ARIC Manuscript Proposal #4449

PC Reviewed: 05/14/24 **Status: _____** **Priority: 2**

1.a. Full Title: The association of midlife clusters of morbidity with cognitive decline and dementia

b. Abbreviated Title (Length 26 characters):

2. Writing Group:
Writing group members:

First author: Marco Egle
Other authors: Elise Kinyanjui, Renee C. Groechel, Valerie Morrill, Keenan Walker, Anna Kucharska-Newton, Thomas Mosley, Silvia Koton, David S Knopman, *Others welcome*

Last author: Rebecca F. Gottesman

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

First author: Dr. Marco T. Egle
Address: Building 10, Room 4D37, 10 Center Drive, Bethesda, MD 20814

Phone: 220-234-2831
E-mail: eglemt@nih.gov

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: Dr. Rebecca F. Gottesman
Address: Building 10, Room 4D37, 10 Center Drive, Bethesda, MD 20814

Phone: 301-435-9321
E-mail: rebecca.gottesman@nih.gov

3. Timeline:

Manuscript prepared and submitted in ~5-6 months.

4. Rationale:

Dementia is a major public health burden affecting more than 55 Million people worldwide. Previous evidence has highlighted the important role of midlife risk factors such as smoking, diabetes, prehypertension, and hypertension implicated in the risk of dementia[1, 2]. In fact, the risk for late-life dementia predicted by diabetes in midlife was almost as high as having at least 1 APOE e4 genotype[2]. More recent evidence has furthermore shown that the prevalence of multimorbidity (≥2 chronic diseases), particularly when starting in midlife, is strongly associated with dementia risk while progressively weakening at later age of onset of multimorbidity[3]. Altogether, these findings emphasize that multiple chronic conditions and their combination as early as in midlife are important determinants for cognitive decline and dementia.

In an age of personalized medicine, it is crucial not only to understand the associations between individual morbidities and dementia, but to understand how combinations of comorbidities might determine risk of dementia, and to evaluate these combinations at the person level. Therefore, instead of focusing on which singular clinical conditions, at the disease or risk factor level, are associated with dementia risk, it may be valuable to evaluate individuals themselves who often have several co-existing conditions. New statistical tools such as cluster analysis allows us to break down the dimensionality of the risk factor data per individual and to allocate people with similar clinical profiles to specific clusters of morbidities, which can subsequently be characterized by defining clinical features. These unsupervised approaches have become more popular over the recent years and have yielded interesting findings with regard to multimorbidity trajectories over time and mortality risk[4], post-stroke disability[5], and dementia risk[6]. In the UK biobank study, investigators identified several clusters of morbidity associated with dementia risk which showed differences when stratified by sex[6]. In women, dementia risk was highest for a cluster characterized by hypertension, diabetes, and coronary heart disease. In contrast, men had the highest dementia risk when being allocated to a cluster defined by diabetes and hypertension. The associations of the clusters with dementia were furthermore modified by genetic APOE-e4 carrier status. These

findings provide new insight into the risk of dementia conversion. One significant limitation, however, is the UK Biobank's lack of racial diversity. Up to 97% of the participants were of White race in the cluster analysis which makes the results of the data-driven cluster approach difficult to generalize.

Using data from the multi-racial ARIC study, we can build upon previous work in primarily White participants to explore whether certain clusters of morbidities in midlife are associated with a higher risk of dementia and cognitive decline. Specifically, given observed similarities between risk factors for stroke and risk factors for dementia, when explored individually, we hypothesize that clusters with strong associations with stroke and its severity would also be especially important in dementia risk. The proposed project is an extension of an existing ongoing project in ARIC where the above-described methods were applied to generate clusters of morbidities at Visit 1, and these clusters were evaluated in association with stroke incidence and severity. In that study, we show that all nine clusters, when compared to a cluster of individuals with no defining risk factor predominance, were associated with 30-year stroke incidence and that the strength of association varied by stroke severity. The association was strongest for any of the clusters which were defined by coronary heart disease, heart failure, atrial fibrillation, and renal dysfunction. In this project, we would like to test whether these same clusters, shown as important in future stroke incidence and its severity, are also associated with dementia conversion and with cognitive decline, and whether the associations vary by demographic and genetic characteristics such as race, sex, educational attainment, and APOE-e4 carrier status. As stroke and dementia are highly linked[7], the new project would allow us to draw new parallels on common patterns of morbidities in midlife which increase the risk of both clinical outcomes.

