Lifestyle Interventions and Independence for Elders

The LIFE Study

NIH U01AG022376

Analysis Protocol

1.0. Design

The LIFE Study is a **multicenter single-blind RCT** involving physical activity vs. a successful aging program, with an average follow-up of 2.7 years (range 1.9-3.5 years or 23-42 months) in 1635 non-disabled, community-dwelling persons age 70-89 years randomized across 9 Field Centers. Since randomization at the Florida site was stratified by the two sub-sites, analyses that control for factors used to stratify randomization will consider these as two different sites, thus resulting in 9 sites.

The **inclusion criteria** are (1) age 70 to 89 years; (2) summary score <10 on the Short Physical Performance Battery (SPPB);[90] (3) sedentary lifestyle; (4) ability to complete the 400 m walk test within 15 minutes without sitting or the help of another person, or the use of a walker; and (5) willingness to be randomized to either intervention group. The **exclusion criteria** reflect conditions that may interfere with the conduct of the physical activity program. LIFE recruited 67.2% women and 21% racial minorities.

1.1. Primary Hypothesis and Primary Outcome

Primary hypothesis: Compared with a successful aging (SA) health education program, a **long-term** structured physical activity (PA) program reduces the risk of major mobility disability, defined as incapacity to walk 400 m.

After a thorough evaluation of possible alternative approaches, LIFE has selected as the primary outcome for the full-scale trial **time to the onset of major mobility disability (MMD)**. This outcome is adjudicated as described below. The objective component of the **major mobility disability** outcome is defined as the inability to complete a 400 m walk test within 15 minutes without sitting or the help of another person. Individuals who complete the walk in more than 15 minutes have an extremely slow pace (<0.45 m/sec), which would make their walking capacity of little utility in daily life.[91] Selecting a higher cut point, such as 30 or 60 minutes, makes the objective assessment impractical and does not add to the clinical significance of the outcome. Major mobility disability is assessed every six months by staff who are **blinded** to the intervention assignment.

1.2. Secondary and Tertiary Hypotheses

Secondary hypotheses:

Compared with random assignment to a successful aging program, random assignment to a long-term structured physical activity program:

1. Improves pre-specified measures of cognitive function based on the Digit Symbol Test (DSST) and the Hopkins Verbal Learning Test (HVLT);
2. Reduces the risk of serious fall injuries;
3. Reduces the risk of persistent major mobility disability;
4. Reduces the risk of the combined outcome of major mobility disability or death;
5. Reduces the risk of disability in activities of daily living (ADLs); and
6. Is cost-effective.

Tertiary hypotheses: Compared with SA, the PA program

1. Reduces the risk of the combined outcome of all-cause mild cognitive impairment or dementia (MCI/D);
2. Improves performance on a composite measure of cognitive function;
3. Improves physical performance within subgroups defined on the basis of ethnicity/race, gender and baseline performance;
4. Improves sleep-wake disturbances and leads to a lower incidence of sleep-wake disturbances;
5. Reduces dyspnea, improves ventilatory capacity, and reduces respiratory muscle weakness;
6. Reduces the risk of hospital admissions for exacerbation of airway disease

(asthma, bronchitis, emphysema, &/or COPD) or pneumonia; and

7. Reduces the risk of combined cardiovascular (CVD) events including:
   - Myocardial infarction (MI),
   - Angina requiring hospitalization,
   - Any stroke (ischemic or hemorrhagic),
   - Transitory ischemic attack (TIA) requiring hospitalization,
   - Hospitalization for carotid artery disease,
   - Hospitalization for congestive heart failure (CHF),
   - Hospitalization for peripheral artery disease (PAD) or outpatient revascularization for PAD,
   - Ruptured abdominal aortic aneurism (AAA), and
   - CVD death;
8. Improves cognitive function within subgroups defined on the basis of baseline level of global cognitive function defined by baseline 3MS (<90 versus $\geq$ 90);
9. Improves lower extremity blood flow, as assessed by ABI;
10. Reduces the overall proportion of 400 m walk failures over time, as determined by an average intervention effect across repeated measurements.

