediate identification of missing data and increase the likelihood of retrieving that data (MD-1). The Shepherd RC will review de-identified data via the Study Portal on a weekly basis for missing data and data integrity. iConquerMS will monitor and provide a weekly report to the Shepherd RC and Dr. Backus regarding dropouts and completion rates.
   - Demographic information and all PROs will be collected via the secure, study-specific partition of the Study Portal. The site RC will assist each participant with orientation to iConquerMS and introduce them to the PROs and process for completion of these. Data is downloadable into a database so that data analyses can be performed. To ensure completion of PROs at the specified times, participants will receive a reminder from the site RC as well as an automated reminder from Study Portal (MD-1). There will be processes in place with iConquerMS to provide for range check, logical consistency, and completion of data. Weekly reports will be generated with de-identified data and sent to the Shepherd RC and Dr. Backus to perform data checks. The Shepherd RC and Dr. Backus will communicate with sites immediately in the event of missing data.
   - Table 5 presents the Communication strategy that will also be employed to prevent missing data.

b. **Statistical methods to handle missing data.** While we will try to minimize missingness, it inevitably occurs. We do not expect large proportions of missingness because of our approach to visits and data

collection. We will use multiple imputation using SAS Proc MI and Proc MI analyze with 5 replicates per imputed dataset. We also will develop a propensity score within treatment group derived via a logistic regression analysis for prediction of the major missingness. The reason for using propensity scores is that if the proportion of missingness is low, covariate adjustment for multiple imputation can be dominated by the characteristics of those with the data and not those without. Item non-response for the scales will be dealt with per the instructions of the validated scales. The half rule will be used if there isn't a standard for the validated questionnaire unless the assumption of non-hierarchy of the question does not hold. If less than 3% of data are missing, the issue will be ignored and "complete-case analysis" will be used.[87] If, as expected, more than 3% of the data are missing (due to refusals or drop-outs, for example), then we will use multiple imputation for all analyses. The current expectation is that the "chained equations" (or, "fully conditional specification") method will be used in imputing the missing values. We expect to use a minimum of 100 sets of imputed values.

c. **Validated methods to deal with missing data that properly account for statistical uncertainty due to missingness.** We will use multiple imputation with Rubin's rules for estimating uncertainty, including imputation uncertainty. As noted above, we will use PROC MI and MI Analyze with five replicates per assessment so that variability will consist of both estimation uncertainty (within estimate variability) and variability between the imputed data sets. Diagnostic checks of the imputations will be used (see 6.e., below)
d. **Recording and reporting reasons for dropout and missing data and account for all patients in reports.** There will be several logs that will be maintained by the site RC, including site Training log, Adverse Events log, and Enrolled Participants log. Participants in both the facility-based (FET) or telerehab (TET) group will record their training each session in a log-book. The Pedometer data will also be monitored to determine adherence to the exercise program. Participants will record missed sessions and the reason for the miss, and if the sessions was made up during the corresponding week. For FET, the site RC will collect the log-books on a weekly basis; for TET, participants will mail the log-book to the site on a monthly basis. The site RC will enter the data into the central database. In addition, for TET, the coach will obtain information about missed sessions during the video-chat sessions and document these in the database. Any missed training or assessment sessions will be reported to the site RC. If any participant is dropped from the study, the data and reason will be documented in the site Enrollment Log. This information will be summarized annually for the site-specific IRBs as well as for the Shepherd (Primary site) IRB, the Advisory Board, and the DSMB, and will be reported in the Final Project report.
e. **Examining sensitivity of inferences to missing data methods and assumptions and incorporate them into interpretation.** Multiple imputation assumes that the data are Missing at Random (MAR). Sensitivity analysis here means checking that the results do not change qualitatively (the conclusion is the same) under other assumptions. We will use the following methods for our sensitivity analyses: (1) a weighting approach (the different imputed values are given different weights) as suggested in Carpenter, J.R., Kenward, M.G., and White, I.R. (2007),[85] The weights are chosen empirically and will mimic a Not Missing at Random assumption; (2) varying the imputed values by multiplying them by a fixed value; this allows one to emulate, say, a pattern-mixture model. Discussions of this appear in van Buuren, S. (2012)[86] and Raghunathan, T. (2016).[87] Note that van Buuren, among others[88] also discusses diagnostic checks on the imputations and these will be used also. If imputations fail the diagnostic checks, we will change our imputation model and re-impute (and re-run the diagnostic checks).

Addressing the heterogeneity of treatment effect (HTE)

a. **Goals of the HTE analyses.** Hypothesis driven analyses expand on the model above by including interaction terms (e.g., an indicator of RRMS v PMS which is then included in interaction terms by the other parameters in the model (to assess, e.g., treatment effectiveness)). Hypothesis generating analyses will use logistic regression where the outcome variable is whether the person got worse (v did not get worse) on, e.g., T25FWT; covariates will include which treatment they received and whether they were randomized to that treatment or chose

(Choice) as well as various demographics (e.g., gender, age) and indicators of medical status (including, e.g., RRMS v PMS, time since diagnosis. This now becomes a risk prediction model and we will follow the recommendations in Kent DM et al (2010).[89] In addition, one of our primary interests is to assess the impact of choice on the performance and benefits of exercise. This is aimed at a key aspect of heterogeneity of the treatment effects as it will have direct bearing on clinical recommendations for broader implementation. It is also important to note that should we fail to demonstrate non-inferiority, then the heterogeneity within and amongst subgroups may become very important for description of study results and recommendations. Finally, given the concern that health literacy may have an impact on outcomes, our study will use tools to assess and classify study participants with regard to health literacy in order to understand the relative contribution of health literacy to our outcomes. Health literacy outcomes will be managed in the same fashion as other important interaction terms (e.g., age, sex, relapse remitting vs. progressive MS) for statistical analyses.

b. **Analysis plan for HTE.** There are two specific hypothesis driven analyses: (1) RRMS v PMS; we hypothesize that those with RRMS will do better than those with PMS regardless of group; this will be analyzed by adding an interaction term to the model shown in SA2 above where the interaction is between treatment effect and type of MS; (2) a) symptom management medications: participants may decrease symptom management medications due to a decrease in their symptoms. This would be viewed as a positive outcome; b) participants who change their disease modifying therapy (DMT) will do worse than those who do not change their DMT; again, an interaction of this with treatment effectiveness will be included in the model discussed under SA2 above. We are also interested in examining heterogeneity amongst key subgroups such as age, gender, etc. These are a byproduct of our sensitivity analyses but will inform the heterogeneity questions.

c. **Basis for all HTE claims based on appropriate statistical contrasts among groups being compared, such as interaction tests or estimates of differences in treatment effect.** In addition to the interaction tests discussed above, we will also provide estimates of treatment effect for each such subgroup (e.g., RRMS, PMS) and the differences between the subgroups along with 95% confidence intervals (CIs). The addition of the effects plus CIs is to provide readers results that may be easier to interpret than the interaction terms in a regression – that is, the interaction tests and the treatment effect estimates are not competitive but supplement each other. Tests of whether the effects are different will also be made.

d. Plan to report all pre-specified analyses and, at a minimum, the number of post hoc analyses, including all subgroups and outcomes analyzed. The process of measuring HTE, and all results, will be reported, including how subgroups were measured/defined (e.g., what counts as a change of medication), the total number of subgroups and the total number of outcomes tested. For all procedures, the bootstrap will be used for internal validation and p-value (and CI) adjustments will be presented using Bonferroni-type procedures.