ARIC Manuscript Proposal #3998

PC Reviewed: 1/11/22 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Emulation of physical activity and diet intervention strategies for dementia prevention: Differences by educational attainment, race and study site in the ARIC Study

b. Abbreviated Title (Length 26 characters): Lifestyle & dementia

2. Writing Group: Chelsea Liu (first author), Yuan Ma, Albert Hofman, Melinda C. Power, Priya Palta, Casey M. Rebholz, James R. Pike, A. Richey Sharrett, M. Maria Glymour, Rebecca F. Gottesman (senior author); others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __CL___ **[please confirm with your initials electronically or in writing]**

First author: **Chelsea Liu**
Address: 73 Brighton Ave., #3, Allston, MA 02134

Phone: (919) 627-3188 Fax: N/A
E-mail: chelsealiu@g.harvard.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: **Rebecca F. Gottesman**
Address:

Phone: Fax:
E-mail: rebecca.gottesman@nih.gov

3. Timeline:
January – May 2022: data analysis
June – August 2022: writing and editing
September 2022: journal submission

4. Rationale:

Prior research has shown that higher levels of physical activity and a healthy diet are associated with lower risk of cognitive decline or dementia (Hu et al., 2020; Palta et al., 2019), and current strategies for dementia prevention largely focus on modification of these proximal factors (Tarique et al., 2018; Norton et al., 2014). The Mediterranean and DASH diets have been shown to delay the onset of cognitive impairment (Morris et al., 2015), and meta-analyses of observational studies and randomized controlled trials have shown that engaging in physical activity is associated with lower risk of dementia (Blondell et al., 2014; Smith et al., 2010). However, modification of risk factors at the individual level does not address the contributions of socioeconomic disparities, a fundamental factor in disease occurrence (Link & Phelan, 1995; Phelan et al., 2010).

Several important disparities have been observed in the study of dementia risk. Educational attainment is an important index of SES that is also a major risk factor for dementia. African American older adults are at higher risk of dementia compared to white older adults, an association that may be largely attributable to differences in educational opportunities (Avila et al., 2019; Weuve et al., 2018; Garcia et al., 2018). Lower educational attainment is associated with worse health behaviors such as smoking, physical inactivity and poor diet (Margolis, 2013). Neighborhood-level factors such as residing in census tracts that meet low-income and low-access thresholds may also prevent an individual from engaging in behaviors that lead to better vascular profiles, such as adhering to good dietary practices (Chen et al., 2016). These associations are likely due to the lack of access to healthcare resources, healthy food, as well as the necessary health literacy to carry out health behaviors that are thought to be “modifiable”. It may be additionally challenging to modify vascular and lifestyle risk factors for stroke survivors, a subpopulation that experiences a greater burden of these risk factors compared to those who have not had a stroke even prior to the stroke event (Liu et al, 2021).

Given the evidence that risk factors for dementia thought to be largely due to behavior are strongly influenced by socioeconomic disparities, further investigation is needed to understand the socioeconomic barriers to successful modification. To formally evaluate this question using observational data, we can use trial emulation methods such as the parametric g-formula in order to adjust for treatment-confounder feedback, which is not accounted for in other regression adjustment methods; this method emulates a trial more closely since time-varying confounding explains a substantial part of the discrepancies between observational studies and target trials (Hernan & Robbins, 2016). In this study, we aim to 1) understand whether hypothetical physical activity and diet modification for dementia prevention in the general population is possible under different levels of educational attainment and in different geographic regions of the U.S.; and 2) determine extent to which educational, racial and geographical disparities in dementia are explained by different distributions of physical activity and diet. We hypothesize that hypothetical interventions to increase physical activity and improve diet would be less effective among those with lower educational attainment, as well as in geographic areas with lower average educational attainment, due to the confluence of social and material conditions that greatly influence whether participants are able to make modifications.

5. Main Hypothesis:

We hypothesize that 1) hypothetical interventions in midlife diet or physical activity are less effective in reducing dementia risk among participants with lower levels of educational attainment and in geographic areas with the lowest proportions of participants categorized as high educational attainment (due to weaker associations between diet/physical activity and dementia in these groups); and 2) a significant proportion of the differences by education, race and study site in dementia risk will not be mediated through physical activity and diet (**Figure 1**).

