Multicultural Healthy Diet to Reduce Cognitive Decline & Alzheimer's Disease Risk
Statistical Analysis Plan 05/05/2018
(version date 9/7/2023)

Introduction

This document includes a brief description of the study design, objectives, and outcomes, and then describes the analysis plan for the study outcomes at the end of the study. There will be no interim analyses of efficacy outcomes because this is a low risk study of a pilot clinical trial of short duration and important end-points such as sustainability and timing of effects may not be observed if the trial is stopped early, we will not perform interim efficacy analyses. The study team, along with the DSMB, will monitor for safety and quality of conduct throughout the study.

Design Overview

This study is a two-arm randomized controlled clinical trial with an 18-month intervention period. Assessment visits cover a 27-month period (run-in, baseline, months 9 & 18 and 27) at the Aging Research Center (Van Etten Building) at the Albert Einstein College of Medicine (Einstein).

Objectives

To investigate whether consuming the Multicultural Healthy Diet (MHD) can be adapted for a multicultural middle aged (40-65 yr) middle income cohort in Bronx, New York and whether it improves cognition compared to consuming a usual diet.

Outcomes

Primary Outcomes

The primary outcome is a global composite cognition score assessed at 9 months post baseline. The primary outcome will be calculated as a z-score composite measure of three ambulatory cognitive tasks relating to three cognitive domains: visuospatial memory (Grid Memory task), processing speed (Symbol Search task) and a dimension of short-term associative memory binding shown to be sensitive to the early Alzheimer disease neuropathology (Color Shapes task). The global composite score for each measurement burst (baseline, 9months, etc) will be an average of three domain-specific Z-scores based on the within person average response on that domain from up to 35 sessions within a given measurement burst.

Secondary Outcomes

The secondary outcomes relate study arm to dietary intake based on self-report and biomarker measures of intake at 9 months post baseline. In addition, our secondary outcomes will consist of examining dietary effects on each of the three individual tasks (Symbol Search, Grid Memory,

Color Shapes) to determine whether observed intervention effects are attributable to specific cognitive domains.

Randomization

Randomization occurs after the baseline visit and body measurements when participants are judged to be eligible to proceed to study activities. Stratified permuted block randomization with variable block sizes will be used to randomize participants to one of the two study arms in a 1:1 ratio. Randomization will be stratified on sex, age (<53 yr, 53-65 yr), and education level (no more than high school or equivalent, more than high school), for a total of 8 stratification factors.

Statistical Analysis Plan

Analysis of the Primary Outcome

The primary outcome is a global composite cognition score assessed at 9 months post baseline. The primary outcome will be calculated as a Z-score composite measure of three ambulatory cognitive tasks, each administered up to 35 times, during each measurement burst: Symbol Search, Grid Memory, and Color Shapes. For a given cognitive task ("domain"), an outcome score is calculated for each of up to 35 sessions which are administered over a 1-week measurement burst, approximately 5 sessions per day (administered throughout the day in 5 target ~2 hour windows). For each Burst, and each of the 3 domains, the assigned Z-score is derived from their burst-average response for that domain, and then the three domain burst-average Z-scores are averaged for that Burst to create the primary composite outcome. Note that the Symbol Search and Dot Memory Z-scores **must be multiplied by -1** before creating the composite score because lower scores are better for these tasks. Thus, an increase in the composite score represents better cognitive performance.

At least 50% (18 out of 35) of the scheduled sessions must be completed for each task in order for a given measurement burst to be considered valid. These 35 scores may take longer than 7 days to accumulate but must be taken without 2-week or larger gap between sessions to be included in the analysis. Each burst has a recorded burst start date; sessions recorded prior to this start date are considered part of the pre-burst training and will be discarded. The first day of the Burst 2 measurement should be inside the visit window (i.e., at least one session no earlier than 3 months and no later than 6 months than the 9-month randomization anniversary) for the Burst 2 measurement to be valid and included in the primary analysis. The domain Z-score is computed using the mean and SD for an individual session, estimated from the baseline (Burst 1) distribution and the measurement burst for an individual i at Burst j is the average task scores across the multiple sessions (up to 35) within that burst. The domain Z-score with fewer than 18 valid sessions will be treated as missing and multiply imputed for the primary analysis, which will be an intent-to-treat analysis. The details are as follows.

