RESEARCH ARTICLE

Generalizability of findings from a clinical sample to a community-based sample: A comparison of ADNI and ARIC

Kan Z. Gianattasio[1] | Erin E. Bennett[1] | Jingkai Wei[1] | Megha L. Mehrotra[2] | Thomas Mosley[3] | Rebecca F. Gottesman[4] | Dean F. Wong[5] | Elizabeth A. Stuart[6] | Michael E. Griswold[7] | David Couper[8] | M. Maria Glymour[2] | Melinda C. Power[1] | for the Alzheimer's Disease Neuroimaging Initiative[*]

[1] Department of Epidemiology, George Washington University, Washington, District of Columbia, USA

[2] Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, California, USA

[3] Department of Neurology, University of Mississippi Medical Center, Jackson, Mississippi, USA

[4] Departments of Neurology and Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA

[5] Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, Missouri, USA

[6] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[7] Department of Biostatistics, University of Mississippi Medical Center, Jackson, Mississippi, USA

[8] Department of Biostatistics, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence
Kan Z. Gianattasio, 950 New Hampshire Ave NW, 5th Floor, Washington, DC 20037.
E-mail: kzhang0316@gwu.edu

Abstract

Introduction: Clinic-based study samples, including the Alzheimer's Disease Neuroimaging Initiative (ADNI), offer rich data, but findings may not generalize to community-based settings. We compared associations in ADNI to those in the Atherosclerosis Risk in Communities (ARIC) study to assess generalizability across the two settings.

Methods: We estimated cohort-specific associations among risk factors, cognitive test scores, and neuroimaging outcomes to identify and quantify the extent of significant and substantively meaningful differences in associations between cohorts. We explored whether using more homogenous samples improved comparability in effect estimates.

Results: The proportion of associations that differed significantly between cohorts ranged from 27% to 34% across sample subsets. Many differences were substantively meaningful (e.g., odds ratios [OR] for apolipoprotein E ε4 on amyloid positivity in ARIC: OR = 2.8, in ADNI: OR = 8.6).

Discussion: A higher proportion of associations differed significantly and substantively than would be expected by chance. Findings in clinical samples should be confirmed in more representative samples.

KEYWORDS
Alzheimer's disease, Alzheimer's Disease Neuroimaging Initiative, Atherosclerosis Risk in Communities, dementia, generalizability

1 | INTRODUCTION

Generalizability is an oft-cited limitation of epidemiological studies, particularly when relying on clinic-based study samples.[1] The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an example of a clinical cohort that collects high-quality imaging, biomarker, genetic, and clinical data.[2] ADNI data are shared publicly, and have allowed researchers to make important progress in many areas, particularly around biomarkers associated with presence and progression of Alzheimer's disease (AD).[3,4] Because ADNI recruitment procedures were designed to obtain a sample emulating that of a clinical trial, the sample is generally healthy, predominantly White, and overwhelmingly well educated.[5,6] Researchers using ADNI typically recognize and acknowledge the potentially limited generalizability of their findings,[1,6-10] but the extent of this limitation remains unclear.[6] Given the richness and high quality of data from clinical cohorts, it is important to understand when findings in these samples are likely to generalize to target populations, and when representative sampling is necessary to make broader inferences.

The Atherosclerosis Risk in Communities (ARIC) study collects cognitive and imaging measures similar to those in ADNI. Importantly, ARIC is a community-based cohort, representing another target population of interest. Our goal was to compare associations among risk factors, cognitive outcomes, and neuroimaging outcomes in ADNI and in ARIC to examine the degree to which associations differ, and whether particular types of associations are more or less comparable. This may help identify whether and when associations estimated in clinic-based samples such as ADNI may generalize more broadly to a community-based setting, using ARIC as an example, in the context of dementia research.

2 | METHODS

2.1 | Study samples and eligibility criteria

2.1.1 | ADNI

ADNI (adni.loni.usc.edu) is a longitudinal study of adults aged 55 to 90 launched in 2003 as a public–private partnership. The primary goal of ADNI is to test whether magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

ADNI has enrolled approximately 1800 individuals to date across four phases: ADNI-1 (2004–2010), ADNI-GO (2009–2011), ADNI-2 (2011–2016), and ADNI-3 (2016–present). Participants were recruited from approximately 60 sites across the United States and Canada to meet a target number of persons who were cognitively normal, had MCI, or had mild AD. Recruitment was restricted to those with Hachinski Ischemic Score[11] ≤ 4 and Geriatric Depression Scale (GDS)[12] score ≤ 6, were in general good health, did not use certain medications, and were able and willing to meet study requirements.[5]

RESEARCH IN CONTEXT

1. **Systematic Research**: Generalizability is an oft-cited limitation of epidemiological studies, particularly when relying on clinic-based samples. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an example of a clinical cohort used for dementia research. Researchers using ADNI acknowledge the potentially limited generalizability of their findings, though no study thus far has examined the potential extent of this limitation.

2. **Interpretation**: We compared associations among risk factors, cognitive test scores, and neuroimaging outcomes across ADNI and the Atherosclerosis Risk in Communities Study, a community-based sample. We found that a substantial proportion of estimated associations differed significantly, and differences were often substantively meaningful. Findings from ADNI and similar, highly selected clinical cohorts cannot be assumed to be directly generalizable to other populations.

3. **Future Directions**: Findings in clinical samples such as ADNI should be confirmed in more representative samples. Future efforts to evaluate whether use of transportability algorithms allows generalizability are warranted.

HIGHLIGHTS

- Findings from clinic-based cohorts may not generalize to the community.
- ADNI is an example of such a cohort used broadly in dementia research.
- We compared associations estimated in ADNI and ARIC, a community-based cohort.
- A substantial proportion of estimated associations differed significantly.
- Findings in clinical samples should be confirmed in more representative samples.

After an initial screening visit, eligible participants return for a baseline visit and follow-up visits every 6 to 12 months.

For this cross-sectional analysis, we used data from the screening and baseline visits from any ADNI phase. Demographic data, apolipoprotein E (APOE) status, blood cholesterol, and triglyceride levels were collected during screening. Functional status and vital signs were collected at baseline. The Boston naming test[13] and animal naming test[14,15] were first administered at screening, and the Mini-Mental State Examination (MMSE)[16] and the word fluency test[14,15,17] were

first administered at baseline (word fluency was not administered in ADNI-1 and Boston naming was not administered in ADNI-3). PET scans using florbetapir AV-45 were administered at baseline for participants recruited in ADNI-GO, ADNI-2, or ADNI-3. 3 Tesla (3T) MRI exams were conducted for a subset of ADNI-1 participants and for all participants from ADNI-GO, ADNI-2, and ADNI-3 at either screening or baseline. Finally, participants were diagnosed at baseline with normal cognition, MCI, AD, or other dementia etiology by a study physician, based on a full review of participants' medical histories, Clinical Dementia Rating (CDR), GDS, Functional Activity Questionnaire (FAQ) scores, laboratory test results, and neuropsychological tests.[18]

2.1.2 | ARIC

The ARIC cohort recruited 15,792 participants aged 45 to 64 at baseline from four US communities between 1987 and 1989. In Forsyth County, North Carolina, White and Black participants were selected using the 1980 census and a two-stage area probability sampling approach, stratified on race and income.[19] In Jackson, Mississippi, investigators randomly selected eligible Black residents from the Mississippi Highway Patrol database.[19] In Minneapolis, Minnesota, investigators used the Hennepin County jury selection database to draw potential participants using a simple random sampling approach.[19] In Washington County, Maryland, participants were randomly sampled from a 1975 health census and Department of Motor Vehicles database.[19]