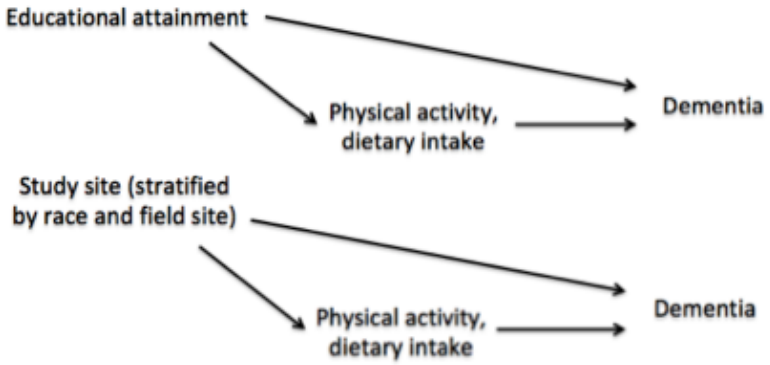

Figure 1. Hypothesized relationships between SES variables, intervention targets (physical activity, dietary intake), and incident dementia

6. Design and analysis:

Study design: We will conduct a secondary data analysis using data from ARIC.

Inclusion/exclusion: We will include Black participants from Forsyth County or Jackson; White participants from Forsyth County, Minneapolis or Washington County. We will include all participants with available education information at baseline, not missing APOE information, and who provided consent for their DNA to be used for research purposes.

Independent variables: Educational attainment is measured in categories with three levels: grade school or high school, high school graduate or vocational school, college and above, with the last group as the reference. Race-center is stratified by race and ARIC field center; we will use the group with the highest average educational attainment level as the reference. Physical activity and diet were measured in visit 1 (1987-1989) and visit 3 (1993-1995). Accordingly, we will only include incident dementia cases that arise after visit 3.

Hypothetical interventions. We will measure leisure-time moderate to vigorous intensity physical activity (MVPA) using participants' self-reported questionnaire responses on the duration and frequency spent in each of up to 4 leisure-time activities from the Modified Baecke Physical Activity Questionnaire (Baecke et al., 1982). Participants who report not participating in MVPA (0 minutes/week) will be categorized as inactive, and all remaining participants will be

categorized as low (1-74 min·wk^{-1}), middle (75-149 min·wk^{-1}), or high ($\geq$ 150 min·wk^{-1}) according to distribution-based tertiles described elsewhere (Palta et al., 2021).

We will use the Healthy Eating Index 2015 (HEI-2015), which was found to be associated with dementia risk (Hu et al., 2020), to measure dietary intake. The HEI-2015 was calculated from participants' responses to the 66-item semi-quantitative Food Frequency Questionnaire (FFQ; Willet et al., 1986), which asked participants how often, on average, they consumed food items of a given portion size in the previous year. The energy intake and total nutrient intake were calculated using data from the U.S. Department of Agriculture. The HEI-2015 score ranges from 0-100, and reflects how well participants adhered to recommendations from the 2015-2020 U.S. Dietary Guidelines for Americans (Krebs-Smith et al., 2020). We will categorize scores into tertiles corresponding to low, middle, or high HEI-2015.

Outcome: The primary outcome of interest is incident dementia after visit 3. Dementia cases were ascertained using the level 3 definition, which include cases seen in clinic and adjudicated by the ARIC expert panel as having dementia (visit 5-7), those determined by telephone assessment or informant interview, as well as those who were determined to have dementia by surveillance based on a prior discharge hospitalization ICD-9 or death certificate code (290.0, 290.1, 290.2, 290.3, 290.4, 290.9, 294.1, 294.2, 294.8, 294.9, 331.0, 331.1, 331.2, 331.8, and 331.9) from the date of the last participant contact up to the most recent follow-up (Knopman et al., 2016). Strokes were ascertained through annual follow-up phone calls to study participants or proxies, surveillance of discharges from local hospitals, and death certificates, with expert adjudication of stroke events (Koton et al., 2014). Follow-up will continue until Visit 7 (2018-2019).

Other variables: Potential confounders to be included in regression models are age at baseline, sex, race-center, APOE ε4 genotype status (0 or 1+ ε4 alleles), and other variables listed below.