The mean and SD are calculated for each domain separately from the baseline measurement (Burst 1). For a given domain, let Y_{i1j}^{d} be the Burst 1 outcome score from domain d, d=1,2,3;

person *i*, i=1,…n; and session *k*=1,…,m_{i1}, where m_{i1} is the number of successfully completed sessions at baseline for person i (which may vary by domain d), then:

$\mu^d = \Sigma_i \bar{Y}^d_{i1.}/n$; $\bar{Y}^d_{i1.} = \Sigma_j (Y^d_{i1k})/ m_{i1}$; SD^d = sqrt ($\Sigma_i(\bar{Y}^d_{i1.}–\bar{Y}^d_{.1.})$^2/(n-1)).

Then for a given measurement burst j=1,2,3,4: where j=1 represents baseline and j=2-4 represents months 9, 18, and 27, respectively; the domain Z-score for subject i and Burst j is denoted by Z^d_{ij} is calculated from Y^d_{ijk} as:

$$Z^d_{ij} = (Y^d_{ij.} – \mu^d)/SD^d.$$

The three primary tasks (variable names) are:

- Z^1 =Symbol search (SYMBOL_SEARCH.median.RT.accurate_trials)
- Z^2 = Grid Memory (DOT_MEMORY.sum.error.distance.overall)
- Z^3 = Color Shapes (COLOR_SHAPES.CorRec.rate)

The global assessment score for person i at Burst j is calculated as the average of the three domain average Z score, with Z^1 and Z^2 multiplied by negative 1 so increase is improvement:

$$Z_{ij} = (-Z^1_{ij} - Z^2_{ij} + Z^3_{ij})/3.$$

The primary analysis will be a between-arm comparison of the change from baseline in global composite cognition score, measured at Burst 2 (approximately month 9) as estimated from the ANCOVA regression model, which also includes adjustment for the stratification factors $Strat_i$ =(sex, age, years of education):

$$Z_{i2} = \beta_0 + \beta_1 Z_{i1} + \beta_2 Arm_i + \beta_3^T Strat_i + e_{i.}$$

In this model, age and years of education are treated as continuous. The stratification variables are added to assure proper confidence interval width, since randomization included stratification involving these variables (Kahan and Morris 2011). The coefficient β_2 quantifies the difference between treatment arms at Burst 2, adjusted for the baseline Burst 1 composite score measure and stratification variables, and will be fit using linear regression.

A supportive analysis of the primary endpoint will be the adjusted analysis examining the treatment effect on the change from baseline at Burst 2 for each of the three individual task domain scores $\bar{Y}^d_{i2}$.

Analysis of Secondary Outcomes

We will consider the treatment effect on the individual tasks (symbol search, color shapes and grid memory) in a longitudinal session-level model with the post-baseline session level outcome Y^d_{ijk} (j>1) regressed on the baseline average $Y^d_{i1.}$, treatment arm, time since baseline $time_{ij}$, time of day t_{ijk}(continuous covariate 0-24, centered at group average), day of week (i.e work day or day off), season (4 level factor variable: winter, spring, summer, fall), practice effects $g(s_{ijk})$ as a function of s_{ijk}(session number 1-35 as a continuous covariate), and, fatigue (visual

analogue). A b-spline with 3 degrees of freedom will be used to model a non-linear practice effect and the likelihood ratio test will be used to assess whether non-linear terms for the practice effect are needed. Splines will also be used to fit a circadian rhythm curve for the effect of the time of day of the session $f(t_{ijk})$. For the continuous tasks (grid memory and symbol search), linear mixed effects model will be fit with a random intercept to account for the correlation of repeated measures within an individual (Model 1). Random intercept variances are allowed to vary by burst. We will also consider whether there is evidence for between person variation in the change over time by including random slope terms for the time effects (Model 2). The treatment effect in Models 1 and 2 will be modeled with a main effect an interaction with the continuous time variable. Thus, the following linear mixed effects models will be fit for each continuous domain:

Model 1: Longitudinal linear mixed effects model for repeated measures across burst and session.

$Y^d_{ijk} = \beta_0 + \beta_1 Y^d_{i1.} + \beta_2 Arm_i + \beta_3 time_{ij} + \beta_4 Arm*time_{ij} + \sum_l \beta_{5l} f(t_{ijk}) + \beta_6 Day_{ijk} + \beta'_5 Season_{ij} + \sum_m \beta_{8m} g(s_{ijk}) + \beta_9 Fatigue_{ijk} + \tilde{\beta}^T_{10} Strat_i + b_{ij} + e_{ijk}$,

Model 2: Longitudinal linear linear mixed effects model for repeated measures across burst and session, including random effects for both slope and intercept.

