Protocol Title:

The Canadian Therapeutic Platform Trial for Multidomain Interventions to Prevent Dementia: Brain Health Support Program Intervention

NCT: NCT05347966

Document Title: Statistical Analysis Plan (SAP)

Document Date: November 12, 2024

The Canadian Therapeutic Platform Trial for Multidomain Interventions to Prevent Dementia: Brain Health Support Program Intervention (CAN-THUMBS UP or CTU: BHSP)

Statistical Analysis Plan (SAP)

Version 2.0 12Nov2024

Executive Lead:	**Howard Chertkow, MD**
Co- Principal Investigators:	**Howard Feldman, MD** **Haakon B. Nygaard, MD, PhD** **Sylvie Belleville, PhD** **Manuel Montero-Odasso, MD, PhD**
Contributing Investigators (alphabetical):	**Nicole Anderson, PhD** **John R. Best, PhD** **Paul Brewster, PhD** **Diane Jacobs, PhD** **Andrew Lim, PhD**

Summary of Changes

Modification	**Rationale**	**SAP Section(s) Impacted**
Revision made to the statistical analysis package to be used	New statistician on project (John Best) uses R as opposed to SAS. Also, revision allows for all analyses and graphics be created in R rather than a combination or SAS and R (as described previously)	8.1 Overview of Analysis Approach 8.2 Analysis of Primary Outcome Software
Revision to primary analytic sample and methods to analyze that sample	Previously, the proposed models would not include the entire intention-to-treat sample due to missingness at baseline in certain outcome measures. Description of multiple imputation prior to analysis added to allow for the entire intention-to-treat sample be incorporated in all primary analyses	4.0 Samples of Interest 8.1 Overview of Analysis Approach
Clarified that regression models will be fit to multiply imputed data where relevant	Clarification added to specify that the multiply imputed data would be used to evaluate changes in secondary and exploratory outcomes (in addition to primary outcome)	8.3 Analysis of Secondary Outcomes 8.4 Exploratory Analyses
Revision made to the analysis plan for computing the cognitive global and sub-domain composite scores from the Neuropsychological Test Battery	Clarified that for purposes of the composite score analysis, the z scores for the timed tasks (Oral Trail Making Test (Part A and Part B) and Color-Word Interference Test (Word Reading and Inhibition) are reverse coded so that a higher score is better	8.4.4 Cognition 8.4.4.1 Global Cognition 8.4.4.2 Subdomains
Updated the measures of interest in the Cogniciti Brain Health Assessment battery	Clarified that the primary outcome measure of interest is the Overall Brain Health Assessment Score. One representative score is calculated from each of the four BHA subtests, and an overall score is calculated from these subtest scores	8.4.4.4 Cogniciti Brain Health Assessment
Included a description of the methodology and analysis plan for calculating the age adjusted polygenic hazard scores (PHS)	Description added based on report provided by the CTU genomics team and additional planned exploratory analyses	8.4.7 Saliva Sample and PHS

Table of Contents

1 Introduction 5

2 Study Aims 5

3 Power and Sample Size Determination 5

4 Populations of Interest 6

5 Enrollment and Participant Flow 6

6 Study Flow CONSORT Diagram 6

7 Evaluation of Demographics and Baseline Characteristics 7

8 Statistical Analysis 7

8.1 Overview of Statistical Approach 7
8.2 Analysis of Primary Outcome 8
8.3 Analyses of Secondary Outcomes 8
8.4 Exploratory Analyses 9

Study Title: The Canadian Therapeutic Platform Trial for Multidomain Interventions to Prevent Dementia: Brain Health Support Program Intervention

Study Acronym: CAN-THUMBS UP or CTU: BHSP

1 Introduction

This initial phase (Phase A) of CTU is a prospective 12-month multi-center longitudinal single-group intervention study to evaluate a web-based educational Brain Health Support Program (BHSP), called Brain Health PRO (BHPro), focused on dementia literacy, self-efficacy, and modifiable lifestyle risk factors. Participants are individuals who are either cognitively unimpaired or have MCI with increased risk related to lifestyle risk factors. At the conclusion of the BHPro intervention, participants will continue in the Platform Trial Cohort (PTC) with the opportunity to consent and enroll in further multidomain intervention trials.

