Adapting the Unified Protocol to Facilitate Activity in Older Adults

NCT04837521

Analytic Plan

Document Date: 2/25/2025

Analytic Plan:

For Aim 1 exit interviews will be analyzed using a rapid coding process designed to provide results quickly for healthcare providers. Coding will be conducted after each cohort and will inform revisions to the UP-5.

Aim 2. will be evaluated using descriptive statistics of treatment satisfaction, percent of sessions attended, retention rate, and therapist adherence. 80% retention and 80% therapist protocol adherence will be evidence of acceptability and feasibility. Change of 2.8 on mean PSFS score will be considered significant.

We will calculate differences pre- and post- intervention on measures of emotional distress and daily activity using general linear models. These calculations, which will include 95% confidence intervals, will provide information about the possible range of effect sizes and whether there is "signal" in improving these outcomes. Based on prior UP studies, we anticipate 80% power to detect at least a moderate effectsize of .60.

Analyses completed after data collection:

All analyses were conducted using R software.

Feasibility and Acceptability

Treatment feasibility was assessed via attrition and session attendance rates. We planned to compare attrition using a generalized linear model with a logit link function (McCullagh, 1984) and session attendance rates using a Z-test of two proportions. Within the TD group, descriptive statistics were also computed for participant homework completion and therapist adherence ratings to the protocol.

Treatment acceptability was assessed via CSQ total scores, and treatment recommendation ratings from the CEQ. CSQ scores were analyzed by a two-way (treatment-by-time) mixed effects linear regression model utilizing the *lme4* package(Bates et al., 2015). Random intercepts were included for participants, and the model was estimated using restricted maximum likelihood. Time was specified as a factor variable with levels for the 1-week and 1-month follow-up visits. Post-hoc contrasts evaluated the significance of differences in estimated marginal means between interventions at each timepoint. The recommendation ratings from the CEQ were analyzed in a two-way (treatment-by-time) mixed effects ordinal logistic regression model with random intercepts per participant, utilizing the *ordinal* package(Christensen, 2023).

Primary Clinical Endpoint

PSFS scores were analyzed using mixed effects linear growth models with time, treatment condition, and their interaction as predictors(Curran et al., 2010). Random effects modeled variation in intercepts, slopes over time, and their covariance. Time was specified as a continuous variable, centered on the date of the final observation, and scaled so that 1-unit represented the median duration between the first treatment session and the 1-month follow-up visit (10.5 weeks). The SG group served as the reference level for the treatment factor. Models were estimated in a Bayesian framework using default priors in the *brms* package(Bürkner, 2017). Four Markov Chain Monte Carlo chains were run with a warmup period of 1,000 iterations and a total of 4,000 iterations per chain, resulting in a posterior distribution of 12,000 candidate models. Convergence was assessed using the potential scale reduction factor (Rhat), with values close to 1 indicating good performance. Density plots of predictions from 100 random draws of the posterior distribution were overlayed against the empirical distribution of the response variable to visually inspect how well our model captured underlying patterns in the data.

To identify the optimal functional form of change over time, we compared nested unconditional models with different specifications of the time trend using leave-one-out cross-validation(Vehtari et al., 2017). Given only three observations per participant, we limited our

comparisons of temporal trends to a simple linear growth model and a piecewise growth model with two linear segments joined by a knot at the 1-week follow-up visit. The model with higher expected log pointwise predictive density (ELPD) indicated greater fit to the data. The number of highly influential observations (i.e., Pareto *k* values > 0.7) was also taken into consideration to assess the reliability of the predictive accuracy estimate. If the difference in ELPD was trivial compared to the standard error of the difference, or the number of influential observations was considerably larger in the piecewise trend, then the simple linear trend was retained for the final conditional growth model to evaluate treatment-by-time interactions.