$Y^d_{ijk} = \beta_0 + \beta_1 Y^d_{i.0} + \beta_2 Arm_i + \beta_3 time_{ij} + \beta_4 Arm*time_{ij} + \sum_l \beta_{5l} f(t_{ijk}) + \beta_6 Day_{ijk} + \beta'_7 Season_{ij} + \sum_m \beta_{8m} g(s_{ijk}) + \beta_9 Fatigue_{ijk} + \tilde{\beta}^T_{10} Strat_i + b3\ time_{ij} + b_4 Arm*time_{ij} + b_{ij} + e_{ijk}$,

where b_{ij} is the person level effect (random intercept term) for participant *i* and Burst *j,* and e_{ijk} is the residual error term assumed to be mean 0, independent across time and individuals and approximately normally distributed. In this model, response Y^d_{ijk} is modeled as a function of the overall baseline level and the term β_4 summarizes the potential treatment effect on the change over time. Models will be further adjusted for age, gender, APOE4 allele, race, and ethnicity in supportive analyses. The difference in time and underlying cognitive function within a burst is assumed negligible, thus the time and season variables in these models will be determined by the first day of the Burst.

The Color Shapes task has only 7 possible values, so a similar generalized estimating equation (GEE) approach will be taken using the Poisson mean model and robust sandwich variance for the number of incorrect responses (0-6).

Our secondary outcomes also include between arm comparisons of MHD-related dietary intake based on self-report and biomarker measures of intake at 9 months post baseline. Outcome models of between-arm differences in dietary intakes indicative of the MHD as measured by nutritional biomarkers, and self-report at Burst 2 will be conducted in a similar manner to the models specified above for the primary cognitive outcomes: namely, ANCOVA models that adjust for the baseline level, Arm and stratification factors will be fit. Tocopherols and carotenoids plasma biomarkers will be additionally adjusted for cholesterol variables per Gross et al (2003). Longitudinal models considering ASA24 data across the 4 bursts will be fit using similar mixed effects models, simplified to analyze only the Burst level average for the ASA24 dietary intakes (no repeated measures within a burst). Analyses will generally be complete case

analyses for the dietary variables. Energy adjustment will be considered where appropriate. Sensitivity analyses with additional adjustment variables associated with missing, as well as exploratory analyses that examine potentially differences in intervention effect by gender, race, ethnicity and age, will also be considered.

The plasma biomarkers include: total folate, alpha and gamma tocopherol, carotenoids (alpha carotene, beta carotene, zeaxanthin+lutein, beta cryptoxanthin, and vitamin B12), lycopene, and fatty acid profile. Serum sodium is an additional secondary outcome, indicative of hydration.

Self-reported dietary outcomes indicative of the MHD diet include the DII score, which is derived from the self-reported data as both unadjusted and unadjusted for energy, water, and the following factors chosen because of hypothesized pro- or anti-inflammatory effects and because they were readily available summaries from the ASA24: total calories (KCAL), protein (PROT), total fat (TFAT), solid fats (SOLID_FATS), carbohydrates (CARB), total folate from food (FF), folate from non-food sources (non-FF), vitamin B12, carotenoids: retinol (RET), beta carotene (BCAR), alpha carotene (ACAR), cryptoxanthin (CRYP), lycopene (LYCO), Lutein+zeaxanthin (LZ); alpha tocopherol (ATOC), fatty acids (EPA, DPA, DHA), total fruits (F_TOTAL), green leafy vegetables (V_DRKGR), beans and peas (V_LEGUMES), whole grains (G_WHOLE), cured meats (PF_CUREDMEAT), poultry (PF_POULT), seafood high in n-3 fatty acids (PF_SEAFD_HI), nuts and seeds (PF_NUTSD), cheese (D_CHEESE), and added sugars (ADD_SUGARS).

Exploratory Analyses

Exploratory outcomes include the treatment effect on the two exploratory tasks (Shopping List and Rotation Span), which will be evaluated with similar longitudinal models developed for the three domains that make up the primary composite Z score (namely Models 1 and 2). Shopping List will be modeled with a linear mixed effects model. Due to the concern of departure in the Rotation Span scores from normality (a slight truncation), this domain will be modeled with generalized estimating equation with robust variance estimate.

For any identified important effects of the diet on cognition, we will explore which of the dietary components were mediating that effect. In particular, for the dietary components that seem to be the most different between arms at Burst 2, we will explore whether the Burst 2 measure (or the change from baseline at Burst 2) are associated with a change in the cognitive outcomes at Burst 2 using models similar to that for the primary analysis of the intervention effect. Supportive analyses may also consider formal mediation analyses or the approaches discussed in Freedman et al. 2010 (i.e., both principal components and Howe's method) as ways to efficiently combine the biomarker and self-reported measures of intake in an analysis of their association with cognitive measures. Exploratory dietary effects also include the treatment effect on HbA1c and self-reported vitamin D and choline. The treatment effect on the inflammatory markers IL6 and CRP will also be considered.