2 Study Aims

A. Our primary aim is to evaluate within-person change from baseline in dementia literacy following participation in BHPro.

B. To evaluate within-person change following participation in BHPro, including change in:
   i. Self-efficacy and attitudes towards dementia and its screening
   ii. Individuals' modifiable risk factors
   iii. Cognition
   iv. Physical activity and sleep quality as measured by actigraphy and EEG wearable devices

C. To evaluate BHPro in relationship to:
   i. Levels of engagement (i.e. percent of chapters completed), ratings of satisfaction, and dropout rates from the intervention
   ii. Levels of engagement in the program as a function of individual characteristics and risk profile
   iii. Levels of engagement as a moderator of within person change in outcomes.
   iv. Association between change in modifiable risk factors and change in cognition

D. To develop a successful comprehensive national recruitment plan to fully enroll the PTC with engagement stakeholder groups including participants, citizen advisors, and community partners

3 Power and Sample Size Determination

Sample size/power calculation was done based on effect size for one-group pre-/post-design. According to Becker, the effect size [1] for such a design should be defined as, assuming the standard deviation (SD) at pre- and post-test are equal,

$$\text{ES}=\frac{\text{mean change}}{\text{SD baseline}}=\frac{\text{mean change}}{\text{SD change}/\sqrt{2(1-r)}}$$

where *r* is correlation between pre- and post-scores. Note that

$$\text{var}(\text{mean change})=2\text{var}(\text{baseline})(1-\text{r})$$

Sample size can then be determined based on the following test statistic

$$T = \frac{\text{mean change}}{\text{SD change}/\sqrt{n}} = \frac{\text{ES}}{\sqrt{2(1-r)/n}}$$

i.e., the corresponding sample size formula for 80% power at 5% significance level to detect ES is given by

$$n = \frac{(1.96 + 0.84)^2}{(\text{mean change/SD change})^2} + 2 = \frac{(1.96 + 0.84)^2}{\text{ES}^2/[2(1-r)]} + 2$$

Note that plus 2 $(\approx 2 \times 1.96^2)$is to approximate Z-test with sample size for t-test[2].

Therefore, the sample sizes for detecting small and moderate magnitudes of ES [3] at 2-sided 5% significance level with 80% power are given below for values of correlation r

ES	*r*	*n*
0.2 (small)	0.2	316
	0.3	277
	0.4	238
	0.5	198
0.5 (medium)	0.2	53
	0.3	46
	0.4	40
	0.5	34

4 Samples of Interest

a. **Enrolled Sample**: Includes all participants who consented to screening.
b. **Intention To Treat (ITT) Sample**: Includes all participants who (1) meet screening criteria (2) and completed at least 1 assessment at baseline.
c. **Modified Intention to Treat (mITT) Sample**: Includes all eligible participants who (1) have completed at least one assessment for the outcome of interest and (2) registered for BHPro.
d. **Completers Sample (CS)**: Includes all eligible participants who have completed the baseline and 12-month assessment for the primary outcome.

5 Enrollment and Participant Flow

5.1 Accrual of the Study

Accrual will be summarized by study site, providing the overall rate of accrual over calendar time. The actual rate and projected rate of accrual will be displayed in a graph (the projected rate assumes uniform accrual over time).

6 Study Flow CONSORT Diagram

A description of participant flow will be provided (see appendix 1). The diagram will describe study status from screening to the end of the study. At each stage, reasons for persons not moving forward will be summarized by frequency and category. The diagram will include the following information:

- Number and reasons for those who screen failed
- Number of participants who completed the Month 6 assessments
- Number of participants who completed the Month 12 assessments
- Number and reasons for participants who discontinued study before the Month 12 assessments

7 Evaluation of Demographics and Baseline Characteristics

Tables will summarize the ITT Sample at baseline, including demographics and medical history. Tables will also summarize the mITT and Completers Samples. Descriptive statistics will be presented as N, mean, standard deviation, minimum, 25th quartile, median, 75th quartile and maximum for continuous variables and frequency tables (row, column percentages) for categorical variables. Corresponding box-whisker plots and histogram plots will also be provided.

