Received: 29 November 2022 | Revised: 19 May 2023 | Accepted: 23 May 2023

DOI: 10.1002/dad2.12453

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

A robust harmonization approach for cognitive data from multiple aging and dementia cohorts

Joseph Giorgio[1,2] | **Ankeet Tanna**[3] | **Maura Malpetti**[4] | **Simon R. White**[5,6] | **Jingshen Wang**[7] | **Suzanne Baker**[8] | **Susan Landau**[1] | **Tomotaka Tanaka**[9] | **Christopher Chen**[9] | **James B. Rowe**[4,10] | **John O'Brien**[5,10] | **Jurgen Fripp**[11] | **Michael Breakspear**[2] | **William Jagust**[1,8] | **Zoe Kourtzi**[3] | **for the Alzheimer's Disease Neuroimaging Initiative, Australian Imaging Biomarkers and Lifestyle flagship study**

[1]Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California, USA

[2]School of Psychological Sciences, College of Engineering, Science and the Environment, University of Newcastle, Newcastle, New South Wales, Australia

[3]Department of Psychology, University of Cambridge, Cambridge, UK

[4]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

[5]Department of Psychiatry, University of Cambridge, Cambridge, UK

[6]MRC Biostatistics Unit, University of Cambridgeshire, Cambridge, UK

[7]Division of Biostatistics, University of California Berkeley, Berkeley, California, USA

[8]Molecular Biophysics & Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[9]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Kent Ridge, Singapore

[10]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[11]The Australian eHealth Research Centre, CSIRO Health and Biosecurity, Brisbane, Queensland, Australia

Correspondence
Joseph Giorgio, Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA 94720, USA.
E-mail: jgiorgio@berkeley.edu

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found in Appendix: Collaborators.

Abstract

INTRODUCTION: Although many cognitive measures have been developed to assess cognitive decline due to Alzheimer's disease (AD), there is little consensus on optimal measures, leading to varied assessments across research cohorts and clinical trials making it difficult to pool cognitive measures across studies.

METHODS: We used a two-stage approach to harmonize cognitive data across cohorts and derive a cross-cohort score of cognitive impairment due to AD. First, we pool and harmonize cognitive data from international cohorts of varying size and ethnic diversity. Next, we derived cognitive composites that leverage maximal data from the harmonized dataset.

RESULTS: We show that our cognitive composites are robust across cohorts and achieve greater or comparable sensitivity to AD-related cognitive decline compared to the Mini-Mental State Examination and Preclinical Alzheimer Cognitive Composite.

Joseph Giorgio and Ankeet Tanna contributed equally to this study.

Finally, we used an independent cohort validating both our harmonization approach and composite measures.

DISCUSSION: Our easy to implement and readily available pipeline offers an approach for researchers to harmonize their cognitive data with large publicly available cohorts, providing a simple way to pool data for the development or validation of findings related to cognitive decline due to AD.

1 | INTRODUCTION

The development of sensitive measures to track the cognitive decline associated with Alzheimer's disease (AD) is important for observational and interventional studies.[1] With the introduction of large longitudinal studies, multi-domain cognitive test batteries that may minimally overlap between different studies have proliferated.[2–6] Given the high dimensionality in testing batteries, researchers often combine multiple test items into cognitive composites to predict diagnostic outcomes in preclinical AD and track longitudinal change.[7–14] Composites cover either a single domain (e.g., memory[15] or executive function[16]) or global cognition[7,8,10,17] but may be limited in their application due to non-overlapping cognitive testing batteries. Therefore, researchers may need to substitute tests in composite construction,[10,11] impeding the ability to harmonize cohorts.

Data harmonization is a field of integrative data analysis that allows researchers to pool data across multiple studies when there is imperfect overlap between data acquired in studies.[18] The harmonized data places variables on the same scale to permit pooling across a large number of studies.[19] Unlike meta-analysis, which only allows researchers to combine summary statistics, harmonization allows researchers to pool raw data for model development and hypothesis testing, mitigating sampling biases and power constraints.[18] Several approaches exist for data harmonization including variable standardization, latent variable models, and imputation approaches.[19,20]

Here, we use an imputation approach to harmonize item-level neuropsychological data, predicting missing data for any individual based on patterns of overlapping data. Several imputation approaches can be applied in psychological research[21,22] with parametric methods popular for deriving missing variables.[23–27] However, parametric approaches are limited by the nature of the missing data[20] with cohort variation having a marked effect on cognitive trajectories.[28] Parametric approaches may hence not be ideal in harmonizing missing data[24] and non-parametric imputation approaches offer a promising method when sampling characteristics differ across cohorts.

Our study had three goals. First, we introduce a computationally efficient tool to non-parametrically harmonize cognitive data across cohorts at the test and subtest level. Next, we use this harmonized data to derive a cross-cohort cognitive composite, testing its sensitivity and robustness to cross-sectional and longitudinal amyloid beta (Aβ)-related change. Finally, we benchmark the sensitivity of the composite to prodromal and preclinical change against the widely used Mini-Mental State Examination (MMSE) or Preclinical Alzheimer Cognitive Composite (PACC). Our work presents a simple approach to harmonize diverse neuropsychological data and generate a robust and sensitive AD cognitive composite.

2 | MATERIALS AND METHODS

2.1 | Study participants

2.1.1 | Cohorts

Cognitive data was harmonized across four cohorts with positron emission tomography (PET) imaging and longitudinal neuropsychological data (assessed every 1 to 2 years). Individuals were included independent of PET imaging or number of assessments. A fifth independent validation sample was used to validate the harmonization approach.

1. The Alzheimer's Disease Neuroimaging Initiative (ADNI; adni.loni.usc.edu). ADNI data used in this analysis were collected from 2005 to late 2019 and included cognitively normal (CN), mild cognitive impairment (MCI), and AD individuals (n = 2513).
2. National University of Singapore (NUS) memory clinic sample includes individuals with no cognitive impairment (NCI), cognitive impairment no dementia (CIND) mild, CIND moderate, vascular dementia (VaD), and AD dementia.[29] To consolidate the stage of clinical impairment to be consistent with other cohorts we assigned NCI individuals as CN, CIND individuals as MCI, and VaD and AD individuals as AD (n = 636).
3. Neuroimaging of Inflammation in Memory and Related Other Disorders (NIMROD), a study performed in Cambridge, UK, that recruited patients from specialist secondary and tertiary care services in the east of England and the Join Dementia Research registry. CN controls were also recruited regionally from volunteer registries.[30] Individuals are defined as CN, MCI, or AD at baseline (n = 89).
4. Berkeley Aging Cohort Study (BACS), a cohort of elderly individuals who had psychometrically normal cognition at baseline, residing in the community in the San Francisco Bay area (n = 188).[31]
5. The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) served as the validation cohort. The AIBL sample

used in this analysis was collected from 2006 to 2021 and was composed of CN individuals and MCI or AD patients (n = 1820;[3] Table 1).