In these cross-sectional analyses, we consider data from participants of ARIC Visit 5 (V5; 2011–2013), which was attended by 6538 participants (5765 participants had died, and 3489 persons were unable or unwilling to attend V5; Table A.1 in supporting information). We define sociodemographic characteristics using data from ARIC baseline, and define all other variables of interest using data collected at V5 or the most recent annual follow-up call preceding V5. Cognitive status (normal, MCI, or dementia) was ascertained at V5 by first applying an algorithm using scores from cognitive and functional assessments, followed by final diagnosis by ARIC physicians and neuropsychologists (Table A.2 in supporting information).[20] All participants with evidence of cognitive impairment at V5, those with a prior ARIC brain scan, and a random sample of cognitively normal participants were invited for 3T MRI scans. Ultimately, 1978 participants completed brain MRI; sampling weights allow weighting of the 3T MRI sample back to the full V5 sample. Of those with V5 3T MRI scans, 345 participants from Jackson, Washington County, and Forsyth County without renal failure, heavy alcohol use, or dementia completed PET scans.

2.1.3 | Eligibility criteria

We restricted both samples to Black and White persons aged 65+ given the small number of participants of other race in these samples and lack of persons under age 65 in the ARIC sample. We also excluded persons missing demographic data or dementia status, and ADNI participants with screening and baseline visit dates more than 12 months apart.

2.2 | Risk factors

Sociodemographic predictors included age, sex, self-reported race (White vs. Black), education (up to high school/ General Education Development (GED) vs. > high school/GED), and marital status (married vs. widowed/divorced/separated/never married). We also considered history of hypertension, measured systolic and diastolic blood pressure, blood cholesterol and triglycerides, APOE ε4 allele status (yes/no), and functional status as potential risk factors of interest. Because ADNI and ARIC used slightly different assessment of functional status,[21] we created two functional status indicators (any vs. no functional limitation): one using four items worded identically or close to identically (FAQ4), and one using an additional five items worded similarly across studies (FAQ9).

2.3 | Cognitive performance

For cognitive outcomes, we focused on tests administered in a consistent manner to both ARIC and ADNI participants: animal naming (60 seconds), word fluency (letter F; 60 seconds), the MMSE, and the Boston naming test (30 items). We computed z-scores for animal naming and word fluency scores using the means and standard deviations from the pooled ADNI and ARIC data, and dichotomized MMSE scores (< 25) and Boston naming (< 25) scores based on commonly used thresholds.[15]

2.4 | Neuroimaging

We confirmed ADNI and ARIC PET scans used the same ligand (florbetapirAV-45[22]) and were conducted similarly.[23–25] Each study published global florbetapir standardized uptake value ratios (SUVRs) based on cortical regions of interest (ROIs). In ADNI, these included parts of the frontal, anterior/posterior cingulate, lateral parietal, lateral temporal, and precuneus regions.[26] ARIC similarly used parts of the frontal, anterior/posterior cingulate, parietal, lateral temporal, and precuneus regions, and in addition included the occipital region.[23] These values were then standardized to a reference region (whole cerebellum in ADNI; cerebellum gray in ARIC). Following convention, we considered only dichotomized SUVR.[23,27–30] We used study-specific cutoffs (1.11 in ADNI, and 1.20 in ARIC)[23,26] because the PET protocols are not identical, and importantly, because study-specific cutoffs are used to define PET positivity in the literature. For MRI outcomes, we confirmed both studies used 3T MRI and processed imaging data using FreeSurfer software.[31] We used three relevant composite MRI imaging outcomes (AD signature ROI cortical region volume,[31] frontal lobe volume, and temporal lobe volume). These outcomes were provided by

ARIC and constructed from raw data for ADNI. All volumetric measures were divided by intracranial volume prior to use in analyses.

Details of between-cohort harmonization of measurement are available in Table A.2.

2.5 | Statistical analyses

We pooled ADNI and ARIC data into a single dataset and ran logistic and linear models estimating (1) associations of risk factors with cognitive testing and PET/MRI outcomes, (2) associations of PET/MRI outcomes with cognitive testing outcomes, and (3) associations of PET outcomes with MRI outcomes. Each model included an indicator for cohort and its interaction with the predictor of interest, allowing for estimation of cohort-specific associations and statistical evaluation of between-cohort differences. Models adjusted for age, sex, education, race, and dementia status (dementia/no dementia), as well as their interactions with cohort to account for potential differences in confounding structure.

To understand how findings changed with increasing homogeneity, we repeated analyses in four subgroups defined by race and cognitive status: (1) all White and Black participants, (2) all White participants, (3) White and Black participants with normal cognition or MCI, (4) white participants with normal cognition or MCI. Because PET scans were not administered for ARIC participants with dementia, analyses involving amyloid beta (Aβ) were limited to subgroups (3) and (4).

Next, to focus on associations that would be considered "positive findings" in at least one cohort, we restricted comparisons to associations that had $P < 0.2$, and more stringently to $P < 0.05$, in at least one cohort. In addition to null hypothesis testing, we also computed the absolute difference in point estimates as a percentage relative to the ARIC point estimate, and we created forest plots of cohort-specific findings to examine differences qualitatively. Finally, for associations that differed significantly between cohorts, we compared estimated cohort-specific marginal expectations to better understand what was driving differences. Because our purpose was to illustrate the potential extent of the differences we may expect to find across the literature, we did not adjust for multiple comparisons.

2.5.1 | Sensitivity analyses

To examine the sensitivity of findings to changes in sample selection, we re-ran all analyses limiting the sample to individuals with complete data across all predictors, cognitive test outcomes, and either MRI outcomes ("common MRI sample"), or PET imaging outcomes ("common PET sample"). We also re-ran all analyses involving the common MRI sample using ARIC MRI sampling weights designed to weight back to the full sample of ARIC V5 participants. Finally, we considered analyses including cohort interactions with only the predictor of interest.

ARIC and ADNI participants provided consent at data collection, and procedures were approved by each study site's institutional review boards. All analyses were conducted in SAS (v.9.4), R (v.4.0.0), and STATA (v.15).

3 | RESULTS

Our eligible sample included 1787 participants from ADNI and 6445 participants from ARIC (Table 1). Differences in sample characteristics across ADNI and ARIC were statistically significant for all except for the proportions of participants scoring < 25 on the MMSE and PET-positive participants. Participants from ADNI were more likely to be male, highly educated, and APOE ε4 positive. Remarkably, though more ADNI participants have MCI or dementia (60% vs. 27%), ADNI participants performed slightly better on cognitive testing (Table 1). These patterns were largely consistent across subsamples used in primary and sensitivity analyses.

Not all participants had complete data on our predictors and outcomes of interest; thus, analytical sample sizes ranged from 493 to 1786 in ADNI and 166 to 6435 in ARIC, depending on the specific subsample and predictor–outcome combination examined (Table 2). In analyses considering all available observations for White and Black participants, 34% of all considered associations were significantly different between ARIC and ADNI (Table 2, Figures A.1a and A.1b in supporting information). This proportion increased to 42% after restricting to associations with P-value < 0.2 in at least one cohort, and to 50% after restricting to associations with P-value < 0.05 in at least one cohort (Table 2, Figure 1). These proportions were lower when analyses were restricted to more homogenous subsamples. However, even after restricting to White participants with normal cognition or MCI—our most stringent restriction—the overall proportion of associations that were statistically different across cohorts was 30% overall (Table 2, Figures A.1a and A.1b); and 44% after restricting to associations with P-value < 0.05 in at least one cohort (Table 2, Figure 1). Differences between cohorts were not clustered in particular risk factor–outcome association types (Figures A.2a and A.2b in supporting information).