1.2. Analysis Plans

For each aim, we present the analysis plan for testing the effect of the intervention on the outcome variable. General considerations for handling missing data are presented at the end of this section.

The primary aim is to assess the long-term relative effect of randomization to a physical activity intervention on time until major mobility disability (defined as inability to walk 400 meters (through objective or adjudicated evidence).

The primary study hypothesis of LIFE will be tested based on a two-tailed significance level of 0.05. In this analysis, the "intention to treat" approach will be used in which participants are grouped according to randomization assignment.

The main comparison of intervention groups with respect to the distribution of time until the first post-randomization occurrence of a primary outcome will be based on survival analysis. To compare intervention arms, we will use a likelihood ratio test from a Cox proportional hazards regression model, stratified by Field Center and gender. The proportional hazard assumption will be examined and alternative models may be used as sensitivity analyses, if necessary. Kaplan-Meier plots will be used to present the survival curves by intervention.

Failure time will be measured from the time of randomization. Follow-up time for participants who have ever had MMD will be calculated as the time in years from randomization to their first occurrence of MMD. Follow-up time for participants who have never had MMD but have had at least one follow-up visit determining no MMD will be calculated as the time in years from randomization to their last known determination of no MMD, regardless of study status after that determination. For example, if a participant with no prior MMD has a determination of no MMD at their two year visit, then dies or withdraws consent three months later, follow-up time will be set at two years. MMD is assessed every six months, so censoring/follow-up times cluster around these six month time points. Because measurement of MMD at 6-month intervals creates intervals that are fairly short and regular in length throughout follow-up, we have followed the recommendations of Leung, Elashoff and Afifi (Censoring issues in survival analysis. Annu Rev Public Health 1997; 18:83-104), who state that “if …the intervals are about 3 to 6 months wide, then we have no reason to complicate the analysis by considering interval censoring.” In the primary analysis, participants who have not had any MMD assessments

will be assigned one hour of follow-up time, since we know that they were able to do the 400m walk at baseline. Additional details on the MMD adjudication process are included in the LIFE MOP (28.4.1 and 28.4.2).

We will examine for differential effects of intervention within the following pre-specified baseline subgroups:

1. ethnicity/race (non-Hispanic white vs. other),
2. gender,
3. baseline physical performance (SPPB < 8 vs. SPPB ≥ 8),
4. age groups (70-79 vs. 80+ years),
5. baseline gait speed based on the 4m walk (<0.8 m/sec vs. ≥ 0.8 m/sec),
6. baseline history of CVD, and
7. baseline history of diabetes.

Effects of the intervention on the primary outcome will be calculated within subgroups and the intervention-by-subgroup interaction will be tested. This test will be implemented by adding the indicator for subgroup and the intervention-by-subgroup interaction to the primary analysis model. The significance of the intervention-by-subgroup interaction will be tested using a likelihood ratio test comparing (1) a model with intervention and subgroup to (2) a model with intervention, subgroup, and intervention-by-subgroup interaction.

This general approach of adding intervention-by-subgroup interactions to pre-specified analysis models that test for the main effect of the intervention will be used to test for differential intervention effects on other outcomes, too. In those cases, the analysis techniques (e.g. ANCOVA rather than survival analysis) may differ but the interactions within these models will be used to evaluate subgroup effects.

Secondary aim #1 will be to assess the relative effect of randomization to the intervention on cognitive function as measured by the DSST and HVLT instruments.

Cognitive data will be collected at the baseline and 2-year visits. The primary analysis will be to test the intervention effect on each outcome, separately. Analysis of the intervention effect will be carried out using analysis of covariance with variables in the model representing field center, gender, the baseline level of the outcome, and the intervention assignment. The effect of the intervention on DSST (overall score) and HVLT (mean of immediate and delayed recall subscales) will be based on a two-tailed significance level of 0.05 and will use the “intent to treat” approach.

