1.a. **Full Title:** Dynamics of Risk Factors for 32-Year Incident Dementia Across Mid And Late-Life in ARIC: Implications for Dementia Prevention

b. **Abbreviated Title (Length 26 characters):**

PAF and Dementia

2. **Writing Group:**

Writing group members:

Paulo H.M. Chaves  
Josef Coresh  
Jennifer A. Deal (Senior Author)  
Rebecca F. Gottesman  
David S. Knopman  
Mark Lee  
Pamela L. Lutsey  
Priya Palta  
Nicholas S. Reed  
Andrea L. C. Schneider  
A. Richey Sharrett  
Moyses Szklo  
Jason R. Smith (Lead Author)  
B. Gwen Windham

Other interested ARIC authors/investigators are welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. **JRS**

**First author:**

Name: Jason R. Smith  
Address: 615 N. Wolfe Street, Baltimore, MD 21205  
Phone: 919-265-3337  
E-mail: jsmit491@jhu.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: Jennifer A. Deal
Address: 2024 East Monument Street, Suite 2-700,
Baltimore, MD 21205
Phone: 410-955-1909
Fax: 410-614-9625
E-mail: jdeal1@jhu.edu

3. Timeline:

Analysis and manuscript development will be completed in 12 months.

4. Rationale:

The number of Americans living with Alzheimer’s dementia is expected to nearly double to 13 million by 2050. Dementias of all types are chronic conditions with a protracted latency period that is associated with diminished quality of life, substantial medical expenditures and societal burden, and high mortality. There are a constellation of strong risk factors for all-cause dementia across the life course, the consequences of which have yet to be fully understood for prevention efforts.

Life course epidemiology as a conceptual framework posits that over and above the independent contribution of chronologic age, factors related to health outcomes vary in predictive value over secular time. Under this framework, we would not expect to see consistent associations between diabetes, hypercholesterolemia, or hearing loss and dementia in midlife and late life. In fact, due solely to the statistics, we should expect to see relative (e.g., ratio) associations attenuated among older adults. The risk of disease tends to increase with age, and as the risk of the outcome increases in the comparison group, the relative risk must decrease. Furthermore, risk factors emerging later in life (e.g., late-life diabetes) have accumulated less exposure time to exert their effects. Importantly, this same statistical phenomenon tends to not be observed on the additive scale. Neglecting this reality can potentially lead to peculiar population-level inferences when relying only on relative measures. When investigating associations between blood pressure and dementia, for example, some studies (e.g.) have concluded high blood pressure in late life is not harmful, and perhaps even apparently protective against dementia risk among older adults, contrary to results from the SPRINT-MIND randomized trial, which found lower risk of mild cognitive impairment with an intensive blood pressure lowering regimen (<120 mmHg) compared to a standard target of 140 mmHg.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Incidence in the Exposed (per 100)</th>
<th>Incidence in the Unexposed (per 100)</th>
<th>Measure of Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk Difference</td>
<td>Relative Risk</td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>60-69</td>
<td>40</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>70-79</td>
<td>60</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>80-89</td>
<td>80</td>
<td>70</td>
<td>10</td>
</tr>
</tbody>
</table>

Table. Hypothetical Example of How Relative Measures of Association May Decrease with Increasing Incidence in the Unexposed Group
With respect to dementia research, measures of association that reconcile the differing prevalence of risk factors and their differing level of risk in the population across time can help derive actionable data for public health and policy planning. In this case, one estimate of interest is the population attributable fraction (PAF); a measure of disease occurrence with respect to the prevalence of the exposure in the population compared to the counterfactual scenario of disease occurrence in the complete absence of the exposure in the population. Stated differently, assuming the exposure has a causal effect on the outcome, the PAF provides the upper bounds for the percentage of cases that would be prevented in the given population if the exposure is prevented or removed entirely. Importantly for this proposal, the PAF is a measure of absolute, not relative, difference.

Although complete removal of an exposure in a population is implausible, estimating the PAF across time is valuable for identifying priorities for future research and targets for prevention over the life course. Estimates of the absolute number of dementia cases that could be prevented in the US, for example, if hypertension were reduced in midlife versus late life could be leveraged for population-level strategic planning and policy prioritization. A well-characterized and biracial community-based sample with ample participant follow-up, the ARIC study is uniquely situated to provide stratified estimates reflecting fluctuations in prevalence of key exposures in mid- and late life which other studies have been unable to provide.