8 Statistical Analysis

8.1 Overview of Statistical Approach

In our one-group pretest-posttest design the major interest is within-person change in the primary and secondary outcomes over time (response variables) in the presence of one or more explanatory variables that are categorical or continuous. The data will be analyzed using a mixed model for repeated measures approach (MMRM)[4]. Example R code using the 'nlme' and 'emmeans' statistical packages is given by:

```
mdl = temp_data |>
                gls(Score ~ factor(Time) + Age + Sex + Education,
                    na.action = na.omit, data = _,
                    correlation = nlme::corSymm(form=~Time |Site/Identifiers),
                    weights = varIdent(form = ~1|Time))
emmeans(mdl, ~Time, contr = "pairwise")
```

This approach will treat sites and participants as random effects, while observational time and their characteristics such as sex, age and education as fixed effects. Potential effect modifications will be examined using this MMRM approach with variable selection based on Akaike/Bayesian information criterion. The MMRM approach effectively handles missing data, with the assumption that data are missing at random. In this approach, information on missing observations is recovered from the observed outcomes via the within-patient correlation structure. In contrast to the flawed method of last observation carried forward[5], which uses only one data point, a MMRM analysis uses all the available data to compensate for the data missing on a particular patient.

However, because there is missing data at baseline for certain outcome measures—including the primary outcome because of a technical issue in the measure administration—the MMRM approach on the observed data will not provide inferences on the intention-to-treat (ITT) sample. This is because the MMRM excludes all individuals without at least one measurement of the outcome of interest. To address this issue, multiple imputation of missing data will be untaken prior to fitting the MMRMs. Forty data sets will be imputed by chained equations (that is, fully conditional specification) using the R package 'mice' [6] using predictive mean matching (continuous variables), polytomous logistic regression (unordered factors), and logistic

regression (binary factors) following 40 iterations of a Gibbs sampler for each imputed data set. The imputation model will include all time-varying primary, secondary and exploratory outcomes, as well as age, sex, education, racial/cultural background, cognitive diagnostic group, and study site. The sequence of missing data imputation will be ordered from the variable with the smallest percent missing to the variable with the most percent missing. A seed will be specified so that the imputation solution is reproducible.

The quality of the imputed values will be assured by evaluating the distribution of imputed values and examination of trace plots for proper mixing and absence of spikes in the iterations. Subsequent analyses will be conducted on each of the imputed data sets with estimates then being pooled using Rubin's rule [7] and degrees of freedom calculated using the Barnard-Rubin adjustment [8].

The primary analysis sample will be the ITT sample, i.e., participants with at least one outcome measure will be included in the analyses. As noted above, this will be achieved by fitting MMRMs on the multiply imputed data. As sensitivity analyses, the MMRMs will be fit to the observed data from the mITT and Completers samples. Sex-stratification will be used in analyses as appropriate.

8.2 Analysis of Primary Outcome

The primary outcome is dementia literacy following participation in the study, as measured by the **Alzheimer's Disease Knowledge Scale (ADKS).** The ADKS is designed to assess knowledge about Alzheimer's Disease (AD) among laypeople, patients, caregivers, and professionals. This self-report questionnaire contains 30 true/false items. The total score is quantitative and ranges from 0-30, with a higher score indicating better knowledge about AD.

An MMRM will be used to evaluate the primary outcome measure of change over time on the ADKS. This model will examine the effect of time (baseline, 6-month, 12-month), adjusted for baseline characteristics including age, sex, race, cognitive status and education.

Inference on change in ADKS from baseline will be conducted via least squared mean estimates as implemented in the R 'emmeans' package [9].

8.3 Analyses of Secondary Outcomes

8.3.1 Change in self-efficacy following participation in the study, as measured by the **General Self-Efficacy Scale (GSE).**

The GSE measures perceived competence in dealing with a range of stressful or challenging situations. This self-report questionnaire contains 10-items, each rated on a 4-point scale (not true at all, hardly true, moderately true, exactly true). The total score is quantitative and ranges from 10-40, with a higher score indicating more self-efficacy.