2.1.2 | Consent statement

All subjects provided informed consent and relevant ethics approval was acquired for each cohort.

2.2 | Harmonization

2.2.1 | Cognitive testing batteries

All four harmonization cohorts included comprehensive batteries of tests that interrogated multiple domains of cognition for up to 13 years of follow-up. For each neuropsychological visit we decomposed tests into subtest-level neuropsychological variables that revealed greater similarity. For example, decomposing the Addenbrooke's Cognitive Examination Revised from the NIMROD study allowed for harmonization of clock-drawing and fluency tasks, which were given as standalone tasks in ADNI. For test and subtests that were similar across testing regimes, and where prior evidence highlighted high correlation between variables (e.g., comparing the California/Hopkins/Rey Auditory Verbal Learning Tests [C/H/RAVLT] to each other), we scaled and aligned these to represent the same test or subtest variable (Table S1 in supporting information). After alignment, the resultant cognitive scores covered 125 variables, with varying degrees of overlap among the four harmonization cohorts. We followed the same alignment and scaling procedure for the 3920 neuropsychological visits in the AIBL Validation sample (Figure 1).

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the literature using traditional sources, including Google Scholar, meeting abstracts, and presentations. Several approaches exist to harmonize differentially sampled cognitive data across cohorts. These publications are appropriately cited.
2. **Interpretation**: We developed a simple approach to harmonize data across international aging and dementia research cohorts of varying sizes. We show that these harmonized data have highly consistent covariance patterns suggesting the data can be used to derive and validate robust cross cohort cognitive composites. We introduce a simple cognitive composite that is highly sensitive to amyloid status throughout clinical syndromes.
3. **Future directions**: Our easy-to-implement pipeline allows researchers to pool their data to confirm hypotheses and validate models across cohorts. Once harmonized the data can be used in the empirical derivation of cognitive composites. Our approach can be extended to incorporate additional data as it becomes available, allowing for an expanded library of neuropsychological variables and sample heterogeneity.

2.2.2 | Imputation of missing neuropsychological variables

Test harmonization across the cohorts used k-nearest neighbors (k-NN), a non-parametric approach to impute missing values by

TABLE 1 Cohort characteristics and demographics. Demographic characteristics for participants in all four cohorts. Statistics derived using one-way ANOVA. Asterisks (*) indicate baseline demographic (i.e., age, sex, education, MMSE) values significantly different to ADNI ($P < 0.001$).

Variable	ADNI	NUS[a]	NIMROD	BACS	AIBL
Sample size	2513	636	89	188	1820
Diagnosis	CN = 888 MCI = 1058 AD = 411 Missing = 156	CN = 29 MCI = 101 AD = 46 Missing = 450	CN = 38 MCI = 28 AD = 22 Missing = 0	CN = 188 MCI = 0 AD = 0 Missing = 0	CN = 1091 MCI = 395 AD = 316 Missing = 0
Total number of visits	10,622	2553	255	824	3920
Follow-up years, mean (SD)	2.78 (3.05)	3.48 (1.43)	2.04 (1.40)	4.2 (3.25)	2.24 (2.87)
Age, mean (SD)	73.14 (7.35)	75.64 (7.29)*	72.15 (8.07)	75.78 (5.834)*	71.81 (7.05)*
Female, male sex	1006/1149	101/86	37/52	108/80	964/856*
Education, mean (SD)	15.87 (3.23)	7.6 (4.7)*	13.46 (2.88)*	16.85 (1.97)*	12.81 (3.11)*[b]
MMSE, mean (SD)	27.22 (2.99)	21.13 (6.29)*	26.64 (3.66)	28.80 (1.27)*	26.81 (3.27)
Number with Aβ PET	1260	185	20	188	1820

Abbreviations: Aβ, amyloid beta; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; ANOVA, analysis of variance; BACS, Berkeley Aging Cohort Study; CN, cognitively normal; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; NIMROD, Neuroimaging of Inflammation in Memory and Related Other Disorders study; NUS, National University of Singapore.
[a]Education/sex only available for NUS subjects with Aβ imaging.; PET, positron emission tomography; SD, standard deviation.
[b]Midpoint of discretized education used for those missing exact education.

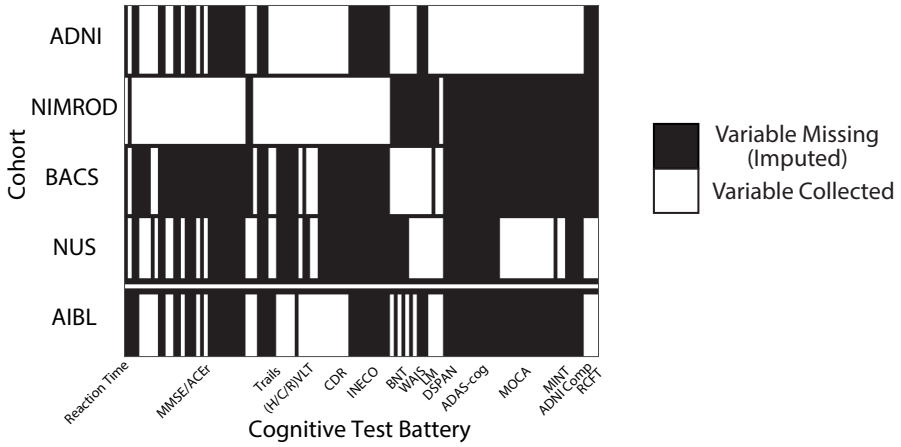

FIGURE 1 Overlapping cognitive variables. Overlap between test and subtest level neuropsychological variables across the four harmonization cohorts (top) and the AIBL Validation cohort (bottom). Regions shaded white are variables that were collected in a given cohort, regions shaded black are variables that are not collected in a given cohort and are imputed using k-NN imputation. ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; BACS, Berkeley Aging Cohort Study; BNT, Boston Naming Test; CDR, Clinical Dementia Rating; DSPAN, Digit Span; (H/C/R)VLT, Hopkins/California/Rey Verbal Learning Tests; INECO, Institute of Cognitive Neurology Frontal Screening; k-NN, k-nearest neighbors; LM, Logical Memory; MINT, Multilingual Naming Test; MOCA, Montreal Cognitive Assessment; NIMROD, Neuroimaging of Inflammation in Memory and Related Other Disorders study; NUS, National University of Singapore; RCFT, Rey Complex Figure Test; WAIS, Wechsler Adult Intelligence Scale

determining the *k* most similar cases (i.e., neuropsychological visit for any participant) and assigned missing values with the observed value from the closest case (i.e., $k = 1$) or the weighted average of the *k* closest cases. The result of the k-NN imputation is a complete set of neuropsychological variables for each visit (Supplementary Methods in supporting information).