For any significant differences by treatment arm, we will examine whether there was evidence of a differential intervention effect across APOE4 groups. Exploratory analyses that examine

potentially differences in identified intervention effects by gender, race, ethnicity and age, will also be considered.

Exploratory analysis will also examine whether there was weight loss at Burst 2 relative to Burst 1 in a complete case analysis:

$$W_{i2} = \beta_0 + \beta_1 W_{i1} + \beta_2 Arm_i + \beta_3^T Strat_i + e_{i.}$$

and if significant changes in weight are identified, analyses will be considered to identify whether any identified intervention effects were potentially mediated by weight loss.

Handling of Missing Data

Analyses of the primary outcome will be done on an intent-to-treat basis; we will assume the data are missing at random (MAR) and use multiple imputation to impute missing measures. Rubin's rules will be used to calculate the variance of the model parameters for each analysis of interest. For secondary outcomes, approaches to address missing data will also assume MAR and approaches to address missing data will generally employ complete case adjustment methods (Little et al., 2022). For missingness rates >10%, sensitivity analyses may be performed using varying assumptions regarding the missing data model to examine whether study conclusions are robust to these differing assumptions. Further details are provided below.

Imputation Models for the Ambulatory Cognitive Scores

For the primary analysis, we will use the multiple imputation chained equation approach (MICE) to impute the necessary missing data to allow an ITT analysis (Azur et al. 2011). Burst measures that had fewer than 18 valid sessions for a given domain average (i.e. for $\bar{Y}_{ij.}^{d}$ for d=1,2,3; j=1,2) or a Burst 2 measure that started outside the Burst 2 window will be treated as missing and imputed.

A sequence of chained equations for MICE will be considered using: all variables in the primary analysis model (i.e., Arm, gender (0=female, 1=male), age at Burst 1, years of education, Burst 1 composite Z-score, and Burst 2 composite Z-score), run-in average variables for each of the tasks 1-3 in the composite outcome ($\bar{Y}_{i0.}^{d}$),(averaging first 10 or fewer for each), MOCA score, indicator for in-person vs telephone MOCA, and interaction between these two MOCA variables, domain-specific Zscores based on the average of the available data from up to the first 10 sessions at Burst 1 for each of the three tasks 1-3 in the composite outcome, and the Burst 2 domain-specific Zscores using up to the first 10 sessions for each of the three tasks. Additional models may be considered as sensitivity analyses.

The imputation will be conducted in R using the `mice` package defaults for each variable type (e.g., predictive mean matching for the missing continuous variables (Van Buuren 2018). The number of multiple imputations will be greater than 25 (large enough to be stable, in terms of the intervention effect estimates and p-values).

Missing data approach for the secondary outcomes

For randomized trials, under the MAR assumption and when only the outcome variable is missing, complete case analyses that additionally adjust for all the variables related to the missingness imputation will be consistent and can be the most efficient analysis (Little et al, 2022). Thus for the secondary endpoints, the adjustment approach will be considered. For the repeated measures models of the cognitive outcomes, the proposed adjusted analysis models are as described above (Model 1 and Model 2). For the dietary biomarker and self-reported ASA-24 outcomes, adjustment for baseline demographics (age at Burst 1, gender, race, ethnicity, years of education, MOCA score, indicator for in-person vs telephone MOCA, and interaction between the two MOCA variables, and BMI. Energy adjustment will be considered as appropriate for the dietary component. In particular, parsimonious outcome models that include adjustment of variables when found to be associated with missingness will be considered. Analysis of tocopherols and carotenoids plasma biomarkers will be additionally adjusted for cholesterol variables per Gross et al (2003).

References

Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011 Mar;20(1):40-9. Doi: 10.1002/mpr.329. PMID: 21499542; PMCID: PMC3074241.

Gross M, Yu X, Hannan P, Prouty C, Jacobs Jr DR. Lipid standardization of serum fat-soluble antioxidant concentrations: the YALTA study. The American journal of clinical nutrition. 2003 Feb 1;77(2):458-66.

Little RJ, Carpenter JR, Lee KJ. A comparison of three popular methods for handling missing data: complete-case analysis, inverse probability weighting, and multiple imputation. Sociological Methods & Research. 2022 Aug 5: 1-31. doi: 10.177/00491241221113873.

Van Buuren, S. (2018). Flexible imputation of missing data. CRC press.