1.a. Full Title: When are representative samples necessary for generalizability? A comparison between ADNI and ARIC

b. Abbreviated Title (Length 26 characters): Sample representativeness

2. Writing Group:
Writing group members:

Kan Z. Gianattasio (first); Erin Bennett; Megha Mehrotra; M. Maria Glymour; Thomas Mosley; Rebecca Gottesman; Elizabeth Stuart; Melinda C. Power (last)

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

First author: Kan Gianattasio
Address: 950 New Hampshire Ave NW, 5th Floor, Washington DC 20052

Phone: 202-994-2572 Fax: 
E-mail: kzhang0316@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: 950 New Hampshire Ave NW, 5th Floor, Washington DC 20052

Phone: 202-994-7778 Fax: 
E-mail: power@gwu.edu

3. Timeline:
October – November 2020: data analysis  
December 2020: Finalize manuscript with co-authors

4. **Rationale:**

The Alzheimer’s Disease Neuroimaging Initiative (ADNI) is a longitudinal study that collects high quality imaging, biomarker, genetic, and clinical data from participants recruited from over 60 sites in the US and Canada.\(^1\) Its data is shared publicly, and has allowed researchers to make important progress on identifying biomarker changes associated with early stage Alzheimer’s disease and its disease progression and risk factors for these biomarker changes and clinical outcomes.\(^2,3\) However, because ADNI recruitment procedures resulted in a highly-selected sample that is predominantly white and disproportionately well-educated,\(^1,4\) certain types of estimated associations from ADNI may not be generalizable to the larger US population. It is therefore important to understand when ADNI findings are likely to be widely generalizable without adjustment, when adjustment for observed characteristics is sufficient, and when representative samples may be more necessary to make broad inferences.

The Atherosclerosis Risk in Communities (ARIC) study collects similar cognitive and imaging measures as similar to those in ADNI. Importantly, because ARIC participants were selected randomly from across four US communities, it is relatively more representative of the general population than is the ADNI sample. Thus, ARIC provides an opportunity against which ADNI can be evaluated for external validity. Specifically, comparing various associations (between sociodemographic factors, cognitive outcomes, and imaging outcomes) in ADNI vs. in ARIC will provide insight into which types of associations between variables can be directly generalized from ADNI to the four ARIC communities. In addition, we may explore whether weighting to a reference standard or use of other transportability methods allows for better generalization of ADNI to ARIC.

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

How do the associations between risk factors and imaging outcomes differ between ADNI and ARIC?

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:**

Cross-sectional analysis using ARIC Visit 5 data and ADNI baseline and screening data from the ADNI GO and ADNI2 phases.

**Inclusion/Exclusion:**
All primary analyses in ARIC will be limited to non-Hispanic white participants at Visit 5. We will expand the sample to include all non-Hispanic white and non-Hispanic black Visit 5 participants in secondary analyses.

For analyses involving PET amyloid-β data, all ARIC Visit 5 participants of the ARIC-PET Amyloid Imaging Study without dementia and with non-missing data on the independent variables listed below will be included.

For analyses involving MRI data, all ARIC Visit 5 participants with MRI and with non-missing data on independent variables listed below will be included.

Similar exclusion/inclusion criteria will be applied to the ADNI baseline visit data.

**Independent variables:**

- Age
- Gender
- Education
- Marital status
- Cognitive performance (MMSE, word fluency, animal fluency, Boston naming)
- Blood pressure
- History of hypertension
- Functional status
- Depression
- ApoE4

**Dependent variables:**

- amyloid-β
- AD signature region volume
- Frontal lobe volume
- Temporal lobe volume

**Hypothesized confounders**

Age, gender, education, cognitive status (normal, MCI, dementia)

**Statistical Analyses:**

We have confirmed close comparability of the variables above in ARIC and ADNI. We will pool the ARIC and ADNI datasets, and run unadjusted and adjusted linear and logistic regressions to examine associations between each of the independent variables with each of the dependent variables. Note that for the purposes of this paper, we are interested only in comparing estimated associations across the datasets, and not in estimating causal effects. All models will include an indicator for dataset (1=ADNI,
0=ARIC) and its interaction with the independent variable of interest; fully adjusted models will include age, gender and education. We will also examine associations between the MRI outcomes (AD signature region volume, frontal lobe volume, temporal lobe volume) and PET outcome (amyloid-β), similarly including an indicator for dataset and its interaction with the predictor of interest, as well as age, gender, and education in the fully adjusted model. Similarity and differences in findings will be summarized in tables and figures. Finally, in the case where there are substantial differences, we may explore transportability of ADNI using statistical techniques developed for this purpose. For example, the ADNI and ARIC samples will be weighted to a common reference sample, and we will re-run the analyses to evaluate whether any previous differences in estimated associations across the two datasets are mitigated.

Sensitivity analyses will incorporate ARIC MRI sampling weights.

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? ___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 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)?

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
  X  A. primarily the result of an ancillary study (list number* 1999.01, 2009.29 and 2017.01)
  ___  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.

References