2.2.3 | Cognitive composite derivation

The composition of the Cross-Cohort Alzheimer Cognitive Composite (CC-ACC) builds on established methods used to create the PACC.[10] To derive the CC-ACC, we used the variance-normalized mean of all variables belonging to the three PACC domains (i.e., memory, executive function, and general cognition; Table S2 in supporting information). The mean of the variance-normalized scores was taken to be the score for that domain. These domain scores were then summed and standardized similar to the PACC derivation (Supplementary Methods). Scores are combined so that a decrease is associated with worsening cognition.

2.3 | AD related cognitive decline

2.3.1 | PET neuroimaging, Aβ positivity, and tau stage

Of the 5246 individuals with neuropsychological testing, 3473 also had baseline Aβ PET imaging (Table 1). PET data were analyzed using cohort-specific pre-processing pipelines to measure amyloid in the Centiloid (CL) scale and Aβ positivity assigned as a value of CL >15.[32] Because data from NUS had not been scaled to Centiloids, a visual assessment was used to define amyloid positivity (Supplementary Methods). Within the Harmonization cohort 576 individuals with Aβ PET (286 Aβ+) underwent 18F-Flortaucipir tau (FTP) PET imaging (n[FTP/Aβ+FTP]: ADNI 444/232; NIMROD 15/15; BACS 117/39). Data were summarized for three Braak staging regions I (entorhinal), III/IV (inferolateral temporal), and V/VI (extra-temporal neocortical). Individuals were assigned as either tau negative (T−) or assigned a tau positive (T+) Braak stage (i.e., individuals are assigned one of four potential tau categories; T−, T+ Braak I, T+ Braak III/IV, T+ Braak V/VI) based on previously published thresholds[33] (Supplementary Methods).

2.3.2 | CC-ACC sensitivity to Aβ and tau pathology

To examine relationships between baseline Aβ and longitudinal changes in CC-ACC, we use linear mixed effects (LME) models stratified by clinical impairment (i.e., CN or MCI/AD). Within each model the response variable is the CC-ACC and fixed effect predictor variables entered as either mean centered continuous variables (i.e., years from baseline, age at baseline, education) or categorical variables (Aβ status, sex, cohort). To examine the sensitivity of the CC-ACC to tau severity we assigned Aβ-positive individuals a Braak tau stage and entered this as a categorical variable of interest (Supplementary Methods).

2.3.3 | CC-ACC prediction of Aβ status in MCI

We compared both baseline and longitudinal CC-ACC scores to the MMSE and PACC (ADNI and AIBL Validation cohorts) in classification of baseline Aβ status for patients with MCI. We used logistic

regression to fit baseline and annualized rate of change in cognitive score as predictors with Aβ status as the target variable. To assess how well each cognitive score discriminated Aβ status we compared mean differences and effect sizes as well as the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. To determine whether the discriminability of CC-ACC and MMSE or PACC differed we bootstrapped our sample 1000 times (with replacement) forcing balanced numbers of Aβ-positive and -negative individuals in each bootstrap. Within each bootstrap we ran the logistic regression and assessed significant differences in AUC using a paired *t* test across bootstraps.

2.3.4 | CC-ACC as a clinically meaningful endpoint for preclinical AD clinical trials

We compared the sensitivity of the CC-ACC to the PACC and MMSE as a clinically meaningful endpoint for a hypothetical clinical trial targeting CN Aβ-positive individuals. We assessed performance using one-sample, two-tailed *t* tests on the annualized rate of cognitive score decline for longitudinal follow-up of 1 to 7 years. Annualized rate of cognitive score change was calculated using the linear least squares fit of CC-ACC/PACC/MMSE and years from baseline. We compared the sample size needed for an arm of a hypothetical clinical trial designed to detect a 25% reduction in annual cognitive score change with a significance of 0.05 and a power of $\alpha = 0.8$. We defined the null hypothesis as the mean and standard deviation of the rate of change calculated from the observed sample, for which the alternative hypothesis is a 25% reduction of the mean of the observed sample.

2.4 | Code and data availability

The MATLAB code and raw data to impute missing item-level variables then calculate the CC-ACC for any new data set are available online (https://github.com/jjgiorgio/cognitive_harmonisation). Source data may be requested from the respective cohorts or Dementias Platform UK.

3 | RESULTS

3.1 | Data harmonization

K-NN imputation resulted in a set of 125 real and imputed cognitive variables. We observed a good agreement between hidden and ground truth C/H/RAVLT total scores after serial imputation (NUS $R^2 = 0.75$ root mean square error [RMSE] = 8.1; NIMROD $R^2 = 0.95$, RMSE = 3.6; BACS $R^2 = 0.53$, RMSE = 8; AIBL $R^2 = 0.95$, RMSE = 3.4). We also observed a high consistency between item-to-item correlations (Figure 2) across cohorts with similar diagnostic categories to ADNI (i.e., CN, MCI, and AD; Harmonization cohorts: ADNI vs. NUS = 96%; ADNI vs. NIMROD = 95%; Validation cohort: ADNI vs. AIBL = 97%; Figure S1 in supporting information). Comparing the entire ADNI sample with BACS the association between variable correlations is weaker ($R^2 = 83.1\%$); however, re-calculating the correlation matrix in ADNI for only individuals who are CN at baseline (i.e., to have a more similar sample to the BACS cohort) significantly improved the cognitive associations ($R^2 = 87.4\%$, Steiger's $Z = 19.4$, $P < 0.001$). There was variability in imputation performance across cohorts, with BACS being the poorest, likely due to the smallest number of overlapping tests in BACS (30 shared variables). However, we observed highly reproducible loadings of the real variables and the imputed ADNI composites across each cohort (Figure 2). Further, we observed negligible effects on imputation quality based on missing not at random mechanisms (Supplementary Results: Effect of Missing Not at Random, Figures S2 and S3 in supporting information) suggesting that it is the number of overlapping variables rather than the proportion of missingness that has the largest impact on imputation quality. Finally, we ran a parametric imputation pipeline (Multiple Imputation by Chained Equations [MICE]) using the *mice* package in R.[34] We observed the correlation structure between variables was not preserved across cohorts using MICE and the imputation quality of the ADNI composites was significantly poorer than our k-NN approach. Together, this suggests that a simple parametric imputation approach does not harmonize neuropsychological variables across cohorts as well as our non-parametric k-NN imputation (Supplementary Results: Parametric Imputation, Figures S4, S5, Table S3 in supporting information). For subsequent analyses the real and imputed values derived from k-NN are used to derive the CC-ACC as a harmonized cognitive composite.