Pre-specified subgroups for the DSST and HVLT instruments include:

1. Baseline 3MSE < 90 versus 3MSE ≥ 90,
2. Baseline physical performance (SPPB < 8 vs. SPPB ≥ 8)
3. Gender, and
4. Age at baseline (70-79 vs 80+ years).

Testing for subgroup effects on these cognitive endpoints at 24-months will be performed by using the F-test for the interaction between the intervention and subgroup. To compare effect sizes between these two tests, results will also be portrayed from parallel analyses based on z-scores formed by dividing the difference between individual scores and the cohort-wide average at baseline by the cohort-wide standard deviation at baseline. Supporting analyses of the individual HVLT subscores will also be reported.

Secondary aim #2 will be to assess the relative effect of randomization to the intervention on serious fall injuries.

The main comparisons of intervention groups with respect to the distribution of time until the first post-randomization occurrence of a serious fall injury will be based on survival analyses as mentioned in the primary aim. The Kaplan-Meier method will be used

to estimate the "survival" functions for participants in different intervention arms. To compare intervention arms, we will use a likelihood ratio test from a Cox regression model stratified by gender. Due to the expected small number of serious fall injuries, we have chosen not to stratify this outcome by field center. Failure time is measured from the time of randomization.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Secondary aim #3 will be to assess the relative effect of randomization to the intervention on persistent mobility disability.

Of primary interest is a comparison of the probability of being classified as having major mobility disability at two consecutive assessments. Persistent MMD will be defined as two consecutive determinations of MMD at 6-month assessment visits. Time until failure will be determined by the time from randomization until the initial failure. If a participant misses an intermediate visits, but still fails at two completed visits in a row, then time is still calculated as the time until the initial failure. Death after an initial MMD determination will be considered persistent mobility disability and event time will be considered the initial time of the MMD failure. Censoring time will be calculated as the time until the last definitive assessment for MMD, for those deemed not to have failed, or their last non-MMD assessment if persistent MMD has not been established. The main comparisons of intervention groups with respect to the distribution of time until persistent MMD will be based on survival analyses as mentioned in the primary aim. The Kaplan-Meier method will be used to estimate the "survival" functions for participants in different intervention arms. To compare intervention arms, we will use a likelihood ratio test from a Cox regression model stratified by gender. Due to the expected small number of persistent MMD cases, we have chosen not to stratify this outcome by field center.

Pre-specified subgroups will be the same as used for the primary outcome.

Secondary aim #4 will be to assess the relative effect of randomization to the intervention on the combined outcome of major mobility disability or death.

The main comparisons of intervention groups with respect to the distribution of time until the first post-randomization occurrence of the combined outcome of major mobility disability or death will be based on survival analyses. The same analysis described in the primary aim will be used. To compare intervention arms, we will use a likelihood ratio test from a Cox regression model, stratified by field center and gender. Failure time is measured from the time of randomization.

Pre-specified subgroups will be the same as used for the primary outcome.

Secondary aim #5 will be to assess the relative effect of randomization to the intervention on disability in activities of daily living.

Our analysis of new disability will be carried out within the subgroup of 1580 participants who reported no need for personal assistance with 6 ADL tasks (moving in and out of a chair, moving in and out of a bed, using toilet, dressing, bathing, and walking across a small room) at baseline. In the absence of a response by the participant, proxy responses will be used. Time until the initial report of need for personal assistance with any of these tasks will be used in the primary analysis. To compare intervention arms, we will use a likelihood ratio test from a Cox regression model, stratified by field center and gender. Failure time is measured from the time of randomization.

An additional analysis of disability will be performed on the disability score from 21-item version of the PAT-D and the 3 subscales (basic ADLs, 7 items; mobility disability, 8 items; instrumental ADLS, 6 tasks). Level of disability will be repeatedly measured as a

continuous, score variable. A comparison of average post-randomization levels of each outcome (overall score and subscales) between intervention groups will be performed using mixed-effects analysis of covariance techniques appropriate for repeatedly measured outcomes. An estimate of the effect size at follow-up visits will be obtained by using a contrast to estimate the difference between mean levels of the outcome for the control and intervention groups at each time point. These models will contain variables representing field center, gender, a follow-up time effect, the baseline level of disability score, the intervention effect and a follow-up time by intervention effect. The covariance between repeated measures will be characterized with an unstructured covariance structure.