5. Main Hypothesis/Study Questions:

- All clusters of morbidity in midlife, when compared to a cluster of individuals with no defining risk factor predominance, are associated with greater risk of incident dementia and steeper cognitive decline
- Midlife clusters defined by coronary heart disease, heart failure, atrial fibrillation and renal dysfunction are most strongly associated with dementia risk

- Midlife risk factor clusters are associated with cognitive decline between midlife and late life (visit 2-visit 5) as well as with cognitive decline in late life (visit 5 to most recent visit). The association with mid-to- late life cognitive decline will be strongest for those midlife clusters defined by coronary heart disease, atrial fibrillation, heart failure and renal dysfunction. We hypothesize that the association with cognitive decline in late life will be strongest for midlife clusters defined by smoking, obesity, and diabetes.

Secondary hypothesis:

- The association between the clusters and dementia/ cognitive decline is modified by sex, race, APOE e4 carrier status, and education. We specifically hypothesize that the associations of disease clusters with dementia risk will be stronger for women, for positive APOE-4 carriers, and those with lower education.
- Midlife clusters of morbidity are associated with dementia before age 75. The association with dementia before age 75 will be strongest for the clusters defined by coronary heart disease, heart failure, atrial fibrillation and renal dysfunction.

6. Design and analysis (study design, inclusion/exclusion, outcome, and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Participant inclusion: Among the 15,792 participants enrolled at ARIC baseline, 15,404 participants from only White and Black race-centers (Washington site White (*N*= 3,907); Forsyth county Black (*N*= 466); Forsyth county White (*N*= 3,484); Jackson county Black (*N*= 3,619); Minneapolis White (*N*= 3,929)) will be included in this study. Participants will also be excluded if they had ischemic stroke prior to baseline (*N*= 282) or prevalent dementia at baseline (n=1)..

Outcome:

1. Dementia incidence between baseline and the latest available visit will be included. We will use the censored level 3 dementia diagnosis as the event indicator and the corresponding adjusted date as the event time.
2. Cognitive Decline

- Three cognitive assessments were used to assess cognition at visits 2 and 4. These were the Delayed Word Recall task, Digit Symbol Substitution from the Wechsler Adult Intelligence Scale -Revised (WAIS-R), and a Letter Fluency task.
- At visit 5 and beyond, cognitive function was measured with a more comprehensive cognitive battery as part of Atherosclerosis Risk in Communities Neurocognitivc Study (ARIC-NCS) study. The cognitive measures were based on a variety of neuropsychological tests which have been described in detail previously.[8]

We will use cognitive factor scores generated as part of prior ARIC efforts from visit 2 onwards[9].

Demographic and risk factors

Demographic and risk factors at ARIC visit 1 will be included in the study (**Table 1**). Demographic measures will include age, sex, race, center, and education level. APOE-e4 carrier status will also be included in the analysis.

Table 1. Risk factor and chronic health conditions included in the cluster analysis.

Variables	Variable descriptor
Cancer	*hom10f*:
Coronary heart disease	*prvchd05*:
Current drinking	*curdrk02*:
Current smoking	cursmk01:
Diabetes	*diabts03*:
Heart failure	*prevhf01*:
Hypercholesteremia	*tchsiu01:*
Hypertriglyceridemia	trgsiu01
Hypertension	*hypert05:*
Obesity	*bmi01:*
Atrial fibrillation	*v1ecg22*
Vital exhaustion*	*hpba01-21*
Peripheral artery disease	pad02
Renal dysfunction	egfrcr_v1

**Please note, Maastricht questionnaire of vital exhaustion was administered at visit 2 but will be included as a visit 1 measure*

Data analysis

The statistical analysis will be conducted using the R software programing language. The cluster analysis and characterization has been carried out previously in a separate ARIC study (MP#4267) evaluating these clusters and incident stroke/ stroke severity; the same clusters will be used in this study.