In primary analyses across subsamples defined by race and cognitive status, the proportion of associations with point estimates differing by > 50% between ADNI and ARIC ranged from 50% to 55% of all associations evaluated, with no evidence of reduced differences in subsamples with more homogenous composition (Table 2, Figures A.3a and A.3b in supporting information). These proportions were similar when restricting to associations with P-value < 0.2 in at least one cohort (range 50%–58%) and with P-value < 0.05 in at least one cohort (range 48%–56%; Table 2, Figure 2). Again, differences in point estimates between cohorts were not clustered within any particular risk factor–outcome association types (Figures A.4a and A4b in supporting information).

Many of the differences in estimated associations were substantively meaningful (Figures 3–4, Figure A.5 in supporting information). For example, the association between having any functional limitation (FAQ9 $\geq$1) and standardized animal naming score is –0.49 in ADNI, compared to –0.17 in ARIC (Figure 3), with differences driven by higher scores among those without a limitation in ADNI than in ARIC

TABLE 1 Sample characteristics across all participants with non-missing data on age, sex, education, race, and dementia status

Variable	ADNI		ARIC	
	N	Mean (SD) / N (%)	N	Mean (SD) / N (%)
Predictors				
Age in years	1787	75.1 (6.0)	6445	75.8 (5.3)
Male	1787	983 (55%)	6445	2649 (41.1%)
Education: greater than HS/GED	1787	1516 (84.8%)	6445	3327 (51.6%)
Black	1787	72 (4.0%)	6445	1527 (23.7%)
Married	1780	1347 (75.7%)	6445	3954 (61.3%)
Cognitive status				
Normal	1787	712 (39.8%)	6445	4734 (73.5%)
MCI	1787	762 (42.6%)	6445	1371 (21.3%)
Dementia	1787	313 (17.5%)	6445	340 (5.3%)
FAQ score ≥1 (out of 9 items)	1776	871 (49.0%)	2476	1700 (68.7%)
FAQ score ≥1 (out of 4 items)	1780	551 (31.0%)	2578	514 (19.9%)
Serum cholesterol (mg/dL)	1331	192.8 (39.9)	6343	181.5 (42.0)
Triglycerides (mg/dL)	1321	139.0 (74.6)	6329	124.9 (60.1)
Hypertension history (ARIC up to visit 5)	1787	861 (48.2%)	6445	4937 (76.6%)
Systolic blood pressure (mm Hg)	1764	133.6 (16.8)	6413	130.6 (18.6)
Diastolic blood pressure (mm Hg)	1764	73.8 (9.2)	6413	66.3 (10.8)
At least one APOE ε4 allele	1716	772 (45%)	6236	1813 (29.1%)
Cognitive test outcomes				
MMSE score < 25	1785	265 (14.8%)	6435	957 (14.9%)
Boston naming score < 25	1451	369 (25.4%)	6196	2089 (33.7%)
Animal naming score	1786	17.7 (6.1)	6343	16.0 (5.1)
Word fluency score	1055	13.7 (4.8)	6251	11.3 (4.5)
MRI outcomes				
AD signature region volume, standardized by intracranial volume	1025	0.04 (0.0)	1954	0.04 (0.0)
Frontal lobe volume, standardized by intracranial volume	1028	0.11 (0.01)	1954	0.11 (0.01)
Temporal lobe volume, standardized by intracranial volume	1026	0.07 (0.01)	1954	0.07 (0.01)
PET imaging outcomes				
PET positive (SUVR > 1.11 in ADNI, and SUVR > 1.2 in ARIC) [a]	832	445 (53.5%)	344	176 (51.2%)

[a] ADNI cortical regions of interest include frontal, anterior/posterior cingulate, lateral parietal and lateral temporal regions, standardized by whole cerebellum reference region; ARIC cortical regions of interest include frontal, anterior/posterior cingulate, lateral parietal, lateral temporal, occipital and precuneus regions, standardized by cerebellum gray.

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein E; ARIC, Atherosclerosis Risk in Communities; FAQ, Functional Activities Questionnaire; HS, high school; GED, General Education Development; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PET, positron emission tomography; SD, Standard deviation; SUVR, standardized uptake value ratio.

(Figure A.5a). Conversely, the association between a 10-year-older difference in age and standardized AD signature ROI cortical region volume is –0.72 in ARIC, but just –0.42 in ADNI (Figure 3), with differences driven by larger volumes among younger individuals in ADNI versus in ARIC (Figure A.5e). Notably, having an APOE ε4 allele was associated with 8.6 times the odds of being PET positive in ADNI, compared to 2.79 times the odds in ARIC (Figure 4), driven by higher probability of PET positivity among APOE ε4 carriers in ADNI than in ARIC (0.78 vs. 0.58; Figure A.5 g). In some cases, estimated associations were significant in only one cohort. For example, APOE ε4 status was significantly

associated with all four cognitive test outcomes in ADNI, but not in ARIC (Figures 3–4). Finally, in rare instances, the direction of estimated associations was inconsistent between ADNI and ARIC (Figures 3–4; Figures A.5b, A.5d, A.5f).

Sensitivity analyses were consistent with our overall findings. Patterns in statistical differences were similar in analyses using the weighted and unweighted common MRI samples (Table A.3.a and A.3.b in supporting information), and analyses excluding cohort interactions with covariates (Tables A.4-A.5 in supporting information). In the common PET sample, proportions of significantly different associations

TABLE 2 Proportion of associations that differ between ADNI and ARIC[a] in analyses using all available observations for each association

Sample (Ns)	All available associations		Associations with $P < 0.2$ in at least one cohort		Associations with $P < 0.05$ in at least one cohort	
	Number of associations evaluated	Number (%) different associations	Number of associations evaluated	Number (%) different associations	Number of associations evaluated	Number (%) different associations
Proportion of associations that differ statistically between ADNI and ARIC ($P < 0.05$)						
White & Black, all cognition (ADNI $N = 630–1786$, ARIC $N = 1517–6435$)[b]	103	35 (34.0%)	84	35 (41.7%)	64	32 (50.0%)
White, all cognition (ADNI $N = 600–1714$, ARIC $N = 1179–4910$)[b,c]	96	26 (27.1%)	73	26 (35.6%)	56	23 (41.1%)
White & Black, normal/MCI (ADNI $N = 515–1474$, ARIC $N = 255–6069$)	123	39 (31.7%)	104	39 (37.5%)	82	38 (46.3%)
White, normal/MCI (ADNI $N = 493–1414$, ARIC $N = 166–4711$)[c]	115	34 (29.6%)	90	34 (37.8%)	73	32 (43.8%)
Proportion of associations that differ qualitatively between ADNI and ARIC (> 50% difference in point estimates)						
White & Black, all cognition (ADNI $N = 630–1786$, ARIC $N = 1517–6435$)[b]	103	51 (49.5%)	84	42 (50.0%)	64	31 (48.4%)
White, all cognition (ADNI $N = 600–1714$, ARIC $N = 1179–4910$)[b,c]	96	50 (52.1%)	73	41 (56.2%)	56	28 (50.0%)
White & Black, normal/MCI (ADNI $N = 515–1474$, ARIC $N = 255–6069$)	123	68 (55.3%)	104	57 (54.8%)	82	46 (56.1%)
White, normal/MCI (ADNI $N = 493–1414$, ARIC $N = 166–4711$)[c]	115	61 (53%)	90	52 (57.8%)	73	39 (53.4%)