A Bayesian analytic approach was desirable in the context of this small sample given that the results offer a more nuanced understanding of uncertainty in estimated treatment effects compared to frequentist methods. Our analyses permitted two separate types of hypothesis tests: (1) whether a consistent effect greater or less than zero existed, and (2) whether an effect was significant in size(Makowski et al., 2019). We evaluated the existence of effects using probability of direction (*pd*) values (i.e., the proportion of the posterior distribution of estimates either positive or negative in value, whichever figure is larger). *pd*-values less than 5% were considered strong evidence an effect existed. To evaluate the significance of the difference in outcomes between treatments, we defined a region of practical equivalence (ROPE) using $\pm 0.1*\sigma_{\text{PSFS baseline}}$ to span a range of negligible effect sizes(Kruschke & Liddell, 2018). If a treatment difference existed (per the *pd*), *and* less than 5% of the full posterior distribution overlapped with the ROPE, the effect was considered significant. For the main effects of time (i.e., within-person change scores[1]), we stipulated that 2.8 points constitutes a minimum clinically important difference (MCID) on the PSFS (Mathis et al., 2019) and calculated the probability that a participant's improvement exceeded this threshold as a benchmark for clinical significance. Standardized effect sizes for between-treatment and within-person effects were calculated by dividing point estimates from the growth model by $\sigma_{\text{PSFS baseline}}$(Feingold, 2009).

Secondary Endpoints – Self-Report Forms

[1] Point estimates for total within-person change (ΔPSFS) reflected the difference in predicted values at Time=0 (i.e., expected PSFS scores at 1-month follow-up) and Time=-1 (i.e., expected PSFS scores 10.5 weeks earlier).

Additional clinical self-report outcomes (three PROMIS scales, BEAQ, CFS, and SMQ) were analyzed similarly to the primary endpoint using Bayesian mixed effects growth models with random intercepts, slopes, and their covariance. First, the optimal functional form of the temporal trend (i.e., linear or piecewise trajectory) was determined for each outcome variable by comparing nested unconditional models as previously described. Then, outcome variables with the same form of trajectory were simultaneously predicted in a multivariate mixed effects model, with level-1 residuals allowed to covary across outcomes, and separate (i.e., uncorrelated) blocks for level-2 random effects[2]. For each outcome variable we assessed the existence of effects as well as the clinical significance of the between-treatment difference in change scores using the ROPE-full method, as defined by $\pm 0.1 * \sigma_{\text{Outcome baseline}}$.

Secondary Endpoints – Physical Activity Measures

SPPB gait speed was analyzed in a two-way (treatment-by-time) mixed effects linear regression model with random intercepts, estimated via restricted maximum likelihood. The ANOVA-like specification was selected instead of the growth model as there were only two timepoints per participant.

In contrast, the large number of repeated observations for weekly average daily steps allowed for modeling potentially complex nonlinear trends. Step counts were analyzed using a multiple-group mixed effects model with a latent basis growth specification, estimated via maximum likelihood with the *nlme* package(Pinheiro et al., 2023). Initially, random effects for intercepts and total growth were included, but the model could not converge; thus, only random intercepts were retained. In the latent basis growth model, the optimal shape of the temporal trajectory is estimated directly from the data, similar to factor loadings in confirmatory factor analysis(Ram & Grimm, 2007). Two basis vectors, A_0 and A_1, each with seven elements (corresponding to the number of measurement occasions per person), described the pattern of change over time separately within SG and TD groups. A_0 was a fixed vector of 1s, while A_1 contained two fixed elements for model identification – the first

[2] Across outcomes, we identified a single outlier value on the CFS that was non-credible (based on the subsequent and succeeding scores for that participant) and replaced that data point using multiple imputation.

coefficient set to 0 and the last set to 1. The remaining five elements of A_1 were freely estimated, with each basis coefficient representing the proportion of total change occurring between the first and last measurement occasions. PerRam and Grimm (2009), the multiple-group growth model can be expressed as follows:

$$Avg\ Daily\ Steps[Visit]_n = g_{0nc} * A_{0c}[Visit] + g_{1c} * A_{1c}[Visit] + e[Visit]_{nc}$$

The subscript *c* denotes the respective treatment group (SG or TD) for individual *n*, g_0 represents the expected starting value of weekly average daily steps, and g_1 represents the expected change in average daily steps from pre-treatment to 1-month follow-up .