Prior work in ARIC has estimated the risk of dementia associated with midlife vascular risk factors, and current projects are estimating weighted communal PAFs for dementia by leveraging meta-analytic risk factor and national prevalence data to probe racial disparities. Moreover the recent Lancet Commission estimated weighted PAFs (some worldwide if data were available) for modifiable risk factors in one life-course period (either early, mid, or late life; whichever was deemed most important).

In this proposal, we aim to extend this work by using a life-course approach to estimate and present the PAFs in both midlife and late life for strong risk factors—including modifiable and non-modifiable—for incident dementia previously identified in ARIC studies, leveraging 30+ years of community-based follow-up data linked with two nationally representative surveys. These estimates will add to the growing literature on identifying periods during which preventing dementia is likely to have greater effects in the population. Particularly within the broader context of limited clinical benefit of dementia treatment demonstrated in randomized trials, these data are urgently needed and will yield an immediate benefit from both a public health and policy perspective.

5. Main Hypothesis/Study Questions:

What are the PAFs for known risk factors for 30-year incident dementia in the US, and does it differ when measured in midlife versus late life?

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).
Study design and sample

To estimate longitudinal associations of relevant risk factors in mid- and late-life and subsequent dementia in the ARIC study, we will employ a nonconcurrent cohort design. Our inclusion criteria will be all ARIC participants with race recorded at Visit 1 (1987-1989). We will then exclude participants: a) who identify as Non-Black and Non-White race, and Black participants at the Minnesota and Maryland field centers (in accordance with ARIC study analysis recommendations due to race-site aliasing), b) with dementia prior to baseline, and c) with missing covariate data. For the proposed analysis we will leverage data from Visit 1 (1987-89) through Visit 7 (2018-19), for a total of 32 years of follow-up.

We will define risk factors measured in midlife as those from Visit 1 and risk factors measured in late life as those from Visit 5 (2011-2013), with the exception of hearing loss—measured among all participants at Visit 6 (2016-2017). Given the wide age range of participants at Visit 1, we will additionally consider the use of age group as the time scale (e.g., restricting to age categories at Visit 1 and Visit 5) for alternative definitions of midlife and late life.

We will then estimate PAFs for dementia within the ARIC study, and PAFs generalizable to the US population. To calculate PAFs generalizable to the US population, we will leverage two population-based surveys: we will derive the prevalence of midlife risk factors from the National Health and Nutrition Examination Survey (NHANES), and late life risk factor prevalence from the National Health and Aging Trends Study (NHATS), where available.

Incident dementia diagnosis

We will define incident dementia diagnosis (“level 3” dementia in the ARIC database) using the standardized ARIC algorithm.10,13 The algorithm leverages: a) longitudinal cognitive data and complete neuropsychological battery data among participants attending clinic visits; b) supplemental cognitive data obtained outside clinic visits, including the Modified Telephone Interview for Cognitive Status (TICS-M) from interviews with the participant, the modified Clinical Dementia Rating scale from interviews with informants at Visit 5, and the six-item screener (SIS) of Alzheimer’s disease 8 (AD8) from Visit 6 and Visit 7; and c) ICD-9/10 claims-based definitions for dementia from hospitalizations or death certificates (with onset date estimated as 6 months pre-hospitalization).

Hypertension

We will parameterize blood pressure at Visit 1 and Visit 5 as a categorical variable using standard thresholds:14 a) hypotension (<90/<60 mm Hg); b) normotension (reference group; not meeting other criteria); c) prehypertension (between >120/>80 mm Hg and <140/<90 mm Hg); and d) hypertension (≥140/≥90 mm Hg or use of antihypertension medication).

Diabetes

We will parameterize diabetes mellitus at Visit 1 and Visit 5 as a binary variable based on clinical guidelines previously used in ARIC studies:15 no diabetes (reference group; not meeting
other criteria), and diabetes (self-reported physician diagnosis of diabetes or non-fasting glucose \( \geq 1200 \text{ mg/dL} \) or fasting glucose \( \geq 126 \text{ mg/dL} \)).