3.2 | CC-ACC association with global cognitive and functional impairment

For a subsample of 2602 individuals from the Harmonization cohort (baseline diagnosis CN = 926, MCI = 1086, AD = 433, missing = 157) with baseline and/or follow-up Clinical Dementia Rating (CDR) we observed a strong association between the CC-ACC and the global CDR across all years (Kendall's tau [τ] all years: $\tau(10875) = -0.63$, $P < 0.0001$) and throughout each year of follow-up. We repeated these analyses for a subsample of 1785 individuals from the AIBL Validation cohort (baseline diagnosis CN = 1086, MCI = 391, AD = 308) observing a highly similar association between the baseline CC-ACC and CDR (Kendall's τ all years: $\tau(3920) = -0.60$, $P < 0.0001$) and throughout each year of follow-up (Figure S6, Table S4 in supporting information).

3.3 | CC-ACC sensitivity to Aβ-related change for CN individuals

The LME model fitting CN individuals from the Harmonization sample ($n = 667$) showed a significant main effect of Aβ status (F[1,2650] = 17.11, $P < 0.001$, $\beta = -0.379$ [−0.558, −0.199]) and a significant interaction between baseline Aβ status and time from baseline (F[1,2650] = 20.722, $P < 0.0001$, $\beta = -0.1197$ [−0.171 −0.0681]; Figure 3a). We