An MMRM will be used to evaluate performance over time on the GSE. This model will examine the effect of time (baseline, 6-month, 12-month), adjusted for age, sex, race, cognitive status and education. The MMRM will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the 'completers' sample.

8.3.2 To evaluate engagement in BHPro, as measured by:

a. Percentage of chapter completion

Descriptive statistics will be used to describe fundamental measures of engagement with BHPro (percentage of chapter completion), overall and by age, sex, gender, educational level, diagnostic group, and severity of cognitive impairment (MCI vs CU) will be tracked.

b. Dropout rates and survival analysis with reasons for early discontinuation

Real time Kaplan-Meier curves will be used to evaluate time to study dropout, overall and by CTU participating site as well as by major diagnostic and demographic groups.

8.3.4 To evaluate user satisfaction of BHPro, as measured by evaluation of:

a. Usability as measured by the **System Usability Scale (SUS)**

The **SUS** is an 11-item questionnaire with 5-point Likert scale. The total raw score ranges from 0-43 with a higher score indicating greater usability and satisfaction with Brain Health Pro.

Descriptive statistics will be used to describe measures of user satisfaction, usability and acceptance of BHPro overall and by age, sex, gender, educational level, diagnostic group, and severity of cognitive impairment (MCI vs CU).

b. Acceptance as measured by the **Technology Acceptance Model Questionnaire (TAMQ)**)

The **TAMQ** is a 20-item questionnaire (adapted for BHPro) with 7-point scale. The total score ranges between 0-120, with a higher score indicating greater acceptance and satisfaction with BHPro.

Descriptive statistics will be used to describe measures of user satisfaction, usability and acceptance of BHPro overall and by age, sex, gender, educational level, diagnostic group, and severity of cognitive impairment (MCI vs CU).

8.4 Exploratory Analyses

8.4.1 Change in attitudes toward dementia following participation in the study as measured by **Sections B and D** of the **Perceptions Regarding Investigational Screening for Memory in Primary Care (PRISM-PC)**.

The **PRISM-PC** questionnaire captures participants' acceptance, perceived harms, and

perceived benefits of dementia screening. Sections B and D will be used to assess attitudes toward dementia and dementia screening. Section B consists of 8-items and measures acceptance of screening. Section D consists of 29 items and measures benefits and harms of dementia screening including 4 constructs (benefits, stigma, suffering, and negative impact of screening on independence). All items are rated on a 5-point scale (strongly agree, agree, I don't know, disagree, strongly disagree). The total score is quantitative and ranges from 36-180. Positive statements were reverse coded so that a total higher score indicates greater acceptance and higher perceived benefit of dementia screening.

An MMRM will be used to evaluate performance over time on the PRISM-PC. This model will examine the effect of time (baseline, 6-month, 12-month) adjusted for baseline characteristics including age, sex, race, cognitive status and education. The MMRM will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the 'completers' sample.

8.4.2 Effect of dose (number of chapters completed in BHPro)

Fall-off in use of the program as time on study increases will be monitored, and Cox regression with time-varying covariates will be used to investigate the extent to which a decrease in engagement with BHPro provides an early warning for risk of study dropout.

The effect of dose will be assessed using regression models with restricted cubic splines to allow for non-linearity in the association of the number of chapters completed within the BHPro program (a measure of dose) and dropout status.

8.4.3 Change in modifiable risk factors following participation in the study, as measured by **BHPro Lifestyle Risk Questionnaires**.

Each lifestyle questionnaires includes a quantitative total score.

Physical Activity: The International Physical Activity Questionnaire-Short Form is a brief 7-item self-report measure that captures types of and intensity of physical activity and sitting time. The total score is calculated by summing the accrued number of points assigned per question. The total score ranges between 0-9, with a higher score indicating more physical activity.