Finally, we will also evaluate the onset of severe disability (defined as the initial report of the need for personal assistance in 3+ ADL's during follow-up of the 6 tasks previously defined) and development or progression of ADL disability (defined as the initial increase from baseline of at least 1 activity in the need for personal assistance for the six ADLs previously defined). Both of the above analyses will be performed on all 1635 participants and to compare intervention arms, for each outcome, we will use a likelihood ratio test from a Cox regression model, stratified by field center and gender. Failure time is measured from the time of randomization.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Secondary aim #6 will be to assess the cost-effectiveness of the intervention.

Cost-effectiveness analyses will be conducted following the guidelines of the Panel of Cost-Effectiveness in Health and Medicine. The ratio of direct costs of the physical activity intervention to the amount of quality-adjusted life years (QALYs) produced is calculated. Health care costs will be estimated and differences between the physical activity and lifestyle intervention groups will be calculated to examine whether any cost-offset may occur. LIFE takes a societal perspective. The trial uses the health education intervention as the comparator for all cost-effectiveness analyses. Results will be described as the incremental cost-effectiveness over the comparator. Sensitivity analyses will be conducted to examine whether the cost-effectiveness results change as a function of any estimates or assumptions made in the process. Decision modeling will be used to estimate long-term cost-effectiveness beyond the 1-year time horizon for which data collection is planned. Future health care costs will be discounted at a rate of 3% for any calculations or projections beyond the first year of follow-up.

Tertiary aim #1 will be to explore the effects of the interventions on MCI/D. Cognitive function will be assessed at baseline and at the 24-month follow-up visit and classification of participants to MCI or dementia will be based on a case adjudication review process. Logistic regression will be used to assess whether the proportion of participants diagnoses with MCI or D varies by intervention assignment. Baseline adjudicated status for MCI, baseline global cognitive function (3MSE score), clinical site, and education status (< HS, HS graduate) will be used as a covariates in this analyses.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Tertiary aim #2 will be to explore the effects of the intervention on composite measure of the cognitive assessment battery. We will construct a composite measure to include all components of the Cognitive Assessment Battery (but not the 3MSE, which is used to screen participants for MCI/D classification). For this composite measure of cognitive performance we will z-transform each score (DSST, HVLT, flanker, N-Back, and

task switching) by dividing its difference from the baseline mean by the baseline standard deviation. Prior to this transformation, 1% Winsorization of the flanker, N-back, and task switching scores will be applied to limit the influence of extreme values. This will involve replacing any scores less than the 1st percentile of the overall distribution of all scores with the value of the 1st percentile and replacing any scores greater than the 99th percentile of the overall distribution of all scores with the value of the 99th percentile. The following summary statistics will be used. DSST performance will be summarized by its overall score. HVLT will be summarized by the average of its z-transformed immediate and delayed recall scores (this average will be re-normalized to have standard deviation 1). Flanker will be summarized by the average of its z-transformed congruent and incongruent reaction times (this average will be re-normalized to have standard deviation 1). N-back will be summarized the average of the z-transformed 1-back and 2-back hits minus errors (this average will be re-normalized to have standard deviation 1). Task switching will be summarized by the average of the z-transformed no-switch and switch reaction times (this average will be re-normalized to have standard deviation 1). Z-scores will be calculated by dividing the difference between individual scores and the cohort-wide mean at baseline by the cohort-wide standard deviation at baseline. The DSST and HVLT are collected at the 24 month visit; the remaining tests are collected at either the 18 or 30 month visits. This analysis will be limited to participants who provide all data necessary for calculating this composite. The average of the z-scores from each of these five summary measures (which will be renormalized to have standard deviation 1), rather than adopting weighted averages, provides slightly greater emphasis on executive function. The relative effect of the PA intervention on this composite will be assessed with analyses of covariance applied to the composite measure, with the baseline composite included as a covariate and a marker to denote whether 18 or 30 month data were included. Differences between intervention arms for each individual test that contributes to the composite measure will be described. A similar approach will be used to assess the impact of the intervention on executive functions using a composite score for the three executive function tests (Flanker, N-back and Task Switching).