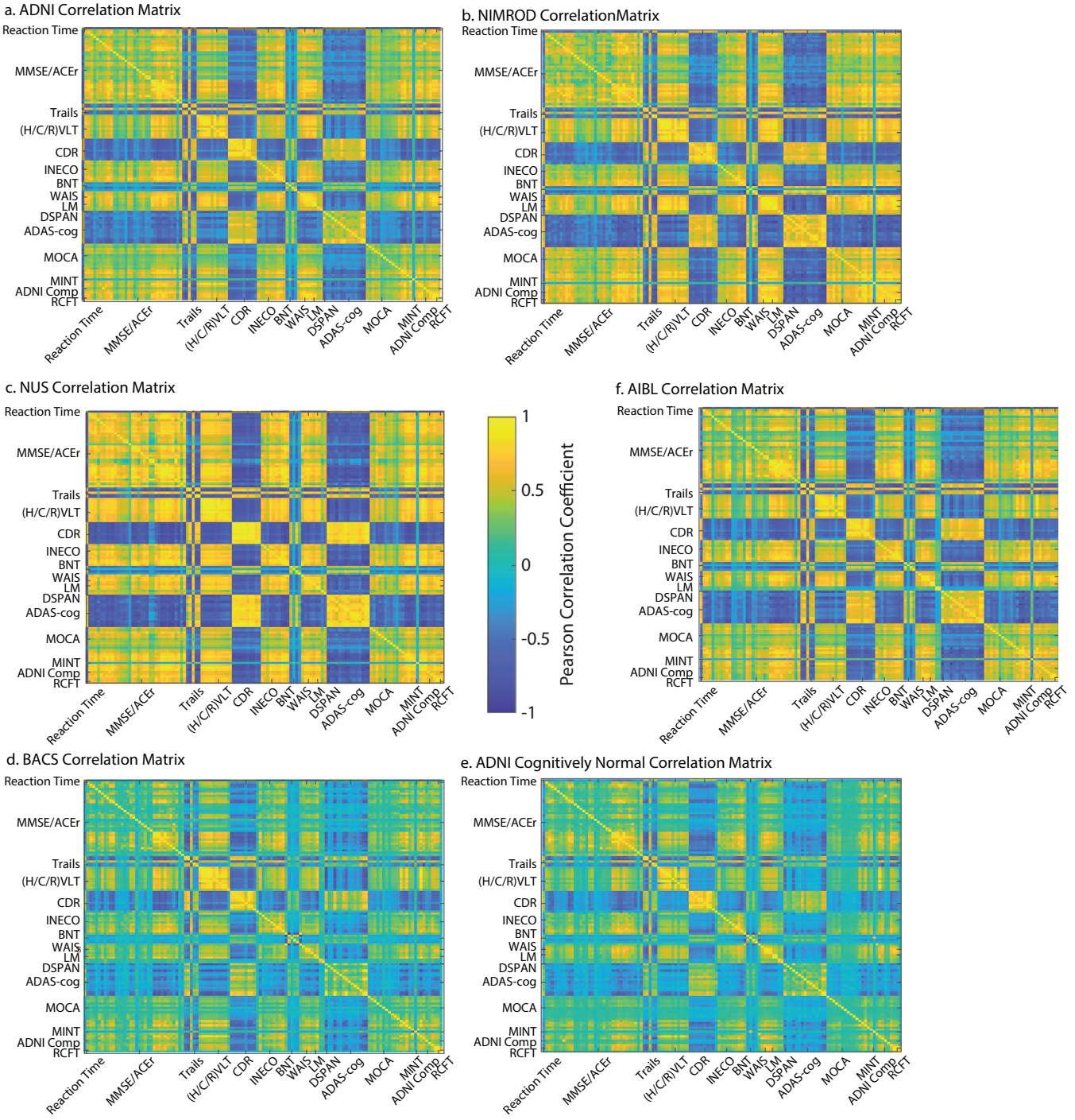

FIGURE 2 Item-to-item correlation matrices. The neuropsychological item to item correlation matrices: a, entire ADNI sample correlation matrix; b, NIMROD correlation matrix; c, NUS correlation matrix; d, BACS correlation matrix; e, ADNI correlation matrix for subsample who were cognitively normal at baseline; f, AIBL Validation cohort. We observed highly reproducible loadings of the real variables and the imputed ADNI composites across each cohort, ADNI-Mem versus C/H/RAVLT total (reference ADNI: $R^2 = 89.3\%$; NUS: $R^2 = 90.5\%$; NIMROD: $R^2 = 94.6\%$; BACS: $R^2 = 84.5\%$; AIBL: $R^2 = 92.6\%$) and ADNI-EF versus log(Trails B) (reference ADNI: $R^2 = 85.1\%$; NUS: $R^2 = 89.6\%$; NIMROD: $R^2 = 89.9\%$; BACS: $R^2 = 78.7\%$; AIBL, Trails B not collected). ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; BACS, Berkeley Aging Cohort Study; BNT, Boston Naming Test; CDR, Clinical Dementia Rating; DSPAN, Digit Span; (H/C/R)VLT, Hopkins/California/Rey Verbal Learning Tests; INECO, Institute of Cognitive Neurology Frontal Screening; k-NN, k-nearest neighbors; LM, Logical Memory; MINT, Multilingual Naming Test; MOCA, Montreal Cognitive Assessment; NIMROD, Neuroimaging of Inflammation in Memory and Related Other Disorders study; NUS, National University of Singapore; RCFT, Rey Complex Figure Test; WAIS, Wechsler Adult Intelligence Scale

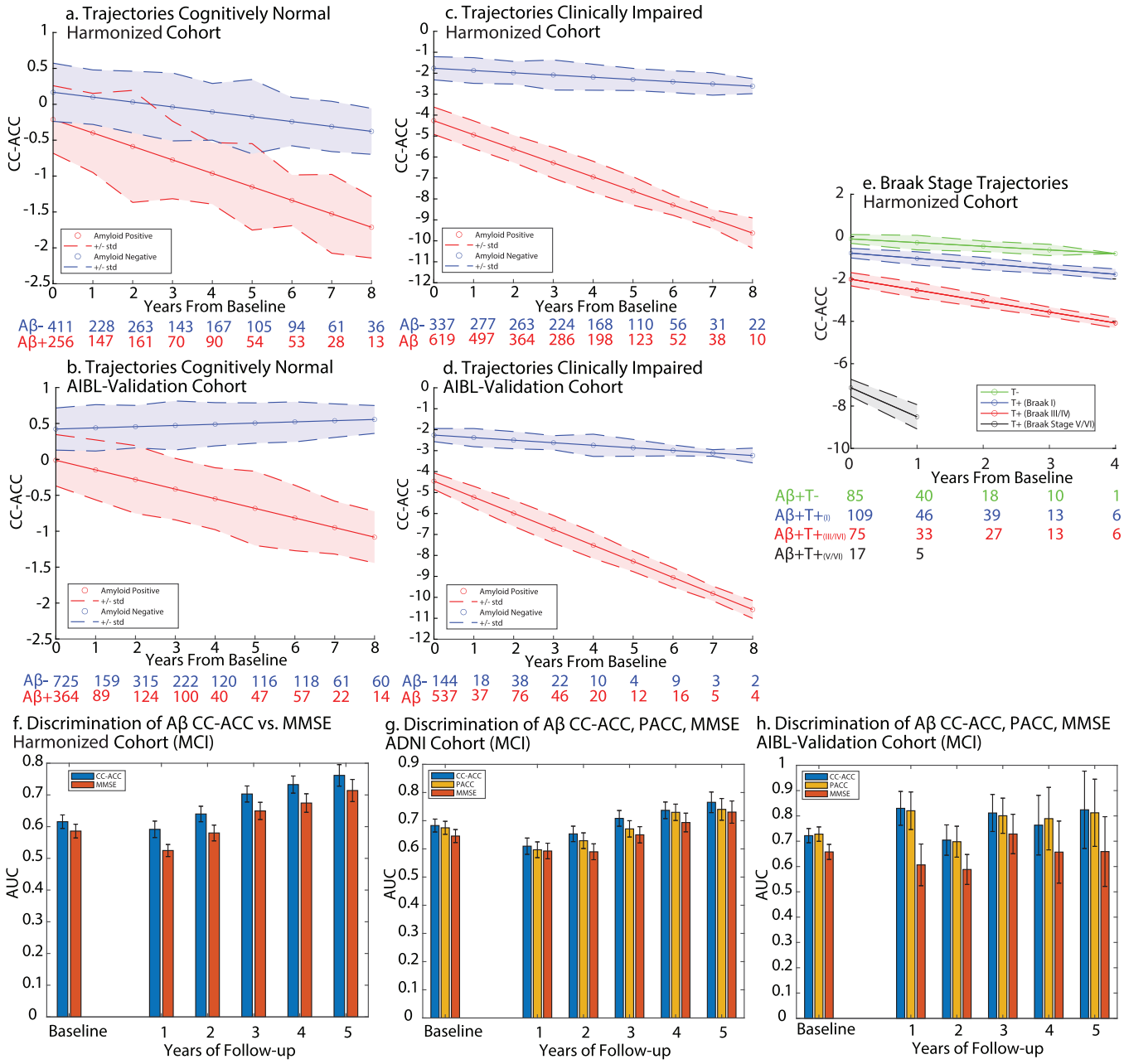

FIGURE 3 Cognitively normal trajectories: a, Harmonization cohort, b, AIBL Validation cohort. Expected trajectories for an Aβ– (blue) and an Aβ+ (red) 70-year-old women with an education of 14 years with no cognitive impairment (i.e., CN) at baseline. Clinically impaired trajectories: c, Harmonization cohort, d, AIBL Validation cohort. Plots show the expected trajectories for an Aβ– (blue) and an Aβ+ (red) 70-year-old woman with an education of 14 years with cognitive impairment (i.e., MCI or AD) at baseline. e, Braak stage trajectories. Expected trajectories for an Aβ+ 70-year-old woman with an education of 14 years who is tau negative (green), tau Braak I positive (blue), tau Braak III/IV positive (red), and tau Braak V/VI positive (black). Error bars represent standard deviation of the residual of the fit from the LME. Numbers below the plots show the sample size at different years of follow-up. Discriminatory power of CC-ACC for baseline Aβ status for MCI patients. f, Harmonization cohort, g, ADNI cohort, h, AIBL Validation Cohort. AUC of the receiver operating characteristic curve for a logistic regression predicting baseline Aβ. Blue bars represent the AUC for either the baseline CC-ACC or the annualized rate of change of the CC-ACC at different durations of follow-up. Red bars represent the AUC for either the baseline MMSE or the annualized rate of change of the MMSE at different durations of follow-up. Yellow bars represent the AUC for either the baseline PACC or the annualized rate of change of the PACC at different durations of follow-up. Error bars indicate the standard deviation of the AUC across bootstraps. Aβ, amyloid beta; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; AUC, area under the curve; CN, cognitively normal; CC-ACC, Cross-Cohort Alzheimer Cognitive Composite; LME, linear mixed effects; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PACC, Preclinical Alzheimer Cognitive Composite