Cognitive Engagement: A 6-item questionnaire assessing types and duration of cognitively engaging activities. The total score is calculated by summing the accrued number of points assigned per question. The total score ranges between 0 and 36, with a higher score indicating more cognitively stimulating activities.

Diet: Shorter 11-item version of the Eating Pattern Self-Assessment Questionnaire developed by CCNA Team 5. The total score is calculated by calculating the accrued number of points assigned per question. The total score ranges between 0 and 22, with a higher score indicating more healthy eating habits.

Sleep: Brief 8-item questionnaire developed by BHPro sleep module content leads that assesses sleep duration, sleep patterns, and difficulties and daytime fatigue over the past 3 months. The total score is calculated by summing the accrued number of points

assigned per question. The total score ranges between 1 and 16, with a higher score indicating healthier sleep.

Social & Psychological Health: An 8-item questionnaire comprising 1 item on loneliness, 1 item on ageism, 1 item on subjective age, 1 item on essentialist beliefs of aging and 4 items on depression, anxiety, stress, and social support from STOP-D. The total score is calculated by summing the accrued number of points assigned per question. If a user selects' "0/Never" for example, they receive 0 points for that question. The total score ranges between -1 and 50, with a higher score indicating poorer social and psychological health (a lower score indicates better health).

Vascular Health: 12-item self-report questionnaire developed by the BHPro Vascular Health team based on the American Heart Association's Life's Simple Seven checklist of the main risk factors for heart disease and stroke. The total score is calculated by summing the accrued number of points assigned per question. The total score ranges between -4 and 8, with a higher score indicating better vascular health.

Vision & Hearing: 10-item questionnaire developed by BHSP Vision and Hearing Module content leads that assesses perceived visual and auditory ability and actions taken to address potential vision/hearing difficulties. The total score is calculated by summing the accrued number of points assigned per question. If a user selects "Yes" for example, they receive 2 points for that question. The total score ranges between 0 and 20 points, with a higher score indicating poorer vision and hearing (a lower score indicates better health). Response: higher scores = more functional impairment as a proxy for risk.

An MMRM will be used to evaluate performance over time on the BHPro Lifestyle Risk Questionnaires. This model will examine the effect of time (baseline, 6-month, 12-month) adjusted for baseline characteristics including age, sex, race, cognitive status and education. The MMRM will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the 'completers' sample.

8.4.4 Cognition

8.4.4.1 Change in **global cognition** will be assessed using a composite outcome score computed from the following measures from Neuropsychological Test Battery:

- Craft Story – Immediate Recall – Total Story Units Recalled - Verbatim
- Craft Story – Delayed Recall – Total Story Units Recalled - Verbatim
- ADAS-COG Word Recall – Total Immediate Recall (trial 1 + trial 2 + trial 3)
- ADAS-COG Word Recall – Delay (total correct)
- Brief Visuospatial Memory Test-Revised (BVMT-R)- Total Recall (trial 1 + trial 2 + trial 3)
- BVMT-R Delayed Recall
- Oral Symbol Digit Modalities Test – Total Raw Score
- Oral Trail Making Test (Part A) – Time To Completion
- Oral Trail Making Test (Part B-A)
- DKEFS Category Fluency – Total Correct (animals only)
- DKEFS Color-Word Interference Test (CWIT) – Condition 2 Word Reading – Time to Complete

- DKEFS CWIT – (Condition 3 Inhibition-Condition 2 Word Reading)

Z-scores for each participant on each measure will be calculated using the overall mean and SD of all participants at baseline as reference. Scores from each individual test will be converted to z-scores that typically range from -3 to 3, with higher scores reflecting better performance, and averaged to form a composite. For purposes of this analysis, the z scores for the timed tasks (Oral Trail Making Test (Part A and Part B) and Color-Word Interference Test (Word Reading and Inhibition)) are reverse coded. If more than 25% of the component measures are missing, the global cognitive composite score will be missing.

An MMRM will be used to evaluate change in the global cognitive outcome from baseline to Month 12, adjusted for baseline characteristics including age, sex, race, cognitive status and education. The MMRM will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the 'completers' sample. Inference on the change will be conducted through least-square means estimates.