Pre-specified subgroups will be the same as used for the secondary cognitive outcome.

Tertiary aim #3 will be to assess the effects of the interventions on the SPPB score and 400m walk speed within subgroups defined by:

1. ethnicity/race (non-Hispanic white vs. other),
2. gender and
3. baseline physical performance (SPPB <8 vs. SPPB ≥8) subgroups.

Mixed effects models for repeatedly measured continuous outcomes (such as described for ADLs) will be used to explore these effects. These models will contain variables representing field center, gender, a follow-up time effect, the baseline level of SPPB/400MW speed, an intervention effect and the time by intervention interaction effect. Estimates of the intervention effect will be obtained within these subgroups and formal tests of interactions between intervention and subgroup variables will be performed. Forest plots will be used to graphically display the results of these subgroup analyses.

Tertiary Aims #4 will assess the intervention effect on sleep-wake disturbance (as measured by the Insomnia Severity Index (ISI), Epworth Sleepiness Scale (ESS), Pittsburgh Sleep Quality Index (PSQI), and Berlin Questionnaire (BQ)) using logistic regression models appropriate for repeated binary outcomes and pre-specified cutpoints for the continuous sleep-wake measures. The model will use generalized estimating equations (GEE), a logit link, and a binomial variance. An exchangeable correlation matrix

and a robust covariance estimate will be used. The proportion of participants reporting sleep-wake disturbances at each visit during follow-up will be compared between intervention groups and the model will contain terms for field center, gender, follow-up visit and an intervention by follow-up visit term. Similar models will be fitted using the sleep-wake disturbance measures as continuous outcomes.

An additional analysis of sleep-wake disturbances will focus on incident reports of problems. The analysis of each endpoint will be limited to the subgroup of participants that reported no difficulties at baseline for that endpoint (LIFE baseline prevalence of sleep-wake issues: 34% reported insomnia; 18% reported daytime drowsiness; 54% reported poor sleep quality; and 33% were at high risk for sleep apnea). Using logistic regression models, we will compare the proportion of participants reporting a new sleep-wake disturbance at either the 6- or 12-month follow-up (a few participants will have 30-month follow-up data). This model will contain terms for field center and gender.

Significance for the intervention effect will be based on the likelihood ratio test, and hazard/odds ratios and 95% confidence intervals will be constructed from the fitted models.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Tertiary Aim #5 will assess the dyspnea outcomes using ordinal logistic regression models with adjustment for field center and gender. The primary dyspnea outcome definition for the analysis will be based on a Borg dyspnea score of <0.5 (no dyspnea), 0.5 to 2 (mild >dyspnea), >2 (moderate-to-severe). Dyspnea score on the ATS scale of <1 (no dyspnea), 1 to 2 (mild dyspnea), >2 (moderate-to-severe), will be considered an outcome definition that is secondary to that generated from the Borg score. Ventilatory capacity will be examined using two types of mixed model analysis of variance models. First, we will examine for different post-baseline mean levels of FEV1 using a model adjusting for the baseline level, time (categorical), field center, and gender. We will account for within-person correlation using a subject random effect. Second, we will examine for a difference in the rates of decline using a random intercept/random slope mixed model. Both models will assume unstructured covariance matrices for repeated measures. Analyses will also be conducted on z-scores, standardized based on normal FEV1 values for combinations of sex, age, height and ethnicity.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Tertiary Aims #6 and 7 will be assessed using Cox proportional hazards models, stratified by field center and gender. Significance for the intervention effect will be based on the likelihood ratio test, and hazard ratios and 95% confidence intervals will be constructed from the stratified model. Failure time is measured from the time of randomization to the occurrence of a pulmonary disorder (Aim 6) or combined cardiovascular event (Aim 7) as determined by the adjudication review process.