**Hypercholesterolemia**

We will obtain fasting plasma total cholesterol, triglyceride, and high density lipoprotein cholesterol levels (all in mg/dL) from Visit 1 and Visit 5; these are measured using enzymatic methods.\(^{16,17}\) Among participants with triglycerides <400 mg/dL, we will derive low density lipoprotein cholesterol (LDL-c) levels (mg/dL) using the Friedewald equation.\(^{18}\)

We will parameterize primary hypercholesterolemia in two ways: 1) LDL-c as a continuous variable; 2) LDL-c as a categorical variable given clinically relevant guidelines that inform secondary prophylactic treatment:\(^{19}\) a) \(< 100 \text{ mg/dL} \ (< 2.6 \text{ mmol/L}; \text{reference group, non-elevated})\); b) 100-159 mg/dL (2.6 – 4.0 mmol/L; mild hypercholesterolemia); c) 160-189 mg/dL (4.1 – 4.8 mmol/L; moderate hypercholesterolemia); d) \( \geq 190 \text{ mg/dL} \ ( \geq 4.9 \text{ mmol/L}; \text{severe hypercholesterolemia})\). Additionally, among those with hypercholesterolemia we might consider further characterization with respect to maintenance of suggested therapeutic targets for LDL-c levels (e.g., when prescribed statins or other lipid-lowering medications).

**Overweight and obesity**

We will derive body mass index (BMI) at Visit 1 and Visit 5 as kilograms/height m\(^2\) as a continuous value, then parameterize obesity as a categorical variable: a) underweight (BMI < 18.5); b) normal weight (reference group; BMI between 18.5 – 24.9); c) overweight (BMI between 25.0 – 29.9; and d) obese (BMI \( \geq 30.0 \)).

**Smoking**

We will parameterize self-reported smoking history at Visit 1 and Visit 5 as a categorical variable: a) never smoker (reference group); b) former smoker; and c) current smoker.

**Alcohol consumption**

Given the questions “Do you presently drink alcoholic beverages?” and “Have you ever consumed alcoholic beverages?”, we will parameterize self-reported alcohol consumption history at Visit 1 and Visit 5 as a categorical variable: a) never drinker; b) former drinker; and c) current drinker (reference group).

**Prevalent and incident stroke**

We will parameterize self-reported history of stroke (i.e., prevalent) diagnosed by a physician at Visit 1 as a binary variable.

Incident strokes, defined as stroke hospitalizations and death by ICD-9/10 codes, are identified by a computer algorithm and then adjudicated by experts.\(^ {20,21} \) We will utilize adjudicated stroke data available from Visit 1-Visit 5.
Prevalent coronary heart disease

We will parameterize prevalent coronary heart disease at Visit 1 as a binary variable: no versus yes (defined by self-reported history of physician diagnosis at baseline or adjudicated fatal coronary heart disease).

Physical inactivity

We will define leisure time physical activity (LTPA) leveraging the standardized interviewer administered questionnaire at Visit 1; an open-ended form concerning up to 4 leisure-time activities or sports played and the duration and frequency spent in each of the 4 leisure-time activities. Given the duration of self-reported activity, and of number metabolic-equivalent (MET; range, 1-12), MET-minutes/week was estimated over the past year by multiplying the frequency, duration, and MET value.

We will parameterize midlife LTPA at Visit 1 as a categorical variable using MET-min/week tertiles derived in Palta et al (2019): a) no LTPA (reference group); b) low LTPA; c) middle LTPA; and d) high LTPA.

Head injury

We will define head injury (with or without loss of consciousness) at Visit 1 and Visit 5 utilizing an ARIC algorithm; this leverages a) ARIC hospitalization surveillance data, b) CMS hospitalization/ED data, and c) self-report head injury questions (history of head injury that resulted in loss of consciousness, extended loss of consciousness, or long-term problems or dysfunction).

We will parameterize head injury at Visit 1 and Visit 5 as a binary variable.

Hearing loss

Pure tone air conduction audiometry is available for Visit 6 (2016-2017). We will calculate a better-hearing ear, 4-frequency pure tone average (PTA) in decibels hearing level (dB HL) using audiometric thresholds at the speech frequencies of 0.5, 1, 2, and 4 kHz.

We will define late life hearing impairment at Visit 6 by categorizing PTA using a clinically defined ordinal variable: a) no hearing impairment (reference group; < 25 dB HL); b) mild hearing impairment (26 - 40 dB HL); c) and moderate or greater hearing impairment, (> 40 dB HL).

Positive depressive symptomology

Consistent with previous ARIC studies, we will parameterize late life depression symptomology, measured by the 11-item Center for Epidemiologic Studies-Depression scale, at Visit 5 as a binary variable (i.e., score ≥ 9 categorized as “positive” for depressive
symptomology, score < 9 as negative). Although this threshold could result in some false positives, this is arguably desirable over more false negatives while retaining the ability to identify clinically relevant depressive symptomology.

**Education**

We will parameterize education at Visit 1 as a categorical variable (i.e., less than high school, high school or high school equivalent, some or full college, and graduate or professional school).