Additional exploratory analysis will be conducted by analyzing the data obtained by first calculating the domain-specific outcomes (see section 8.4.4.2 below), and then aggregating the domain-specific outcomes to calculate the global cognitive outcome score. Methods similar to the above will be used for the analysis.

8.4.4.2 Change in the **subdomains of memory, processing speed, and executive functions** will be assessed using domain-specific composites.

The memory subdomain is comprised of the following tests:

- Craft Story – Immediate Recall – Total Story Units Recalled - Verbatim
- Craft Story – Delayed Recall – Total Story Units Recalled - Verbatim
- ADAS-COG Word Recall – Total Immediate Recall (trial 1 + trial 2 + trial 3)
- ADAS-COG Word Recall – Delay (total correct)
- BVMT-R Total Recall (trial 1 + trial 2 + trial 3)
- BVMT-R Delayed Recall

The processing speed subdomain is comprised of the following tests:

- Oral SDMT – Total Raw Score
- Oral TMT (Part A) – Time To Completion
- DKEFS CWIT – Condition 1 Color Naming – Time to Complete

The Executive Functions subdomain is comprised of the following tests:

- Oral Trail Making Test (Part B – Part A)
- DKEFS Category Fluency – Total Correct (animals only)
- DKEFS CWIT- (Condition 3 Inhibition- Condition 2 Word Reading)

Z-scores for each participant on each measure will be calculated using the overall mean and SD of all participants at baseline as reference.

To calculate each domain-specific cognitive composite, scores from each individual test within the subdomain will be converted to z-scores that typically

range from -3 to 3, with higher scores reflecting better performance, and averaged to form a composite. For purposes of this analysis, the z scores for the timed tasks (Oral Trail Making Test (Part A and Part B) and Color-Word Interference Test (Word Reading and Inhibition)) are reverse coded. Composite scores are computed when 50% or more of the component test scores are available.

An MMRM will be used to evaluate change from baseline to Month 12 in each of the domain-specific composites z-scores, adjusted for baseline characteristics including age, sex, race, cognitive status and education. The MMRM will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the 'completers' sample. The change will be estimated using the least-square mean contrast.

8.4.4.3 Change in performance based on **"Burst" Cognitive Testing (MyCogHealth Mobile App)** that addresses: 1) Within-person change in asymptote, daily variation and within-burst practice effects of "burst" cognitive testing; 2) Between- and within-person factors including self-rated emotional reactivity and its association with variation in cognitive performance over the BHPro intervention

Mixed effects modelling for burst cognitive testing will be used to disambiguate sources of performance variance within-day, across days, and across burst-to-burst timescales and to include within-person factors (i.e., stress, mood) that influence variation in performance. The data will be evaluated within-person cognitive change, cognitive variation, and differential retest/learning effects at the individual level as potential digital biomarkers.

We will investigate the association of participant demographic characteristics, cognitive status, and BHPro compliance with performance and change in MyCogHealth performance parameters. Exploratory analyses will examine concordance between the MyCogHealth outcome measures and the NTB, Cogniciti BHA, and BHPro Lifestyle Risk factors.

8.4.4.4 Change in performance based on online self-administered cognitive testing with the **Cogniciti Brain Health Assessment (BHA)**.

To evaluate the feasibility of online cognitive assessment, we will examine the number of completed self-administered assessments using the Cogniciti BHA at Baseline, Month 6, and Month 12. We will examine the association of participant demographic characteristics, clinical diagnosis, and BHPro compliance with performance and change from baseline on the Overall Brain Health Assessment Score and each of the subtests of the Cogniciti BHA.