[a]Models adjusted for age, sex, education, race (where applicable), and dementia status (where applicable), as well as cohort interactions with each covariate.
[b]Excludes associations involving PET outcomes.
[c]Excludes associations involving race as a predictor.
Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; ARIC, Atherosclerosis Risk in Communities; MCI, mild cognitive impairment; PET, positron emission tomography.
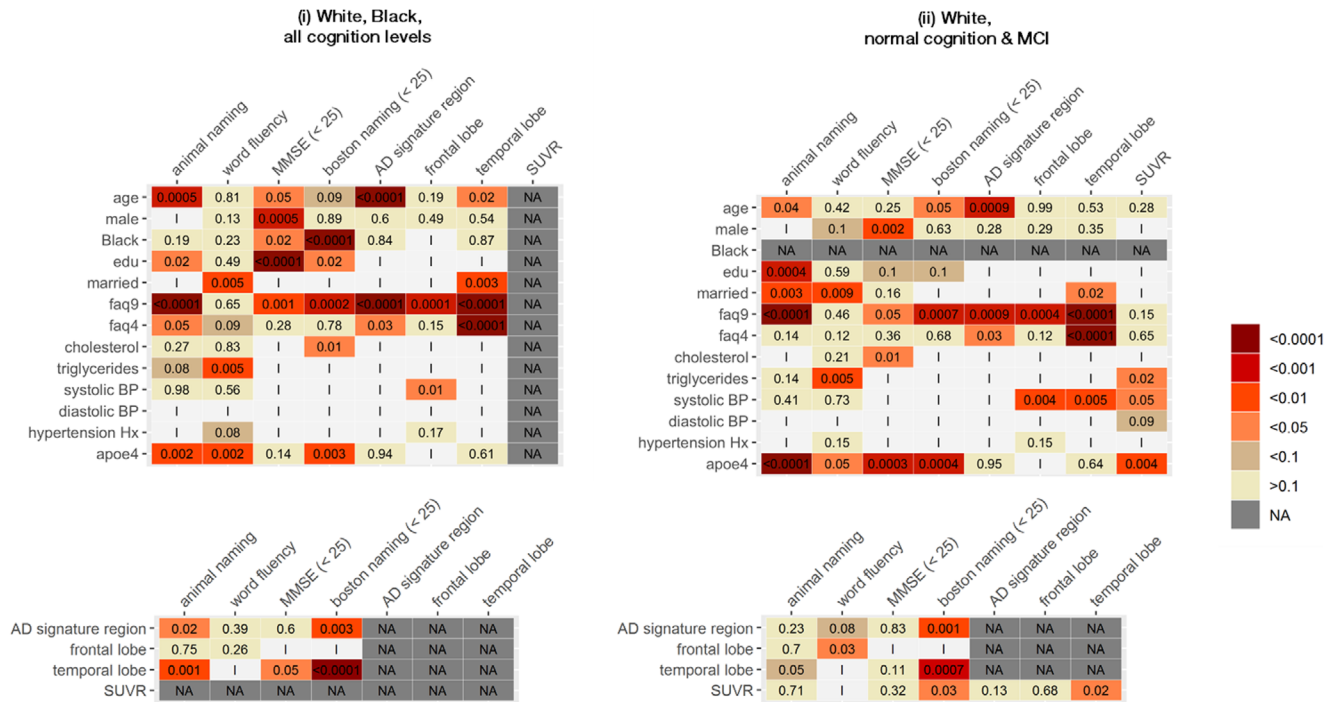
were lower and closer to what would be expected by chance, though this is likely due to loss of power from smaller sample sizes (ADNI $n = 485–507$, ARIC $n = 162–248$; Table A.3.c). The proportion of associations with point estimate differences exceeding 50% were marginally higher in analyses using the common MRI sample (Tables A.3.a and A.3.b), and similar in analyses using the common PET sample (Table A.3.c), as well as in analyses excluding cohort interactions with covariates (Tables A.4-A.5).

4 | DISCUSSION

In this study assessing the generalizability of associations across ADNI and ARIC, we found that the proportion of associations that dif-

fer significantly and the proportion of associations that differ qualitatively across the two samples are substantial, much higher than would be expected by chance. Furthermore, many of these differences were substantively meaningful. Statistically significant differences were observed less frequently in samples of more homogenous participants defined by race and/or cognitive status, though differences remained higher than would be expected by chance alone. There were no apparent patterns in risk factors or outcomes for which associations were more or less likely to differ between cohorts, thus we are unable to make recommendations on which types of associations may be assumed to be generalizable with confidence.

Our study is the first, to our knowledge, to directly examine the extent to which findings observed in a clinical sample are comparable to those from a community-based cohort to better understand whether

(i) White, Black, all cognition levels

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe	SUVR
age	0.0005	0.81	0.05	0.09	<0.0001	0.19	0.02	NA
male	I	0.13	0.89	0.6	0.49	0.49	0.54	NA
Black	0.19	0.23	0.02	<0.0001	0.84	I	0.87	NA
edu	0.02	0.49	<0.0001	0.02	I	I	I	NA
married	I	0.005	I	I	I	I	0.003	NA
faq9	<0.0001	0.65	0.001	0.0002	<0.0001	0.0001	<0.0001	NA
faq4	0.05	0.09	0.28	0.78	0.03	0.15	<0.0001	NA
cholesterol	0.27	0.83	I	0.01	I	I	I	NA
triglycerides	0.08	0.005	I	I	I	I	I	NA
systolic BP	0.98	0.56	I	I	I	0.01	I	NA
diastolic BP	I	I	I	I	I	I	I	NA
hypertension Hx	I	0.08	I	I	I	0.17	I	NA
apoe4	0.002	0.002	0.14	0.003	0.94	I	0.61	NA

(ii) White, normal cognition & MCI

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe	SUVR
age	0.04	0.42	0.25	0.05	0.0009	0.99	0.53	0.28
male	I	0.1	0.002	0.63	0.28	0.29	0.35	I
Black	NA	NA	NA	NA	NA	NA	NA	NA
edu	0.0004	0.59	0.1	0.1	I	I	I	I
married	0.003	0.009	0.16	I	I	I	0.02	I
faq9	<0.0001	0.46	0.05	0.0007	0.0009	0.0004	<0.0001	0.15
faq4	0.14	0.12	0.36	0.68	0.03	0.12	<0.0001	0.65
cholesterol	I	0.21	0.01	I	I	I	I	I
triglycerides	0.14	0.005	I	I	I	I	I	0.02
systolic BP	0.41	0.73	I	I	I	0.004	0.005	0.05
diastolic BP	I	I	I	I	I	I	I	0.09
hypertension Hx	I	0.15	I	I	I	0.15	I	I
apoe4	<0.0001	0.05	0.0003	0.0004	0.95	I	0.64	0.004

Legend: <0.0001; <0.001; <0.01; <0.05; <0.1; >0.1; NA

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe
AD signature region	0.02	0.39	0.6	0.003	NA	NA	NA
frontal lobe	0.75	0.26	I	I	NA	NA	NA
temporal lobe	0.001	I	0.05	<0.0001	NA	NA	NA
SUVR	NA	NA	NA	NA	NA	NA	NA

