1.a. Full Title: Comparison of existing methods for algorithmic classification of dementia status

b. Abbreviated Title (Length 26 characters): Dementia classification

2. Writing Group:
   Writing group members (alphabetical):
   Kan Z. Gianattasio, MPP (first); Maria M. Glymour, MS, ScD; Melinda C. Power, ScD (last); Qiong Wu, PhD; others welcome

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

First author: Melinda C. Power
Address: Department of Epidemiology and Biostatistics
         Milken Institute School of Public Health
         George Washington University
         950 New Hampshire Ave, NW
         Washington, DC 20052
Phone: (202)994-7778
E-mail: power@gwu.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: Melinda C. Power
Address: Department of Epidemiology and Biostatistics
         Milken Institute School of Public Health
         George Washington University
         950 New Hampshire Ave, NW
         Washington, DC 20052
Phone: (202)994-7778
E-mail: power@gwu.edu

3. Timeline:
   Analyses will be completed within 1 year of receipt of data.

4. Rationale:
Dementia ascertainment is time-consuming and costly, making it difficult to implement in large, representative cohort studies. This fact hinders efforts to use the results from large population surveys to describe and monitor trends and disparities in the prevalence and incidence of cognitive impairment. Recognizing this, several groups of researchers have developed algorithms to use existing data from the large and nationally-representative Health and Retirement Study (HRS) to algorithmically classify dementia status in cohort participants.\textsuperscript{1–4} HRS provides an ideal setting for algorithm development. First, a strategically selected subset of participants from HRS were evaluated for dementia at four time points between 2001 and 2009 as part of the Aging, Demographics, and Memory Study (ADAMS).\textsuperscript{5,6} Thus, data from ADAMS provides gold-standard dementia diagnoses against which to train and evaluate algorithms. Second, as HRS is nationally-representative, algorithmic diagnoses can be (and have been) used in HRS to monitor trends in cognitive impairment and dementia at the national level,\textsuperscript{7–9} and to examine the associations of cognitive impairment and dementia with other health factors.\textsuperscript{10–15}

However, researchers hoping to use such algorithms in their studies face a difficult choice. Each algorithm was developed independently, often in the context of other objectives. Thus, reporting of performance metrics is inconsistent, and whether there are substantial differences in performance remains unknown. In addition, the algorithm with the best metrics when applied to the data used to create the algorithm may not be the best algorithm to apply widely because of overfitting. Few reports on the existing algorithms provide performance metrics achieved when the algorithm is applied to data aside from the training data. Moreover, the algorithm with the best overall performance metrics may be ill-suited to desired applications, particularly efforts to describe disparities, if performance differs across sub-populations. Currently, it is unclear whether there are big differences in the performance of existing algorithms across sub-groups defined by race/ethnicity, age, sex, or education. This limits the goal of addressing race/ethnic disparities in Alzheimer’s disease and related dementia (ADRD), which is a national priority codified in the 2011 U.S. National Alzheimer’s Project Act (NAPA).\textsuperscript{16} Recommendations from the resulting 2013 and 2016 ADRD workshops explicitly call for evaluation of under-diagnosis of ADRD and implementation of surveillance for monitoring disparities and trends in ADRD incidence among disparities populations.\textsuperscript{17,18}

Our goal is conduct a head-to-head comparison of performance for five existing algorithms for algorithmic classification of dementia, overall and by socio-demographic subgroups, to begin addressing the long-term goal of monitoring ADRD trends and disparities. We have completed our first stage of analysis, which was to compare overall performance metrics across algorithms when applied to: (a) training data comprising a sample of HRS participants commonly used to develop these algorithms, i.e. those who underwent dementia ascertainment at ADAMS Wave A, and (b) validation data comprising HRS participants who were outside the sample used to develop the algorithms, i.e. those who underwent dementia ascertainment at ADAMS Waves B, C, and D. In the training data, overall classification accuracy ranged from 80% - 87%, sensitivity ranged from 53% to 90%, and specificity ranged from 79% to 96% across the five algorithms. Though overall classification accuracy was similar in the validation data (range: 79% to 88%), sensitivity was much lower (range: 17% to 61%), while specificity was higher (range: 82% to 98%) compared to the training data. These differences suggest that overall, the algorithms are overfitted to the training data, and/or that the algorithms are better at identifying prevalent versus incident dementia. There are also differences in performance metrics by sociodemographic groups. For example, in both the training and validation data, sensitivity was generally lower, while specificity was higher, for non-Hispanic whites compared to both non-Hispanic blacks and Hispanics. Similarly, all algorithms uniformly had higher classification accuracy for individuals with higher education (high school or more) compared to those with less education, and had uniformly higher specificity and overall classification accuracy for those aged under 80 compared to older individuals.