The primary outcome measure for the Cogniciti BHA is the Overall Brain Health Assessment Score. One representative score is calculated from each of the four BHA subtests, and an overall score is calculated from these subtest scores. 1. Spatial Working Memory: Number of responses. This score is the number of

responses (i.e., the number of clicks) required to find all six pairs of shapes, summed across the three trials.
2. Face-Name Association: Recognition accuracy. This is an overall accuracy score for the 24 test items, expressed as a percent. It is calculated as the number of correct hits (i.e., “yes” responses to intact pairs) plus the number of correct rejections (i.e., “no” responses to recombined pairs) divided by the total number of items.
3. Stroop Interference: Incongruent reaction time. The score on this test is the median reaction time, expressed in milliseconds, on the incongruent trials. Of the 30 incongruent trials, reaction times were included only for those trials on which the correct response was given.
4. Letter-Number Alternation: Time to completion. This score is the total time required to complete the alternating letter-number sequence, expressed in seconds.
5. Overall BHA Score. Each of the 4 subtest scores are converted to z-scores based on normative data from Troyer et al 2014[10]. The overall score is calculated as the mean of the four z- scores and is converted to a percentile to aid interpretation.

Exploratory analyses will also examine concordance between the Cogniciti BHA outcome measures and examiner-administered neuropsychological assessments.

8.4.4.5 Change in modifiable risk factors and relationship to change in cognition.

Multivariate mixed linear regression models will be used to investigate the dynamic relationship between baseline level and longitudinal change in health behaviours (as assessed by the lifestyle risk questionnaires) and performance on the Cogniciti BHA and MyCogHealth cognitive tasks. For those models that investigate change in performance on MyCogHealth, session-level ratings of participant mood, sleep quality, and exercise will be investigated as mediators of the association between changes in health behaviours and cognition. All analyses will be adjusted to age, sex, and education. Regression models will be fit to the ITT sample (that is, the same multiply imputed dataset described in section 8.1), to the mITT sample, and to the ‘completers’ sample. A sensitivity analysis will stratify the cohort by cognitive status.

8.4.5 National recruitment success

8.4.5.1 Enrollment rates per month and year across regions with centralized and local recruitment strategies

Descriptive statistics will be used to describe enrollment rates including enrollment per month based on age, sex, gender, ethnicity, and region (rural/vs urban), and what means of recruitment (existing cohort/clinic, postcards, ads, earned media) were used. We will also analyze utility of postcards (how many signed up for study as a function of postcards sent) and response rate based on age, sex, gender, ethnicity, rural/vs urban.

8.4.5.2 Screen fail rate and reasons

Descriptive statistics will be used to describe total screen fail rate and reasons and by age, sex, gender, ethnicity and region.

8.4.5.3 Projected enrollment rates vs actual enrollment rates

Descriptive statistics will be used to compare projected enrollment rates versus actual enrollment rates. Information will be presented in the form of a projected enrollment graph and actual enrollment graph showing levels of enrollment as various recruitment strategies were deployed.

8.4.6 Change in levels of physical activity and sleep quality, as measured by:

a. Actigraphy

Ambulatory actigraphy will be assessed as a measure of total daily physical activity and sleep fragmentation and whether their modification lessens the risk of cognitive decline. The degree of sleep fragmentation will be quantified for each participant using the metric kRA, a novel metric that was developed based on modelling actigraphic data as a series of state transitions between rest and activity, and quantifying the transition probabilities[11]. The triaxial actigraphy data, sampled at 50-60Hz, are used to compute activity counts within each 15 second epoch using a published approach, and arousal is indicated by a non-zero activity count within an epoch, after a sustained period of zero counts. Higher kRA indicates greater sleep fragmentation. In addition, we will quantify total daily physical activity.

b. EEG wearable devices

Sleep stages will be extracted from the overnight EEG data using the manufacturer's algorithms, and will be compared pre- and post-BHSP to quantify effects of the BHSP intervention on sleep architecture. We will compute standard metrics of sleep architecture including total sleep time, sleep efficiency, and wake time after sleep onset, as well as proportion of time spent in each of wake, stage N1, stage N2, stage N3, and REM sleep.

c. Cardiopulmonary sensors (ANNE devices)

For participants in the optional cardiopulmonary sensor sub-study, we will quantify the apnea hypopnea index (AHI), oxygen desaturation index (ODI), hypoxemia burden, and time with oxygen saturation below 90% (O2<90). A participant will be considered to have sleep apnea if their AHI is greater than or equal to 15. In addition, we will quantify 24-hour resting heart rate, and average heart rate, which are important physiological consequences of sleep apnea.