Note that the composite events for airway and cardiovascular disease are defined in Section 12.1 of the LIFE Protocol. A secondary CVD event analysis will be based on an outcome comprised of CVD death or hospitalization for MI or stroke.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Tertiary aim #8 will be to explore whether the relative effect of the PA intervention on measures of cognitive function varies according to baseline level of global cognitive function. Separate estimates of the intervention effect on measures of cognitive function

will be obtained for participants grouped by baseline 3MS (<90 versus ≥ 90). Tests of interaction will be used to compare any differences between these groups.

Tertiary aim #9 will be to explore whether the intervention improves lower extremity blood flow, as assessed by ABI. A comparison of average post-randomization levels of ABI scores between intervention groups will be performed using mixed-effects analysis of covariance techniques appropriate for repeatedly measured outcomes. An estimate of the effect size at the 30-month visit and close-out will be obtained by using a contrast to estimate the difference between mean levels of the outcome for the control and intervention groups at each time point. These models will contain variables representing field center, gender, a follow-up time effect, the baseline level of ABI score and an intervention effect. The covariance between repeated measures will be characterized with an unstructured covariance structure.

Tertiary aim #10 will be to assess the relative effect of randomization to the intervention on the proportion of 400 m walk failures over time, as determined by an average intervention effect across repeated measurements.

This analysis will use the repeated 6-month indicators of 400 m walk status (rather than the time until the initial failure) and compare the average proportion of participants in each intervention group that fail the 400 m walk across all time points using a marginal model, a method that uses generalized estimating equations (GEE) and accounts for the dependency between repeated measures. The GEE analysis will use a logit link, a binomial variance, and an exchangeable correlation matrix when estimating model parameters, but will use a robust covariance when performing hypothesis tests for the overall intervention effect. Odds ratios for the association between 400m walk status and intervention will be estimated after adjusting for field center, gender, and a follow-up time effect.

Pre-specified subgroups will be the first four (i.e. ethnicity, gender, physical performance and age) specified for the primary outcome.

Missing data:

The advanced age and health frailty of the LIFE study population will result in missing data for most outcomes. The approach to missing data described herein is general and is intended to inform approaches for handling missing data in the LIFE study, not to serve as an explicit prescription unique to each outcome.

To identify factors that provide information as to the probability of missing outcomes, we will first compare the baseline characteristics of participants who do and do not have specific follow-up measures. Sensitivity analyses to determine how conclusions from primary outcome models may be affected by missing data will initially be performed by including covariate predictive of missing observations in such models. Such sensitivity analyses are intended as a conservative reexamination of data to explore whether reasonable assumptions placed on missing data might alter an observed finding, but primary consideration will be given to the original analysis of the aim. Non-significant intervention effects on outcomes (i.e., $p > 0.1$) will be subject to sensitivity analysis at the discretion of the analyst(s).

For longitudinal analyses that use maximum likelihood estimation, the original analysis plans have specified that all observed outcome values be used so as to allow for missingness to be dependent on previously observed outcomes (i.e., Missing At Random, MAR). To these models, we will add covariates that significantly predict missingness. Thus, these planned sensitivity analyses will account for the possibility that missing outcomes are dependent upon either observed covariates or previously observed

outcomes that are present in those analyses.

For some outcomes that have only a single follow-up visit, multiple imputation will be used to impute missing outcomes based on covariates observed at baseline. Sensitivity analyses of the primary intervention results for these outcomes will be performed under reasonable imputation models for the outcome of choice.

When missing outcomes are dependent on unobserved outcomes, potentially biased estimates of intervention effects due to differential missingness may occur. If this situation is suspected, then for continuous outcomes a multiple imputation approach will be developed that uses various underlying distributional assumptions for the missing observations within the imputation procedure to evaluate if overall conclusions from analyses change based on reasonable assumptions for the underlying distribution of the missing outcomes. For survival analyses, because follow-up will vary from 24 to 40 months, there will be right censoring of follow-up time; however, some individuals will drop out without having complete follow-up. For these analyses, inverse probability weighting will be used to perform sensitivity analyses of the primary results relative to assumptions about those that dropped out.