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe
AD signature region	0.23	0.08	0.83	0.001	NA	NA	NA
frontal lobe	0.7	0.03	I	I	NA	NA	NA
temporal lobe	0.05	I	0.11	0.0007	NA	NA	NA
SUVR	0.71	I	0.32	0.03	0.13	0.68	0.02

NA: Association not applicable to sample
I: Association not significant to p<0.05 in either cohort

FIGURE 1 Heatmap of *P*-values for statistical significance in differences across ADNI and ARIC estimated associations between risk factors and cognitive test/imaging outcomes, and between imaging outcomes and cognitive test outcomes; limited to associations with *P* < 0.05 statistical significance independently in at least one study. AD, Alzheimer's disease; APOE, apolipoprotein E ε4 positive; BP, blood pressure; FAQ, Functional Activities Questionnaire; MMSE, Mini-Mental State Examination; SUVR, standardized uptake value ratio

clinical sample-based findings are generalizable to other populations. The necessity for this work has long been recognized given ADNI's ongoing contributions to research on relationships between cognitive measures, neuropsychiatric symptoms, biomarkers, risk factors, and AD pathology and progression,[4,6] and facilitation of efforts around early AD detection, improving diagnostic classification, monitoring of disease progression, and improving clinical trial efficiency.[3,4,32] However, lack of generalizability of ADNI findings to other populations may critically limit the applicability of findings from ADNI. For example, given ADNI's generally healthy and highly educated population, certain findings from ADNI may be less relevant to more typical older adults, given evidence on the association between comorbidity burden and faster cognitive decline,[33–35] and potential modification of the association between biomarkers and cognitive decline by education.[36–39] Thus, while ADNI and similar studies remain an invaluable resource for AD research, findings should not be assumed to be directly applicable to other target populations. Efforts to confirm findings from ADNI and other clinic-based samples across different, less-selected population-based studies are warranted.[1,6,40]
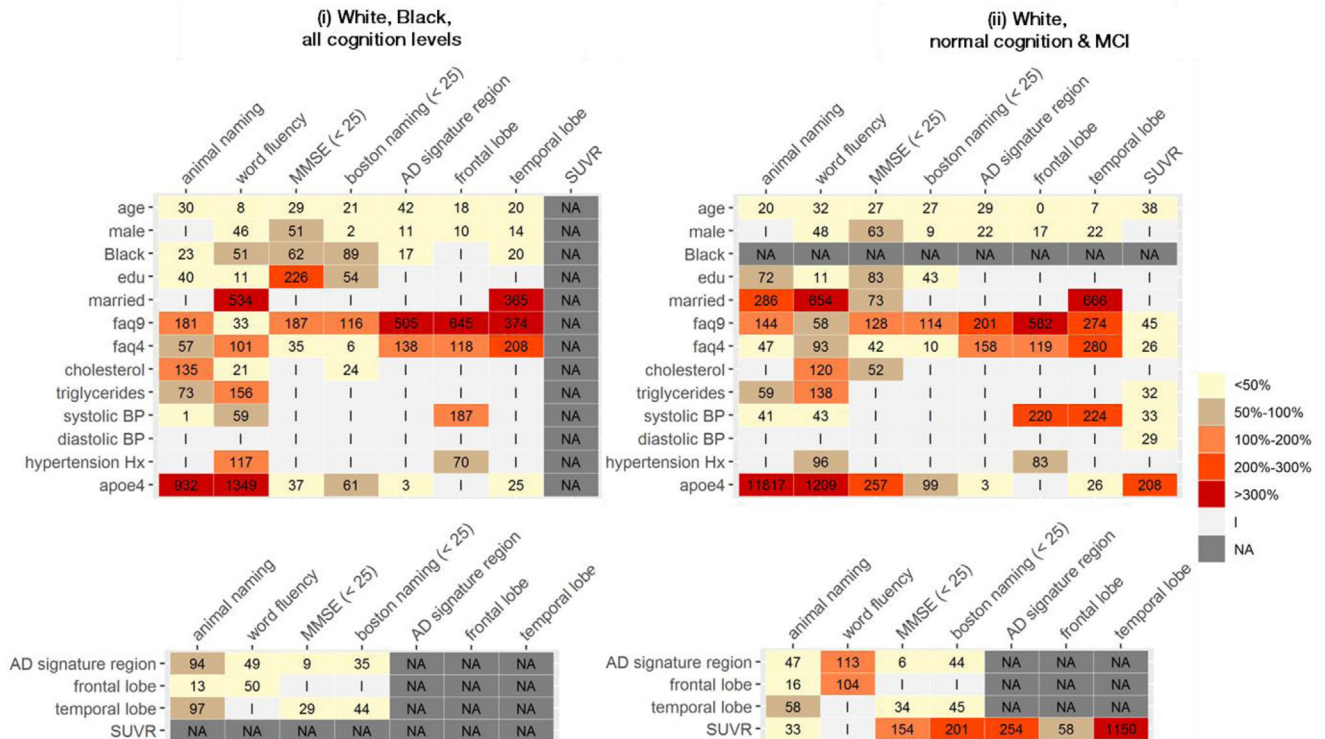
This study contributes to the long-standing debate around the necessity and value of recruiting representative samples for research. Many scholars would agree that the value of representative sampling depends on the research goals.[40-43] Specifically, the prevailing consensus is that while representative sampling is necessary for conducting descriptive studies[42,43] or for informing health-care planning,[40] it should not generally be required for studies estimating associations between variables or estimating causal effects.[42–44] Our findings suggest that inferences based on highly selected clinical samples may not be relevant to the broader population, even if unbiased.

As we were careful to include only variables measured in comparable ways across studies, differences in estimated associations are likely due to imbalanced distribution of characteristics that modify associations of interest across cohorts. In this case, if we can identify and measure all characteristics that modify the association of interest and differ between ADNI and the target population, we may be able to apply transport estimators to produce unbiased estimates of associations in the target population.[45–50] However, the assumptions required for transport are strong and often untestable, and existence of unknown or unmeasured modifiers, as well as positivity violations in the distribution of modifiers, would preclude use of transport algorithms for inference. We plan to examine the utility of transport algorithms in future work.

Our study has strengths. We were deliberate in our selection of risk factors, cognitive test outcomes, and imaging outcomes to identify those measured similarly, or which we felt could be harmonized across cohorts (Table A.2). Additionally, we conducted various sensitivity analyses to ensure that our findings are robust.