Baseline levels of physical activity, sleep fragmentation, slow wave sleep, and indices of sleep apnea and resting heart rate will be compared with those found after receiving 12 months of engagement with the BHSP using the MMRM approach, adjusting for covariates such as age, sex and education. In addition, an interaction effect of these

measures and time will be assessed while burst cognitive testing scores are used as the dependent variable and physical activity, sleep fragmentation, slow wave, sleep, and indices of sleep apnea and resting heart rate are categorized according to their quantiles.

8.4.7 Saliva sample collection, to characterize the distribution of age adjusted polygenic hazard scores (PHS) within the BHSP study group.

The PHS model developed by Desikan [12] will be used to derive age-specific risk estimates for the development of Alzheimer's Disease within the CTU BHSP cohort.

Methodology:
To accomplish this exploratory outcome, the Desikan PHS model will be employed, which is based on the Cox proportional hazards assumption, to calculate risk scores for Alzheimer's disease (AD) using individual genetic profiles from the CTU BHSP cohort. The logarithm of hazards associated with 31 SNPs identified by Desikan, along with APOE variant status, is incorporated into the risk score calculation using the standard equation for Polygenic Risk Score (PRS) calculation (Equation 1). This process generates a risk score for each individual based on their genetic profile.

Analyses:
To validate the predictive value of our risk scores, we will initially examine the distribution of scores within three phenotypic classes: Cognitively Unimpaired (CU), Subjective Cognitive Impairment (SCI), and Mild Cognitive Impairment (MCI). We will also assess the distribution of risk scores between Affected (MCI) and Control (CU, SCI) groups. We will also evaluate the predictive value of the risk scores by incorporating them into a Generalized Linear Model, aiming to distinguish the Affected condition from the Control group.

To enhance the predictive value of the PRS, several strategies are planned for implementation. Specifically, incorporating APOE e4 homozygote status as a separate factor in the model may enhance predictive accuracy, given the substantial impact of this risk genotype on Alzheimer's disease. Additionally, restricting the model to utilize only the top and bottom percentiles of the PRS may improve performance, as these percentiles exhibit more pronounced differences. Furthermore, we intend to generate age-adjusted PRS, similar to the approach taken by Desikan. While Desikan used age of onset for PRS adjustment, which is unavailable in our dataset, we plan to attempt age data imputation on our samples with binned PRS values.

Additional exploratory analyses will examine the effect of PRS on other BHSP outcome variables (e.g. BHPro engagement, lifestyle risk domains, cognition, sleep).

Descriptive statistics will also be used to describe the distribution of age adjusted polygenic hazard scores.

Software

R version 4.4 (r-project.org) will be used for data analysis.

References

1. Becker BJ. Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology.* 1988;41(2):257-278.
2. Guenther WC. Sample size formulas for normal theory t tests. *The American Statistician.* 1981;35(4):243-244.
3. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155-159.
4. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal.* 2008;42(4):303-319.
5. Molnar FJ, Hutton B, Fergusson D. Does analysis using “last observation carried forward” introduce bias in dementia research? *Cmaj.* 2008;179(8):751-753.
6. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *Journal of statistical software.* 2011;45:1-67.
7. Rubin DB. *Multiple imputation for nonresponse in surveys.* New York, NY: John Wiley & Sons; 1987.
8. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika.* 1999;86(4):948-955.
9. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.1. 2024.
10. Troyer AK, Rowe G, Murphy KJ, Levine B, Leach L, Hasher L. Development and evaluation of a self-administered on-line test of memory and attention for middle-aged and older adults. *Frontiers in aging neuroscience.* 2014;6:335.
11. Lim AS, Yu L, Costa M, et al. Quantification of the fragmentation of rest-activity patterns in elderly individuals using a state transition analysis. *Sleep.* 2011;34(11):1569-1581.
12. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. *PLoS Med.* 2017;14(3):e1002258.

Appendix 1: BHSP Consort Diagram