Statistical analysis: In separate baseline tables, we will describe differences by 1) educational attainment (<HS, HS, >HS) and 2) race-center in the following variables: age, sex, APOE ε4 genotype status, visit 1 BMI, family history of CVD, visit 1 and visit 3 physical activity score (total MET-minutes per week), visit 1 and visit 3 dietary intake, and the proportion of participants who had incident dementia by the end of follow-up. We will assess death as a competing event.

Trial emulation via g-formula. We will use the parametric g-formula (McGrath et al., 2019), a time-varying extension of standardization, conduct a survival analysis for time-to-dementia in a hypothetical trial in which strategies (see below) are implemented for diet and physical activity. We will assess the change in dementia risk under the conditions of 1) no intervention, 2) each individual sustained intervention (i.e. high levels of MVPA for both visit 1 and visit 3; repeat for HEI-2015), and 3) joint sustained interventions (i.e. high levels of MVPA and HEI-2015 for both visit 1 and visit 3). We will then assess whether the effect of the hypothetical interventions differs by educational attainment or race-center. The risk of dementia will be estimated using the parametric g-formula with baseline covariates of age, marital status, education, employment, race-center, and family history of CVD (either parent's history of diabetes, stroke, CHD, or

premature history of CHD); and time-varying covariates of smoking, BMI, hypertension, diabetes, diet and physical activity.

G-formula specifications. We will use the g-formula to estimate the 30-year risk of incident dementia by educational attainment and by race-center under each of the strategies detailed in the table below.

Variable	**Strategies**
Physical activity (Visit 1, 3)	High MVPA *, middle MVPA, low MVPA, no change
Dietary intake (Visit 1, 3)	High HEI-2015, middle HEI-2015, low HEI-2015, no change

*Example interpretation for the "high PA" strategy: What is the direct causal effect of high educational attainment on dementia risk, had we intervened to give everyone high MVPA at visit 1 *and* visit 3.

To estimate the g-formula, we will specify a model for the outcome (incident dementia) given the independent variables (educational attainment, race-center) and covariate history and models for time-varying covariates (diet, physical activity, blood pressure, fasting blood glucose, smoking and BMI) given covariate history. The outcome model will be a pooled logistic regression model. In the outcome model, we will add interaction terms between the independent variable and the variables "intervened" upon (diet, physical activity) for each set of analysis in order to specify a more flexible model. Next, we will specify models for all time-varying covariates, including diet, physical activity (PA), systolic blood pressure (SBP), fasting plasma glucose (FPG), smoking (SMK) and BMI. To estimate survival under the "natural course" (no intervention on diet or physical activity), we will specify a model for the independent variables of interest (educational attainment or race-center) as a function of baseline covariates. All covariate models will be modeled using linear regressions, with the exception of the natural course models, which will be estimated using multinomial logistic regressions.

Independent variable: educational attainment

Outcome model:

Dementia = edu + diet + MVPA + edu*diet + edu*PA + age + sex + race-center + family history of CVD + APOE + BP + FPG + SMK + BMI

Time-varying covariate models:

Diet (visit i) = edu + age + race-center + family history of CVD + APOE + BMI (visit i – 1) + SMK (visit i – 1) + FPG (visit i – 1) + BP (visit i – 1)

MVPA (visit i) = edu + age + race-center + family history of CVD + APOE + BMI (visit i – 1) + SMK (visit i – 1) + FPG (visit i – 1) + BP (visit i – 1)

BP (visit i) = edu + age + race-center + family history of CVD + APOE + BMI (visit i – 1) + SMK (visit i – 1) + fasting blood glucose (visit i – 1)

FPG (visit i) = edu + age + race-center + family history of CVD + APOE + BMI (visit i – 1) + SMK (visit i – 1) + BP (visit i – 1)

SMK (visit i) = edu + age + race-center + family history of CVD + APOE + BMI (visit i – 1) + FPG (visit i – 1) + BP (visit i – 1)

BMI (visit i) = edu + age + race-center + family history of CVD + APOE + FPG (visit i – 1) + SMK (visit i – 1) + BP (visit i – 1)

(**Natural course model)** Edu = age + sex + race-center + family history of CVD + APOE

Independent variable: race-center

Outcome model: Dementia = race-center + diet + MVPA + race-center*diet + race-center*PA + age + sex + edu + family history of CVD + smoking + APOE

Time-varying covariate models: same as above

(**Natural course model)** Race-center = age + family history of CVD + APOE

Mediation analysis. Using single and multiple causal mediation methods (Huang & Yang, 2017), we will estimate the proportion of disparity in dementia risk that could be eliminated if participants with low educational attainment had the same distribution of diet and physical activity as participants with high educational attainment. To avoid adjusting for variables on the causal pathway between educational attainment and dementia, the mediation models will be adjusted for age, sex, race-center, APOE ε4 genotype (carrier vs. non-carrier), and family history of CVD. We will repeat these analyses using race-center as the socioeconomic variable of interest (accordingly, we will adjust for educational attainment instead of race-center in the mediation models).