Our work also has limitations. First, the subsample of ARIC participants that underwent PET imaging was small (*N* = 345) and excluded those with dementia; thus, we were unable to assess these

(i) White, Black, all cognition levels

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe	SUVR
age	30	8	29	21	42	18	20	NA
male	I	46	51	2	11	10	14	NA
Black	23	51	62	89	17	I	20	NA
edu	40	11	226	54	I	I	I	NA
married	I	534	I	I	I	I	365	NA
faq9	181	33	187	116	505	645	374	NA
faq4	57	101	35	6	138	118	208	NA
cholesterol	135	21	I	24	I	I	I	NA
triglycerides	73	156	I	I	I	I	I	NA
systolic BP	1	59	I	I	I	187	I	NA
diastolic BP	I	I	I	I	I	I	I	NA
hypertension Hx	I	117	I	I	I	70	I	NA
apoe4	932	1349	37	61	3	I	25	NA

(ii) White, normal cognition & MCI

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe	SUVR
age	20	32	27	27	29	0	7	38
male	I	48	63	9	22	17	22	I
Black	NA	NA	NA	NA	NA	NA	NA	NA
edu	72	11	83	43	I	I	I	NA
married	286	654	73	I	I	I	666	I
faq9	144	58	128	114	201	582	274	45
faq4	47	93	42	10	158	119	280	26
cholesterol	I	120	52	I	I	I	I	I
triglycerides	59	138	I	I	I	I	I	32
systolic BP	41	43	I	I	I	220	224	33
diastolic BP	I	I	I	I	I	I	I	29
hypertension Hx	I	96	I	I	I	83	I	I
apoe4	11817	1209	257	99	3	I	26	208

Legend: <50% | 50%-100% | 100%-200% | 200%-300% | >300% | I | NA

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe
AD signature region	94	49	9	35	NA	NA	NA
frontal lobe	13	50	I	I	NA	NA	NA
temporal lobe	97	I	29	44	NA	NA	NA
SUVR	NA	NA	NA	NA	NA	NA	NA

	animal naming	word fluency	MMSE (< 25)	boston naming (< 25)	AD signature region	frontal lobe	temporal lobe
AD signature region	47	113	6	44	NA	NA	NA
frontal lobe	16	104	I	I	NA	NA	NA
temporal lobe	58	I	34	45	NA	NA	NA
SUVR	33	I	154	201	254	58	1150

[a] Percentage difference relative to ARIC estimates, computed as |ARIC estimate − ADNI estimate|/|ARIC estimate|
NA: Association not applicable to sample
I: Association not significant to p<0.05 in either cohort

FIGURE 2 Heatmap of percentage differences in point estimates of associations[a] between risk factors and cognitive test/imaging outcomes, and between imaging outcomes and cognitive test outcomes; limited to associations with $P < 0.05$ statistical significance independently in at least one study. AD, Alzheimer's disease; APOE, apolipoprotein E ε4 positive; BP, blood pressure; FAQ, Functional Activities Questionnaire; MMSE, Mini-Mental State Examination; SUVR, standardized uptake value ratio

associations across the full spectrum of cognitive status and had low power to detect interactions. Second, we assessed the comparability of ADNI findings to only one other cohort, which itself is not generalizable to the national population: though ARIC participants were initially recruited using random sampling, they were recruited from just four US communities from the late 1980s. However, the baseline characteristics of the overall V5 sample and the V5 MRI and PET subsamples are reasonably similar to those of the full ARIC cohort, suggesting they remain relatively representative of the original communities from which they were sampled (Table A.1). ARIC has the rare advantage among community-based studies of having outcomes data comparable to ADNI. Without considering other comparable datasets, we are unable to make broad conclusions about generalizability of specific associations; researchers should test the comparability of associations of interest in other cohorts where comparable data are available. Third, the ADNI sample is predominantly White (96%), and most Black ARIC participants were recruited from Jackson or Forsyth County. This conflates the effects of race and place, and raises questions about whether associations with race have comparable interpretation across samples. Furthermore, due to the small sample of Black participants in ADNI, we were unable to examine whether and how between-cohort differences may differ by race. Fourth, we cannot rule out differences due to imperfect harmonization, particularly in the Boston naming test, history of hypertension, FAQ9, and PET SUVR (Table A.2). Notably, there is

no universal consensus about how summary/global cortical PET SUVR measures should be constructed; the differences between ADNI and ARIC reflect the variation in PET SUVR constructs that we may see elsewhere. As such, evaluation of between-cohort differences using the PET variables as originally constructed and made available for use in research by ADNI and ARIC is useful and valuable. Finally, because our goal here was only to assess the comparability of various cross-sectional associations, we do not recommend making any substantive inferences based on the sample of estimated associations presented here.