observed no potential cohort effects, with a non-significant three-way interaction among cohort, Aβ status, and years from baseline F(1,2645) = 1.228, $P = 0.2677$. CN individuals from the NUS cohort were removed from this analysis as there are limited Aβ+ samples ($n = 4$). Cohort-level variation is shown in Figure S7 in supporting information. Repeating these analyses in the CN subsample of the AIBL Validation cohort ($n = 1089$) showed a significant main effect of Aβ status (F[1,2746] = 50.66, $P < 0.0001$, $\beta = -0.4345$ [−0.55, −0.314]) and a significant interaction between baseline Aβ status and years from baseline (F[1,2746] = 62.191, $P < 0.0001$, $\beta = -0.1505$ [−0.1879, −0.113]; Figure 3b). Comparing the harmonized sample with the AIBL Validation sample showed no significant three-way interaction among cohort, Aβ status, and years from baseline (F[1,5398] = 1.782, $P = 0.181$). Thus, the CC-ACC is sensitive to baseline Aβ status and Aβ-related cognitive decline and is robust across cohorts.

3.4 | CC-ACC sensitivity to Aβ-related change for clinically impaired individuals

LME models fitting clinically impaired individuals from the Harmonization sample ($n = 956$; baseline diagnosis: MCI = 683, AD = 273) showed a significant main effect of Aβ status (F[1,4308] = 154.99, $P < 0.0001$, $\beta = -2.5014$ [−2.895, −2.107]) and a significant interaction between baseline Aβ status and years from baseline (F[1,4308] = 155.73, $P < 0.0001$, $\beta = -0.56343$ [−0.65195, −0.4749]; Figure 3c). To test for a significant effect of cohort on the interaction of Aβ and time from baseline we truncated the sample to include only up to 4 years from baseline because of differences in maximum follow-up duration and included the interaction of cohort and education as confounds. We observed no significant three-way interaction among cohort, Aβ status, and years from baseline (F[2,3848] = 2.5, $p = 0.082$). Cohort-level variation is shown in Figure S7. Similar results were seen in the clinically impaired subsample of the AIBL Validation cohort ($n = 681$; baseline diagnosis: MCI = 390, AD = 291) with a significant main effect of Aβ status (F[1,995] = 103.55, $P < 0.0001$, $\beta = -2.2164$ [−2.6438, −1.789]) and a significant interaction between baseline Aβ status and years from baseline (F[1,995] = 51.583, $P < 0.0001$, $\beta = -0.64948$ [−0.8269, −0.47203]; Figure 3d). Comparing the harmonized sample with the AIBL Validation cohort showed no significant three-way interaction among cohort, Aβ status, and years from baseline (F[1,5303] = 0.9, $P = 0.34$). Thus, for clinically impaired individuals, the CC-ACC is sensitive to baseline Aβ status as well as Aβ-related cognitive decline and is robust across cohorts.

3.5 | CC-ACC sensitivity to Aβ+ T−, Aβ+ T+ (Braak I, Braak III/IV, Braak V/VI) stage

Finally, we investigated if the CC-ACC differed for Aβ-positive individuals based on the cortical burden of tau. First, we assigned Aβ-positive individuals as either tau negative (T−; $n = 85$) or tau positive (T+; $n = 201$; stages: T+: Braak I $n = 109$; Braak III/IV $n = 75$; Braak V/VI $n = 17$). From the LME we observed a main effect of tau status (F[3,533] = 58.8, $P < 0.0001$) and a significant interaction between baseline tau status and time from baseline (F[3,533] = 6.8, $P < 0.0005$; Figure 3e). Therefore, the CC-ACC is sensitive to a multimodal biological staging of AD based on Aβ status and the severity of tau pathology.

3.6 | CC-ACC sensitivity to Aβ status for MCI patients

Comparing mean differences in baseline CC-ACC for Aβ-positive versus Aβ-negative MCI individuals ($n = 688$) we observe significantly lower CC-ACC at baseline for Aβ-positive individuals. There was also a significantly greater rate of CC-ACC decline for Aβ-positive individuals for 1 to 5 years of follow-up. Results in MCI patients ($N = 395$) from the AIBL Validation cohort were similar, with significantly lower CC-ACC at baseline for Aβ-positive individuals and a significantly greater rate of CC-ACC decline for Aβ-positive individuals for 1 to 4 years of follow-up (Table S5 in supporting information). Results of the logistic regressions to discriminate Aβ status showed that the CC-ACC performs better at classifying Aβ status than the MMSE at all time points for both the Harmonization and AIBL Validation cohorts. In general, the CC-ACC also outperformed the PACC across follow-up in the ADNI and AIBL Validation samples (Figure 3f–h; Table S6 in supporting information). Therefore, we show that the CC-ACC is of equal or greater sensitivity to detect Aβ status than the MMSE and PACC at prodromal stages of AD (i.e., MCI).

3.7 | CC-ACC as a clinically meaningful endpoint for preclinical AD clinical trials

To compare the utility of the CC-ACC to detect cognitive decline for Aβ-positive CN individuals to the current "standard" cognitive composites, we compared the CC-ACC to the ADNI-PACC/MMSE and the AIBL-PACC/MMSE for Aβ-positive CN individuals from ADNI and AIBL. We observed similar performance in detecting cognitive decline between the CC-ACC and the ADNI-PACC in the same group of individuals. Further, we observed that the CC-ACC has marginally better statistical power than the ADNI-PACC delivering a reduction in required sample sizes to detect change in all but one of our time windows (Table 2). Highly similar results were seen in the AIBL CN sample with the CC-ACC delivering a reduction in sample sizes in all but one window (Table 2). Taken together we show that the CC-ACC achieves benchmark performance compared to the PACC in matched samples from the Harmonization cohort and the AIBL Validation cohort. The CC-ACC performed substantially better than the MMSE in both samples across follow-up durations.

TABLE 2 Duration of follow-up to observe cognitive decline in CC-ACC versus PACC for cognitively normal Aβ+ individuals. Mean, standard deviation, and statistics of the *t* test against zero using either the CC-ACC, PACC, or MMSE for varying follow up duration. *n* is the number of cognitively normal Aβ+ individuals from each sample used to calculate the test statistics. Sample size is number of individuals required to observe a hypothetical 25% decrease in cognitive decline using the CC-ACC, PACC, or MMSE. Italicized cells represent rates of decline that are significantly less than zero ($P < 0.05$).