1.Summary of the already performed cluster analysis

1.1 Factor transformation into numerical coordinates

A multiple correspondence analysis (MCA) was first employed as a dimensionality reduction technique to transform the risk factors into numerical coordinates of multiple space dimensions.

1.2 Measuring the similarity between individuals based on their clinical profiles

The Euclidian distance measure was then used on the transformed data structure to quantify the similarity in the coordinates between the individuals.

1.3 Cluster analysis to stratify the population into distinct groups

We then employed an unsupervised hierarchical cluster analysis using Ward's agglomerative criterion with k-means consolidation (HCPC function, from FactoMineR package, Vienna, Austria) to assign individuals to more homogenous groups characterized by similar clinical profiles. The optimal number of clusters was automatically determined based on the gain of the within-inertia (variance within the clusters).

1.4 *Cluster characterization*

The following metrics were used: (1) the observed vs. expected ratio; 2) the intra-cluster prevalence. The observed/expected ratios was computed by dividing the prevalence of a given condition within a cluster by its prevalence in the overall population. A condition was associated with a cluster when the observed/expected ratio was ≥1.5, and when the intra-cluster prevalence of a condition was ≥ 0.25, which means that at least 25% of individuals with a certain condition are found in a cluster[5]. We identified 9 clusters characterized by the following defining conditions and with the following sample size: cluster 1 (no defining condition, relatively healthy cluster, *N*= 6332), cluster 2 (smoking, *N*= 2973), cluster 3 (cancer, *N*=719), cluster 4 (peripheral artery disease, *N*=560), cluster 5

(diabetes, obesity, hypertriglyceremia, hypertension, *N*= 3466), cluster 6 (coronary heart disease, *N*=528), cluster 7 (atrial fibrillation, *N*=27), cluster 8 (heart failure, *N*=615), and cluster 9 (renal dysfunction, *N*=184).

2.Association of the clusters with dementia

In the primary analysis, the association between the clusters in midlife and incident dementia will be tested by Cox proportional hazard regression models. These models will be adjusted for age, sex, race-center, APOE-e4 carrier status, and educational level. The assumptions underlying the regression models such as the proportional hazards assumption and the linearity between log hazard and the covariates will be tested. In a sensitivity analysis, it will be tested whether the association is modified by the covariates sex, education, race, and APOE-e4 carrier status.

Given that mortality is a strong competing risk factor in the analysis, a sensitivity analysis will be conducted which tests the associations between the clusters and dementia with a competing risk model (dementia vs. no dementia) using the function “crr” of the “cmprsk” R package and of the “tidycmprsk” package extension. Competing risk analysis is a type of survival analysis which estimates the marginal probability of an outcome event in the presence of one or several competing events based on the cumulative incidence function (CIF). In this analysis, the competing event is lifetime mortality. The competing risk models will be adjusted by age, sex, race-center, APOE-e4 carrier status, and educational level.

In a separate multinomial logistic regression analysis, it will be assessed whether the clusters in midlife are more strongly associated with early vs. late dementia conversion. The following factor levels will be used: dementia conversion before age 75; dementia conversion at or after age 75; no dementia conversion. The factor level “no dementia conversion” will be the reference level. The covariates age, sex, education, race-center and APOE-e4 genotype will be added as covariates.

3.Association of the clusters with cognitive decline

3.1 Preliminary analysis and data preparation:

We will assess any systematic pattern of missingness in the data using the MICE R package and follow the statistical recommendation provided in the ARIC_Manual_30_220919.