In summary, we found that a substantial proportion of estimated associations among predictors, cognitive performance, and neuroimaging outcomes differed between the ADNI and ARIC cohorts. Thus, findings from ADNI and similar highly selected clinical cohorts should not be assumed to be directly generalizable to other populations and confirmation of findings in clinical samples such as ADNI in more representative samples is necessary. Future efforts should evaluate the use of transportability algorithms to infer associations from ADNI or other clinical samples to different target populations.
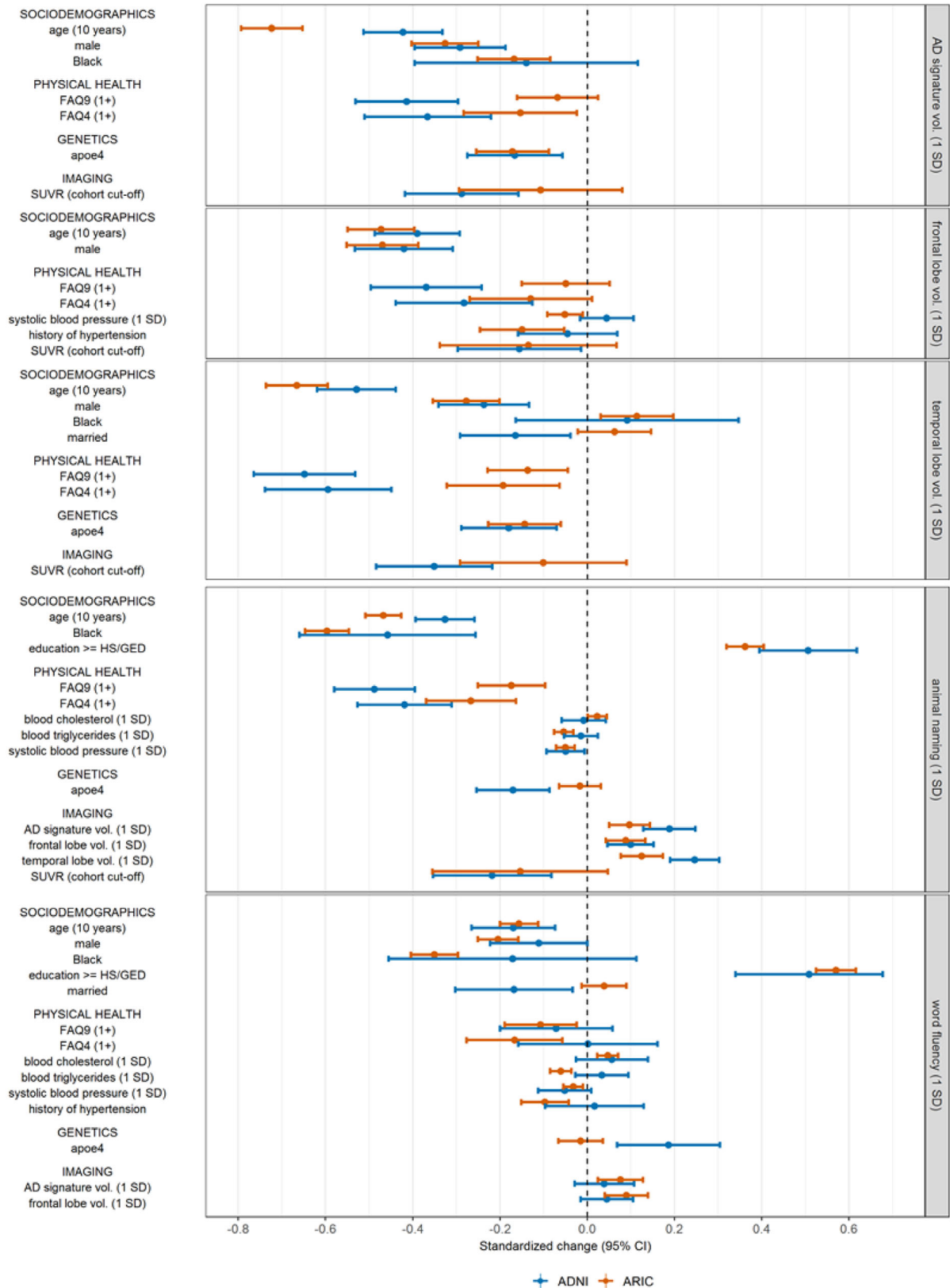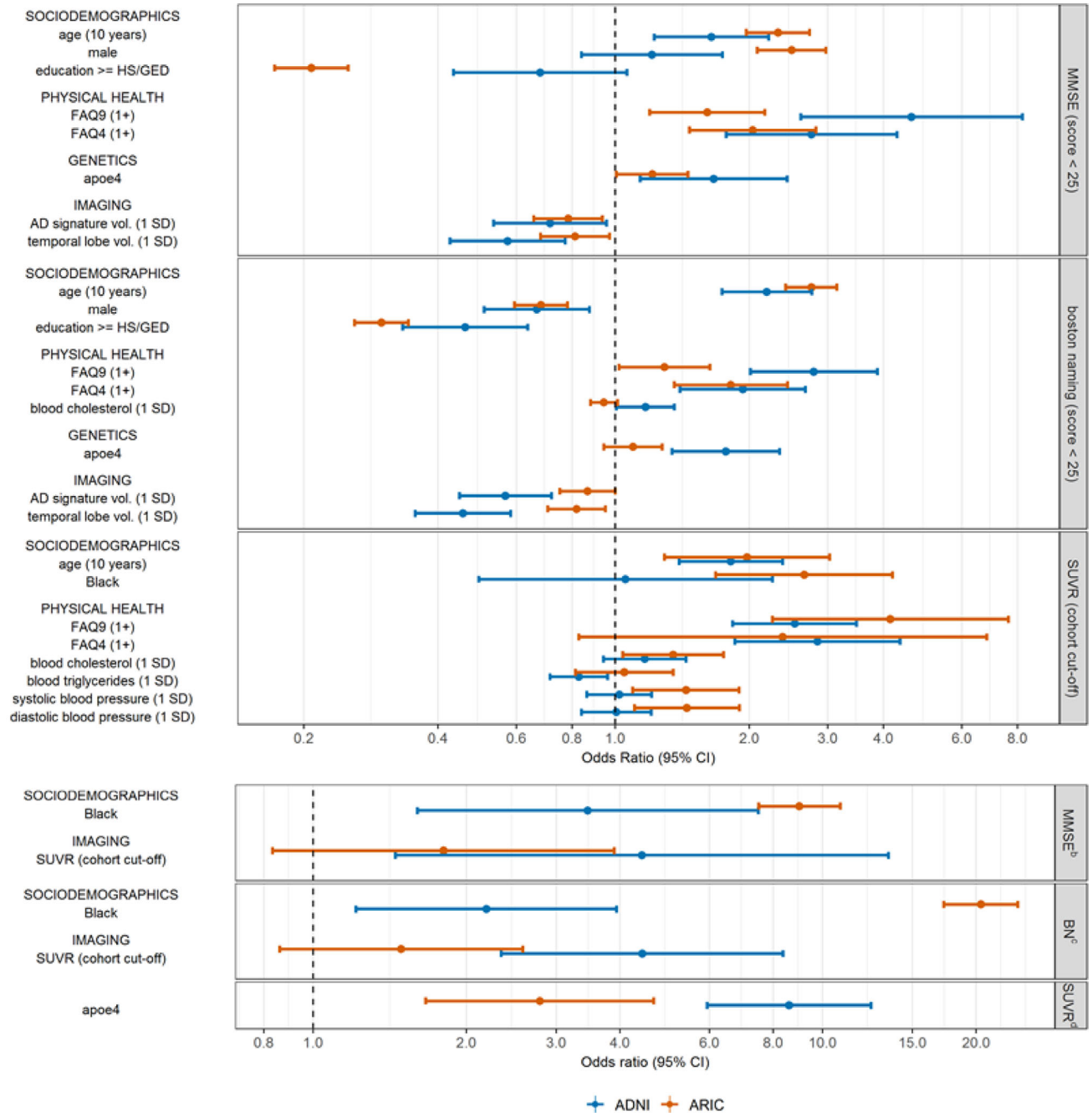
^a for associations involving SUVR, this excludes observations of participants with dementia

FIGURE 3 Comparison of estimated associations in Alzheimer's Disease Neuroimaging Initiative versus in Atherosclerosis Risk in Communities using all available observations^a that are significant to *P* < 0.05 in at least one cohort; continuous outcomes. AD, Alzheimer's disease; APOE, apolipoprotein E ε4 positive; BP, blood pressure; FAQ, Functional Activities Questionnaire; GED, General Education Development; HS, high school; SD, standard deviation; SUVR, standardized uptake value ratio

a for associations involving SUVR, this excludes observations of participants with dementia

b MMSE (score < 25)

c boston naming (score < 25)

d SUVR (cohort cut-off)

FIGURE 4 Comparison of estimated associations in Alzheimer's Disease Neuroimaging Initiative versus in Atherosclerosis Risk in Communities using all available observations[a] that are significant to P < 0.05 in at least one cohort; binary outcomes. AD, Alzheimer's disease; APOE, apolipoprotein Eε4 positive; BP, blood pressure; FAQ, Functional Activities Questionnaire; GED, General Education Development; HS, high school; SD, standard deviation; SUVR, standardized uptake value ratio