Descriptive analyses will be conducted using SAS version 9.4 (Cary, NC). We will use the gFoRmula R package (McGrath et al., 2019) to estimate the g-formula, and the mediation R package for time-to-event outcomes and multiple mediators (Huang & Yang, 2017) to conduct mediation analyses.

Limitations or challenges: First, we acknowledge that the covariates available for this analysis may be insufficient to eliminate substantial confounding. Second, diet has not shown a strong association with incident dementia in ARIC, and this may limit the precision of our estimation of it effects within education or center strata. Also, the epidemiological literature, particularly in the study of "lifestyle" factors, focus largely on modeling the risk factors for disease and less on the social production of disease (Krieger, 1996). Social and material conditions that are largely predetermined due to historical processes relating to race and class—and, consequently, educational opportunities—significantly influence disease risk, and this effect may be magnified over the life-course and in the study of late-life diseases such as dementia. Current preventions strategies for dementia largely focus on modification of lifestyle. If we show that even if everyone had hypothetically "modified" physical activity and diet to the optimal levels, the risk of dementia is still higher in those with lower education, it would suggest that modification of these downstream variables, as measured in ARIC, is not sufficient. In summary, the extent to

which physical activity and diet are truly "modifiable" is highly dependent on socioeconomic factors such as access to resources, neighborhood and zip code, etc. Results will need to be interpreted carefully with attention to the mechanism through which these factors are modified. Stratifying our analysis by educational attainment and race-center is intended to further elucidate the contributions of those socioeconomic disparities. Quantifying these relationships is the first step to ensuring that future interventions are effective and feasible for individuals who lack access to healthcare or other resources and are at higher risk of the disease.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __x_ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No
(The file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __x__ Yes ___ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __x__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html
___x___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

- Adherence to Dietary Patterns and Risk of Incident Dementia: Findings from the Atherosclerosis Risk in Communities Study (Hu)
- Leisure-time physical activity sustained since midlife and preservation of cognitive function: The Atherosclerosis Risk in Communities Study (Palta)
- **1898b** – Midlife cardiovascular health and 20-year cognitive decline: Atherosclerosis Risk in Communities Study results (Terraf)
- **2284** – Lifetime Socioeconomic Position and Cognitive Decline (Kucharska-Newton)
- **1858b** – Midlife occupation and cognitive decline: the ARIC study (Liu)
- **3581** – The Moderating Influence of Education and Lifestyle on Genetic Risk for Dementia (Lee)
- **3677** – An Evaluation of Life's Simple 7 Score in Midlife in Offsetting the Genetic Risk of Dementia (Tin)

- **3210** – Life Course Individual and Neighborhood Socioeconomic Status and Risk of Dementia and MCI in the Atherosclerosis Risk in Communities (ARIC) Neurocognitive Study (NCS) (George)
- **2311** - Individual and contextual socioeconomic profile and physical function in late life: the Atherosclerosis Risk in Communities (ARIC) Study (Palta)
- **1742** - Education and Cognitive Change over 15 Years: The Atherosclerosis Risk in Communities Study (Schneider)
- **2120c** - Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort (Gottesman)

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes __ x __ No

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* _____)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _

*ancillary studies are listed by number https://sites.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References