**APOE ε4**

A non-modifiable risk factor, we will analyze APOE ε4 genotype as a binary variable (0, 1+ ε4 alleles). If numbers permit (particularly for ε4/ε4), we will consider parameterization as a categorical variable (0, 1, 2 ε4 alleles).

**Covariates**

In addition to the above risk factors, we will incorporate demographic covariates collected at Visit 1 of age (derived from date of birth), sex, and race-center.

**Statistical analysis**

We will first describe the demographic and clinical characteristics of the overall study sample, then compare characteristics of those with versus without dementia in ARIC. Additionally, we will describe and compare the prevalence of midlife and late life risk factors in their respective study. To compare groups we will use two-sided t tests (normally distributed continuous variables), Mann-Whitney U tests (non-normal continuous variables), and chi-square tests (categorical variables) with alpha of 0.05; because of strong a priori hypotheses for our included risk factors, we will not adjust for multiple comparisons.

Next we will construct several Cox proportional hazards models using continuous time analysis to estimate multivariable-adjusted hazard ratios of incident dementia. Here, we will first estimate relative hazards of 32-year incident dementia by Visit 7 (2018-2019) among midlife risk factors (collected at Visit 1; 1987-1989). Then, among those that did not develop dementia by Visit 5 (2011-2013), we will estimate relative hazards of 8-year incident dementia by Visit 7 (2018-2019) among late life risk factors collected at Visit 5 (2011-2013). Using the raw data we will also derive estimates of cumulative incidence and relative risk (RR) and risk difference (RD).

Where appropriate, we will handle missing covariate data with multiple imputation. Participants that do not develop dementia over follow-up will be right censored at the latest date of assessment for Visit 7, TICS-M, informant interview, or last point of contact with participants before end of Visit 7. We will use residual plots to check the proportional hazards assumption; if it is not met, we will leverage more flexible models (e.g., stratification or interaction with time).

Leveraging the midlife and late life estimates of relative risk of dementia, we will next derive attributable fractions (AF) for each risk factor.
\[ AF = \frac{R_{\text{exposed}} - R_{\text{unexposed}}}{R_{\text{exposed}}} = \frac{RR - 1}{RR} \]

This will estimate the percentage of 32-year incident dementia among each exposure that are attributable to the exposure in midlife, and similarly for 8-year incident dementia for late life.

Because the prevalence of many of these risk factors varies by age, however, we will then use the estimated prevalence of each risk factor in the US population (in midlife, using NHANES; in late life, using NHATS) and the calculated AF to derive the unadjusted PAF for each risk factor in midlife and late life, respectively,

\[ pAF = \frac{P_{\text{population}} - P_{\text{unexposed}}}{P_{\text{population}}} = \frac{P_{\text{population}} \times (RR - 1)}{P_{\text{population}} \times RR} \]

where \( P \) is the prevalence of the exposure in the population. We will then use additional formulas (e.g.,\(^{26}\)) to obtained adjusted PAF estimates.

This will practically estimate (assuming a causal relationship) the percentage of all-cause dementia cases in the US that could be prevented if each risk factor were modified in midlife and late life.

In secondary analysis, we will also consider a) estimating PAFs in midlife and late life for dementia before age 75 (i.e., early dementia), and b) grouping together risk factors (e.g., vascular risk factors) and estimating PAFs.

**Sensitivity analyses**

We will perform several sensitivity analyses. First, we will potentially use competing risk proportional hazard models to examine associations and PAFs with non-dementia death and incident stroke as competing risks. Second, to evaluate the impact of potential collider bias we will re-run analyses after excluding those with prevalent strokes at Visit 1. Last, to examine disparities in dementia incidence among participants identifying as Black race we will stratify regression models by race.

**Limitations**

First, because smoking level, and specifically pack-years, are strongly related to depression (i.e., smoking cessation is sometimes treated with antidepressants), there could be residual confounding from our parameterization of smoking status. Although pack-years is available at Year 1, given patterns missing data over time, we are less confident in use of a pack-years variable at Visit 5 created for prior work by members of the authorship team (MP#2262). Similarly, there could be residual confounding attributable to our characterization of alcohol consumption.

Third, because depression is not a main outcome in ARIC there is limited data on depression medication, which can potentially affect cognition. And fourth, because coronary heart disease is
not centrally adjudicated we are unable to measure the PAF for coronary heart disease in late life.

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? ____ Yes  ___X___ 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”? ____ 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)?


    Racial and Ethnic Differences in the Population Burden of Dementia Attributable to Modifiable Risk Factors among Americans (Lee and Lutsey) (MP#3849)


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