		ADNI MMSE			ADNI PACC			ADNI CC-ACC		
Years of follow-up	*n*	Mean (std)	T-stat <0 (P)	Sample size	Mean (std)	T-stat <0 (P)	Sample size	Mean (std)	T-stat <0 (P)	Sample size
1	127	−0.091 (1.698)	−0.605 (0.546)	>2300	0.027 (0.983)	0.311 (0.756)	NA	0.015 (0.750)	0.223 (0.824)	NA
2	124	−0.153(1.278)	−1.33(0.186)	>2300	−0.081 (0.763)	−1.187 (0.237)	>2300	−0.090 (0.764)	−1.309 (0.193)	>2300
3	*49*	*−0.242 (0.744)*	*−2.278 (0.027)*	*1188*	*−0.241 (0.465)*	*−3.636 (0.001)*	*468*	*−0.189 (0.400)*	*−3.308 (0.002)*	*565*
4	*70*	*−0.116(0.423)*	*−2.287(0.025)*	*1683*	*−0.120 (0.311)*	*−3.233(0.002)*	*843*	*−0.113, (0.235)*	*−4.027 (<0.001)*	*545*
5	*41*	*−0.211(0.588)*	*−2.298(0.027)*	*978*	*−0.208(0.403)*	*−3.304(0.002)*	*474*	*−0.201, (0.375)*	*−3.438 (0.001)*	*438*
6	*41*	*−0.139(0.427)*	*−2.079(0.044)*	*1193*	*−0.168(0.271)*	*−3.952(<0.001)*	*332*	*−0.177, (0.217)*	*−5.225, (<0.001)*	*191*
7	*20*	*−0.239(0.374)*	*−2.851(0.01)*	*312*	*−0.289(0.318)*	*−4.066(0.001)*	*154*	*−0.300, (0.255)*	*−5.257, (<0.001)*	*93*
		AIBL MMSE			AIBL PACC			AIBL CC−ACC		
1	87	−0.236 (1.603)	−1.389 (0.168)	>2300	−0.154 (0.753)	−1.915 (0.059)	>2300	−0.074 (0.708)	−0.981 (0.329)	>2300
2	115	−0.052 (0.91)	−0.641 (0.523)	>2300	−0.022 (0.363)	−0.654 (0.514)	>2300	−0.051 (0.383)	−1.418 (0.159)	>2300
3	*99*	−0.073 (0.472)	−1.537 (0.127)	>2300	*−0.087 (0.299)*	*−2.893 (0.005)*	*1172*	*−0.094(0.309)*	*−3.035 (0.003)*	*1065*
4	37	−0.118 (0.446)	−1.666 (0.104)	1812	−0.086 (0.262)	−2.005 (0.053)	913	*−0.115 (0.331)*	*−2.107 (0.042)*	*827*
5	*45*	*−0.23 (0.679)*	*−2.325 (0.025)*	*1095*	*−0.160 (0.350)*	*−3.059 (0.004)*	*477*	*−0.174 (0.338)*	*−3.447 (0.001)*	*376*
6	*52*	*−0.144 (0.355)*	*−3.058 (0.003)*	*768*	*−0.102 (0.232)*	*−3.162 (0.003)*	*516*	*−0.091 (0.264)*	*−2.478 (0.017)*	*840*
7	22	−0.061 (0.249)	−1.146 (0.265)	2105	−0.064 (0.171)	−1.760 (0.093)	704	*−0.076 (0.142)*	*−2.513 (0.020)*	*347*

Abbreviations: Aβ, amyloid beta; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; CC-ACC, Cross-Cohort Alzheimer Cognitive Composite; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; PACC, Preclinical Alzheimer Cognitive Composite.

3.8 | CC-ACC executive function and memory subcomponents are sensitive to Aβ-related change

LME models for longitudinal change in CC-ACC measures of executive function and memory in CN individuals from the Harmonization sample ($n = 667$) showed a main effect of Aβ status for both executive function ($F[1,2650] = 14.30$, $P < 0.001$, $\beta = -0.28$ [−0.426, −0.135]) and memory ($F[1,2650] = 13.48$, $P < 0.001$, $\beta = -0.30$ [−0.466, −0.142]). Further, we observed a significant interaction between baseline Aβ status and time from baseline for both executive function ($F[1,2650] = 13.09$, $P < 0.001$, $\beta = -0.067$ [−0.104 −0.031]) and memory ($F[1,2650] = 24.18$, $P < 0.001$, $\beta = -0.097$ [−0.134 −0.061]; Figure 4a, c). For the CN AIBL Validation cohort ($n = 1089$) we showed similar results to the Harmonization cohort, observing a significant main effect of Aβ status for executive function ($F[1,2746] = 24.17$, $P < 0.001$, $\beta = -0.248$ [−0.348, −0.149]) and memory ($F[1,2746] = 40.08$, $P < 0.001$, $\beta = -0.419$ [−0.549, −0.289]). Further, we observed a significant interaction between baseline Aβ status and years from baseline for executive function ($F[1,2746] = 46.75$, $P < 0.001$, $\beta = -0.110$ [−0.142, −0.079]) and memory ($F[1,2746] = 62.27$, $P < 0.001$, $\beta = -0.131$ [−0.164, −0.0985]; Figure 4b, d).

Finally, the LME models in clinically impaired individuals from the harmonization sample ($n = 956$ baseline MCI; $N = 273$ AD) showed a main effect of Aβ status for both executive function ($F[1,4306] = 118.35$, $P < 0.001$, $\beta = -1.514$ [−1.787, −1.24]) and memory ($F[1,4306] = 195.67$, $P < 0.001$, $\beta = -1.61$ [−1.83, −1.38]). Further, we observed a significant interaction between baseline Aβ status and time from baseline for both executive function ($F[1,4306] = 121.5$, $P < 0.001$, $\beta = -0.354$ [−0.417 −0.291]) and memory ($F[(1,4306] = 134.56$, $P < 0.001$, $\beta = -0.252$ [−0.295 −0.21]; Figure 4e, g). For the clinically impaired AIBL Validation cohort ($n = 681$; baseline diagnosis: MCI = 390, AD = 291) we showed similar results to the Harmonization cohort, observing a significant main effect of Aβ status for executive function ($F[1,996] = 76.09$, $P < 0.001$, $\beta = -1.434$ [−1.756, −1.11]) and memory ($F[1,996] = 155.84$, $P < 0.001$, $\beta = -1.474$ [−1.705, −1.242]) with a significant interaction between baseline Aβ status and years from baseline for executive function ($F[1,996] = 46.71$, $P < 0.001$, $\beta = -0.422$ [−0.543, −0.3007]) and memory ($F[1,996] = 55.69$, $P < 0.001$, $\beta = -0.341$ [−0.431, −0.2514]; Figure 4f, h). Comparing how well each subdomain discriminated Aβ status for CN and MCI individuals throughout different duration of follow-up showed that executive function performed the poorest at discriminating baseline Aβ status (Supplementary Results: Sub-domain Discrimination of baseline Aβ status, Figure S8 in supporting information). Thus, our harmonization approach derives single domain composite scores that are sensitive to AD pathology for both CN and clinically impaired individuals.