CONFLICTS OF INTEREST

ORCID

Kan Z. Gianattasio https://orcid.org/0000-0002-8665-2856

REFERENCES

1. Kukull WA, Ganguli M. Generalizability: the trees, the forest, and the low-hanging fruit. *Neurology*. 2012;78(23):1886-1891.
2. Alzheimer's Disease Neuroimaging Initiative. Study Design. *Alzheimer's Disease Neuroimaging Initiative*. http://adni.loni.usc.edu/study-design/. Published 2017. Accessed September 24, 2019.
3. Veitch DP, Weiner MW, Aisen PS, et al. Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's Dement*. 2019;15(1):106-152.
4. Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: reviewing progress toward improved AD clinical trials. *Alzheimer's Dement*. 2017;13(4):e1-e85. https://doi.org/10.1016/j.jalz.2016.11.007
5. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209.
6. Weiner MW, Aisen PS, Jack Jr CR, et al. The Alzheimer's Disease Neuroimaging Initiative: progress report and future plans. *Alzheimer's Dement*. 2010;6(3):202-211. https://doi.org/10.1016/j.jalz.2010.03.007.The
7. Cummings JL. Integrating ADNI results into Alzheimer's disease drug development programs. *Neurobiol Aging*. 2010;31(8):1481-1492.
8. Johnson JK, Gross AL, Pa J, Mclaren DG, Park LQ, Manly JJ. Longitudinal change in neuropsychological performance using latent growth models: a study of mild cognitive impairment. *Brain Imaging Behav*. 2012;6(4):540-550.
9. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin AJ. Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. *BMC Geriatr*. 2015;15(1):1-9.
10. Moonga I, Niccolini F, Wilson H, Pagano G, Politis M. Hypertension is associated with worse cognitive function and hippocampal hypometabolism in Alzheimer's disease. *Eur J Neurol*. 2017;24(9):1173-1182.
11. Hachinski VC, Iliff LD, Zilhka E, et al. Cerebral Blood Flow in Dementia. *Arch Neurol*. 1975;32(9):632-637.
12. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. 1983;17(1):37-49.
13. Roth CR. Boston naming test. In: Kreutzer JS, DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer; 2010:430-433.
14. Shao Z, Janse E, Visser K, Meyer AS. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front Psychol*. 2014;5:1-10.
15. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd ed. New York, NY: Oxford University Press; 2006.
16. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
17. Julayanont P, Nasreddine ZS. The Montreal Cognitive Assessment (MoCA): Concept and Clinical Review. In: Larner AJ, ed. *Concept and Clinical Review*. Cognitive Screening Instruments. Springer; 2017:139-195.
18. Alzheimer's Disease Neuroimaging Initiative. ADNI 1 Clinical Protocols. 2008. http://adni.loni.usc.edu/wp-content/themes/freshnews-dev-v2/documents/clinical/ADNI-1_Protocol.pdf. Accessed July 7, 2020.
19. The Atherosclerosis Risk in Communities Study. Atherosclerosis Risk in Communities Study Protocol Manual 2: Cohort Component Procedures. https://sites.cscc.unc.edu/aric/sites/default/files/public/manuals/Cohort_Procedures.1_2.pdf. Published 1988. Accessed June 18, 2020.
20. Knopman DS, Gottesman RF, Sharrett AR, et al. Mild cognitive impairment and dementia prevalence: the Atherosclerosis Risk in Communities Neurocognitive Study. Alzheimer's Dement Diagnosis. *Assess Dis Monit*. 2016;2:1-11.
21. Mayo AM, Use of the Functional Activities Questionnaire in Older Adults with Dementia.; 2016. https://www.alz.org/careplanning/downloads/functional-activities-questionnaire.pdf.

22. Wong DF, Rosenberg PB, Zhou Y, et al. In Vivo Imaging of Amyloid Deposition in Alzheimer's Disease using the Novel Radioligand [18 F]AV-45 (Florbetapir F 18). *J Nucl Med.* 2010;51(6):913-920.

23. Gottesman RF, Schneider ALC, Zhou Y, et al. The ARIC-PET amyloid imaging study. *Neurology.* 2016;87(5):473-480.

24. Hughes TM, Wagenknecht LE, Craft S, et al. Arterial stiffness and dementia pathology: atherosclerosis risk in communities (ARIC)-PET study. *Neurology.* 2018;90(14):E1248-E1256.

25. The Laboratory of Neuro Imaging (LONI). ADNI2 PET technical procedures manual AV-45 (Florbetapir F 18) & FDG. 2011;45:35. https://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_PET_Tech_Manual_0142011.pdf.

26. Landau S, Jagust W. Florbetapir processing methods. Alzheimer's Disease Neuroimaging Initiative.

27. Ossenkoppele R, Jansen WJ, Rabinovici GD, et al. Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *Jama.* 2015;313(19):1939-1949.

28. Landau SM, Hong A, Fero A, Jagust WJ, Amyloid negativity in clinically diagnosed Alzheimer's disease and MCI. *Alzheimer's Dement.* 2015;86:1377-1385. https://doi.org/10.1016/j.jalz.2015.06.079

29. 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 Neurol.* 2017;74(10):1246-1254.

30. Nägga K, Gustavsson AM, Stomrud E, et al. Increased midlife triglycerides predict brain β-amyloid and tau pathology 20 years later. *Neurology.* 2018;90(1):E73-E81.

31. Knopman DavidS, Griswold ME, Lirette ST, et al. Vascular imaging abnormalities and cognition: mediation by Cortical Volume in nondemented persons: aRIC-NCS Study. *Stroke.* 2015;46(2):433-440.

32. Alzheimer's Disease Neiruoimaging Initiative. About ADNI. http://adni.loni.usc.edu/about/. Accessed March 26, 2019.

33. Solomon A, Dobranici L, Kareholt I, Tudose C, Lăzărescu M. Comorbidity and the rate of cognitive decline in patients with Alzheimer dementia. *Int J Geriatr Psychiatry.* 2011;26(12):1244-1251.

34. Aubert L, Pichierri S, Hommet C, Camus V, Berrut G, De Decker L. Association between comorbidity burden and rapid cognitive decline in individuals with mild to moderate alzheimer's disease. *J Am Geriatr Soc.* 2015;63(3):543-547.

35. Bäckman L, Jones S, Small BJ, Agüero-Torres H, Fratiglioni L. Rate of cognitive decline in preclinical Alzheimer's disease: the role of comorbidity. *Journals Gerontol - Ser B Psychol Sci Soc Sci.* 2003;58(4):228-236.

36. Mungas D, Gavett B, Fletcher E, Farias ST, DeCarli C, Reed B. Education amplifies brain atrophy effect on cognitive decline: implications for cognitive reserve. *Neurobiol Aging.* 2018;68:142-150.

37. Yaffe K, Weston A, Graff-Radford NR, et al. Association of plasma β-amyloid level and cognitive reserve with subsequent cognitive decline. *JAMA.* 2011;305(3):261-266.

38. Soldan A, Pettigrew C, Li S, et al. Relationship of cognitive reserve and cerebrospinal fluid biomarkers to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Neurobiol Aging.* 2013;34(12):2827-2834.

39. Arenaza-Urquijo EM, Molinuevo JL, Sala-Llonch R, et al. Cognitive reserve proxies relate to gray matter loss in cognitively healthy elderly with abnormal cerebrospinal fluid amyloid-β levels. *J Alzheimer's Dis.* 2013;35(4):715-726.

40. Ebrahim S, Smith GD. Commentary: should we always deliberately be non-representative?. *Int J Epidemiol.* 2013;42(4):1022-1026.

41. Elwood JM. Commentary: on representativenes. *Int J Epidemiol.* 2013;42(4):1014-1015.

42. Nohr EA, Olsen J. Commentary: epidemiologists have debated representativeness for more than 40 years-has the time come to move on?. *Int J Epidemiol.* 2013;42(4):1016-1017.

43. Richiardi L, Pizzi C, Pearce N. Commentary: representativeness is usually not necessary and often should be avoided. *Int J Epidemiol.* 2013;42(4):1018-1022.

44. Rothman KJ, Gallacher JEJ, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol.* 2013;42(4):1012-1014.

45. Dahabreh IJ, Hernán MA, Extending inferences from a randomized trial to a target population. arXiv:180500550. 2018:1-21. https://arxiv.org/pdf/1805.00550.pdf.

46. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *Am J Epidemiol.* 2017;186(8):1010-1014.

47. Stuart EA, Ackerman B, Westreich D. Generalizability of Randomized Trial Results to Target Populations: design and Analysis Possibilities. *Res Soc Work Pract.* 2018;28(5):532-537.

48. Mercer AW, Kreuter F, Keeter S, Stuart EA. Theory and Practice in Nonprobability Surveys. *Public Opin Q.* 2017;81:250-279.

49. Rudolph KE, Díaz I, Rosenblum M, Stuart EA. Estimating population treatment effects from a survey subsample. *Am J Epidemiol.* 2014;180(7):737-748.

50. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Stat Sci.* 2014;29(4):579-595.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.