As a next step, we would like to evaluate the performance metrics of the algorithms when applied to data from an unrelated study, which will provide further validation for how algorithmic dementia classification
accuracies compare overall and by socio-demographic subgroups. The Atherosclerosis Risk in Communities Study (ARIC) is an obvious choice, as it collects similar mental and physical functioning data as those used in the algorithms, is a bi-racial cohort, and importantly, includes in-person dementia ascertainment. External validation in the ARIC study will aid our evaluation of which algorithm may be preferred for what purpose, and furthermore provide insight for future algorithm development and refinement to achieve higher predictive accuracies, particularly among racial/ethnic minorities.

5. **Main Hypothesis/Study Questions:**

**Study Aim:**
Evaluate the predictive performance of existing algorithms developed to classify dementia status using Health and Retirement Survey (HRS) participants, when applied to an external validation dataset from ARIC.

**Hypothesis:**
We hypothesize that the performance of the algorithms in classifying dementia will be worse when applied to the ARIC data compared to the HRS data.

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**
Validation study, using data from Visit 5 and Visit 6 if/when available

**Exclusion criteria**
In all analyses, we will exclude participants missing data necessary to compute algorithmic diagnoses. We will also limit our analyses to black and white ARIC participants.

**Cognitive Outcomes**
*Prevalent and Incident Dementia:* We will classify persons as having dementia in Visit 5 (and Visit 6, if/when available) according to adjudicated diagnosis based on Visit 5 (and Visit 6, if/when available) cognitive assessment.

**Predictors/Covariates**
We will use variables that best correspond to those used in the existing dementia classification algorithms derived using HRS data. These include demographic information (age, gender, race, education), limitations in activities of daily living (e.g. dressing, eating, preparing own meals), cognitive test scores for items in the TICS and mini-mental state exam (e.g. dates recall, word recall, object naming) Informant Clinical Dementia Rating items (e.g. subject forgets names of friends/relatives, subject forgets recent events, subject ability to handle money).

**Statistical Analysis**
*Primary Analyses:* We will apply the existing algorithms to the ARIC data to evaluate each the accuracy of each algorithm in classifying participant dementia status in Visit 5 using common cognitive and physical functioning assessments. For the regression-based algorithms,1–3 we will compute the predicted probability of having dementia for each ARIC participant based on the published regression coefficients, and then assign algorithmic dementia classifications, considering a variety of classification thresholds. For the two
cognitive score cut off-based algorithms, we will simply use the corresponding cognitive test scores from ARIC to compute total cognitive score and apply the score cut-offs determined by the original authors to classify dementia status. We will compute the sensitivity, specificity, and overall accuracy of classifications determined by each algorithm, both overall and by socio-demographic sub-groups. We will repeat this analysis using Visit 6 data if/when it becomes available.

Sensitivity Analysis:
First, we may apply inverse probability weights to weight back to the original sample demographics prior to computing classification metrics, as ARIC was originally representative of the communities from which the participants were sampled.

Second, we may re-estimate each regression-based algorithm in the ARIC dataset, to understand sensitivity of algorithm development to differences in the data collected and data collection procedures between the HRS and the ARIC study. We will then re-calculate performance metrics based on these new versions of the existing algorithms to consider sensitivity to survey differences.

Third, contingent on Visit 6 data availability, we will construct an alternative dataset using Visit 6 excluding participants who had a dementia diagnosis in Visit 5 to understand the usefulness of each algorithm in identifying incident versus prevalent dementia cases.

Limitations/Challenges
The primary limitation of this study are the small differences in the data collected and the data collection procedures between the HRS (used for algorithm development) and the ARIC study. Should we find large disparities in performance metrics in the HRS validation/training datasets versus the ARIC dataset, it would be challenging to determine whether these disparities are a result of overfitting of the algorithms, or a result of underlying differences in the availability and measurement of the relevant predictors and/or the standards and procedures used for diagnosing dementia. Conversely, however, a finding of similar algorithm performance metrics when applied to the ARIC and HRS data, despite their differences, would provide support for the generalizability and applicability of these algorithms more broadly in dementia-related research.

7.a. Will the data be used for non-CVD analysis in this manuscript?  _X__ Yes  ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? _X__ Yes  ____ No
(This 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 file ICTDER03 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/search.php

     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)?

#2542 – Validation of a Medicare-claims based algorithm to identify frailty among older adults
#1700 – Cognitive function and incident dementia: The Atherosclerosis Risk in Communities Study

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

11.b. If yes, is the proposal

     ___ A. primarily the result of an ancillary study (list number* 2008.06__)
     ___ 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 at http://www.cscc.unc.edu/aric/forms/

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.

13. Per Data Use Agreement Addendum for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes ___X__ No.

References