3.2 Mixed effects model to model cognitive function over time

As described in ARIC Manual 30, we will use a mixed effects model to model the trajectory of cognitive decline. We will opt for a linear mixed effect model to analyze cognitive scores instead of a generalized estimating equation (GEE) for the reasons described in ARIC Manual 30. The mixed effect model will include both a random intercept and slope. Models with different types covariance structures, including the unstructured covariance matrix, for the random effects will be evaluated. Using the log likelihood estimation, we will compare the different types of mixed effects models and choose the model with the best model fit to the cognitive data. We will also run 2 additional mixed effect models on change in cognitive function between visit 2 and visit 5 (i.e. cognitive function between midlife and late life) as well as between visit 5 and the latest available visit (i.e. cognitive function in late life only). With regard to the model's diagnostics, the scatter plots of the residuals and the LOWESS line will be evaluated. We will specifically examine the following scatterplots: (1) overall; (2) racial subpopulation; (3) different levels of exposures; (4) different levels of exposure stratified by racial subpopulation. We will furthermore assess the normality of errors in the model by inspecting the model's residuals in a histogram or Q-Q plot. In case the underlying model assumptions are not sufficiently met, we may have to adjust the model.

3.3 Linear regression model to assess the cluster's association with cognitive decline

Linear regression model will be employed to test the association between cluster of morbidity and the trajectories of cognitive decline derived from linear mixed model while accounting for age, sex, education, race-center, and APOE-4 status. The statistical assumptions underlying the linear regression model will be graphically inspected and met. These assumptions are (1) the normality of the residuals, (2) homoscedasticity, (3) linear relationship, (4) multicollinearity (5) influential points. In case the underlying assumptions of normality and homoscedasticity are not met, we may use a more robust linear regression model with 2000 bootstrap replications to derive the 95% confidence intervals of the regression estimates. We will also assess the interaction effect between

cluster x sex, cluster x education, cluster x race and cluster x APOE-4 status and stratify the analysis by those clinical variables accordingly.

Anticipated methodologic limitations or challenges?

- The way we define disease burden in this proposal is not conventional. Instead of quantifying the overall health burden (2 or more diseases) in a patient (i.e. disease-centered approach), we are interested in a data-driven approach, i.e., hierarchical clustering analysis, which identifies distinct groups of individuals (i.e. individual-centered approach) that are characterized by specific disease clusters. These clusters will hopefully provide valuable information on dementia risk and cognitive decline.
- Attrition is a significant challenge when working with the cognitive measures in the ARIC data. We will follow the recommendations proposed in the ARIC statistical manual as closely as possible. We will employ multivariate imputations strategies in the context of attrition as recommended in the manual.

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/aricproposals/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)?

- MP #4267 Egle et al. The association between midlife and late life risk factor and chronic disease clustering and stroke incidence and stroke severity
- MP #2120 Gottesman et al. Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort

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

11.b. If yes, is the proposal

A. primarily the result of an ancillary study (list number*_2008.06, ARIC-NCS)

___ 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.
Understood.

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.

Bibliography

1. Shin JH. Dementia Epidemiology Fact Sheet 2022. Ann Rehabil Med 2022;46:53-59.
2. Gottesman RF, Albert MS, Alonso A, et al. Associations between midlife vascular risk factors and 25-year incident dementia in the Atherosclerosis Risk in Communities (ARIC) cohort. JAMA Neurology 2017;74:1246-1254.
3. Ben Hassen C, Fayosse A, Landre B, et al. Association between age at onset of multimorbidity and incidence of dementia: 30 year follow-up in Whitehall II prospective cohort study. BMJ 2022;376:e068005.
4. Vetrano DL, Roso-Llorach A, Fernandez S, et al. Twelve-year clinical trajectories of multimorbidity in a population of older adults. Nat Commun 2020;11:3223.
5. Egle M, Wang W-C, Fann YC, et al. Sex Differences in the Role of Multimorbidity on Poststroke Disability: The Taiwan Stroke Registry. Neurology 2024;102:e209140.
6. Calvin CM, Conroy MC, Moore SF, Kuzma E, Littlejohns TJ. Association of Multimorbidity, Disease Clusters, and Modification by Genetic Factors With Risk of Dementia. JAMA Netw Open 2022;5:e2232124.
7. Koton S, Pike JR, Johansen M, et al. Association of Ischemic Stroke Incidence, Severity, and Recurrence With Dementia in the Atherosclerosis Risk in Communities Cohort Study. JAMA Neurol 2022;79:271-280.
8. 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.
9. Gross AL, Power MC, Albert MS, et al. Application of Latent Variable Methods to the Study of Cognitive Decline When Tests Change over Time. Epidemiology 2015;26:878-887.