1. Hu EA, Wu A, Dearborn JL, et al. Adherence to Dietary Patterns and Risk of Incident Dementia: Findings from the Atherosclerosis Risk in Communities Study. *J Alzheimers Dis*. 2020;78(2):827-835. doi:10.3233/JAD-200392
2. Palta P, Sharrett AR, Gabriel KP, et al. Prospective Analysis of Leisure-Time Physical Activity in Midlife and Beyond and Brain Damage on MRI in Older Adults. *Neurology*. 2021;96(7):e964-e974. doi:10.1212/WNL.0000000000011375
3. Tariq S, Barber PA. Dementia risk and prevention by targeting modifiable vascular risk factors. *J Neurochem*. 2018;144(5):565-581. doi:10.1111/jnc.14132
4. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol*. 2014;13(8):788-794.
5. Morris MC, Tangney CC, Wang Y, et al. MIND diet slows cognitive decline with aging. *Alzheimers Dement*. 2015;11(9):1015-1022. doi:10.1016/j.jalz.2015.04.011
6. Blondell SJ, Hammersley-Mather R, Veerman JL. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC Public Health*. 2014;14:510. Published 2014 May 27. doi:10.1186/1471-2458-14-510
7. Smith PJ, Blumenthal JA, Hoffman BM, et al. Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials. *Psychosom Med*. 2010;72(3):239-252. doi:10.1097/PSY.0b013e3181d14633
8. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995;Spec No:80-94.
9. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav*. 2010;51 Suppl:S28-S40. doi:10.1177/0022146510383498
10. Avila JF, Vonk JMJ, Verney SP, et al. Sex/gender differences in cognitive trajectories vary as a function of race/ethnicity. *Alzheimers Dement*. 2019;15(12):1516-1523. doi:10.1016/j.jalz.2019.04.006
11. Weuve J, Barnes LL, Mendes de Leon CF, et al. Cognitive Aging in Black and White Americans: Cognition, Cognitive Decline, and Incidence of Alzheimer Disease Dementia. *Epidemiology*. 2018;29(1):151-159. doi:10.1097/EDE.0000000000000747
12. Garcia MA, Saenz J, Downer B, Wong R. The role of education in the association between race/ethnicity/nativity, cognitive impairment, and dementia among older adults in the United States. *Demogr Res*. 2018;38:155-168. doi:10.4054/DemRes.2018.38.6
13. Margolis R. Educational differences in healthy behavior changes and adherence among middle-aged Americans. *J Health Soc Behav*. 2013;54(3):353-368. doi:10.1177/0022146513489312
14. Chen D, Jaenicke EC, Volpe RJ. Food Environments and Obesity: Household Diet Expenditure Versus Food Deserts. *Am J Public Health*. 2016;106(5):881-888. doi:10.2105/AJPH.2016.303048
15. Liu C, Roth DL, Gottesman RF, Sheehan OC, Blinka MD, Howard VJ, Judd SE, Cushman M. Change in Life's Simple 7 Measure of Cardiovascular Health After Incident Stroke: The REGARDS Study. *Stroke*. 2021;52(3):878-886. doi:10.1161/STROKEAHA.120.030836

16. Richardson MT, Ainsworth BE, Wu HC, Jacobs DR Jr, Leon AS. Ability of the Atherosclerosis Risk in Communities (ARIC)/Baecke Questionnaire to assess leisure-time physical activity. *Int J Epidemiol*. 1995;24:685–693.
17. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985; 122: 51–65.
18. Knopman DS, Gottesman RF, Sharrett AR, et al. Mild cognitive impairment and dementia prevalence: the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)*. 2016;2:1-11
19. Koton S, Schneider ALC, Rosamond WD, Shahar E, Sang Y, Gottesman RF, Coresh J. Stroke incidence and mortality trends in US communities, 1987 to 2011. *JAMA.* 2014; 312:259-68.
20. McGrath S, Lin V, Zhang Z, et al. gfoRmula: An R Package for Estimating the Effects of Sustained Treatment Strategies via the Parametric g-formula. *Patterns (N Y)*. 2020;1(3):100008. doi:10.1016/j.patter.2020.100008
21. Huang YT, Yang HI. Causal Mediation Analysis of Survival Outcome with Multiple Mediators. *Epidemiology*. 2017;28(3):370-378. doi:10.1097/EDE.0000000000000651
22. Krebs-Smith SM, Pannucci TE, Subar AF, et al. Update of the Healthy Eating Index: HEI-2015 [published correction appears in J Acad Nutr Diet. 2019 Aug 20]. *J Acad Nutr Diet*. 2018;118(9):1591-1602. doi:10.1016/j.jand.2018.05.021
23. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. The American journal of clinical nutrition 1982;36:936-942.
24. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575-1581. doi:10.1249/MSS.0b013e31821ece12
25. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254