4 | DISCUSSION

In this report, we use a simple imputation approach on test and subtest cognitive data to show that diverse AD cognitive test batteries can be harmonized across cohorts. The exemplar cognitive composite from our harmonized data (i.e., CC-ACC) showed high sensitivity

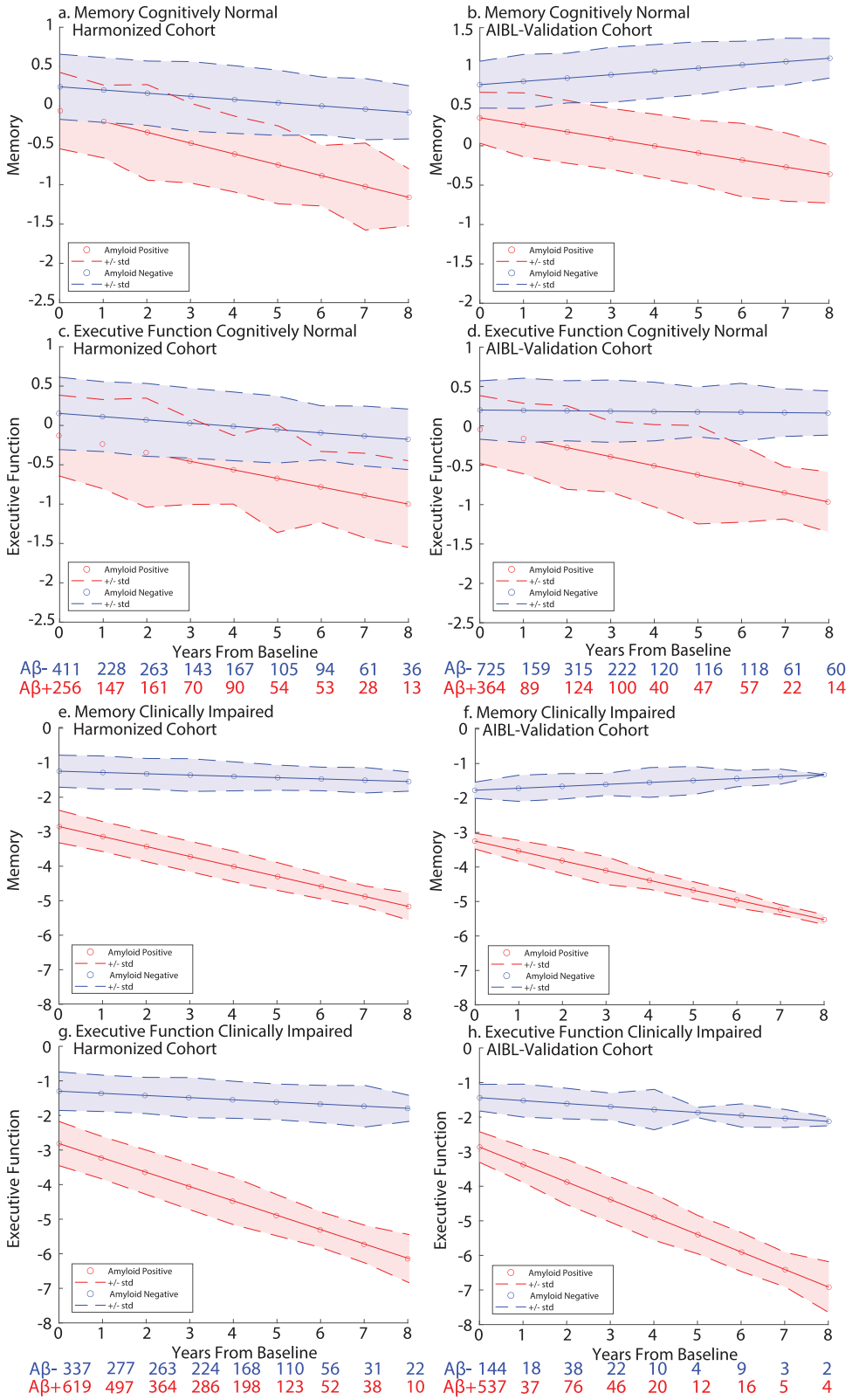

FIGURE 4 Executive function and memory trajectories. CN memory and executive function trajectories a, c. Harmonization cohort; b, d. AIBL Validation cohort. Expected trajectories for an Aβ− (blue) and an Aβ+ (red) 70-year-old women with an education of 14 years with no cognitive impairment (i.e., CN) at baseline. Error bars represent standard deviation of the residual of the fit from the LME model. Numbers below the plots show the sample size at different years of follow-up. Clinically impaired memory and executive function trajectories. e, g, Harmonization cohort; f, h, AIBL Validation cohort. Expected trajectories for an Aβ− (blue) and an Aβ+ (red) 70-year-old woman with an education of 14 years with cognitive impairment (i.e., MCI or AD) at baseline. AD, Alzheimer's disease; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; CN, cognitively normal; LME, linear mixed effects; MCI, mild cognitive impairment

to Aβ-related cognitive decline both in the preclinical and prodromal stages of AD. Critically, we show that this Aβ-related decline is robust across cohorts and has better or equal performance compared to the PACC. Furthermore, we show that the CC-ACC is also sensitive to a fine-grained pathological staging based on both Aβ positivity and tau severity at baseline. These findings provide strong evidence that our harmonization approach can be used to derive specific and sensitive cognitive composites across very different datasets. Our easy-to-implement tool can be used to test hypotheses and validate models across cohorts as well as retrospectively pool samples to improve statistical power and detect subtle treatment effects on cognition. We provide code and instructions to readily apply our approach on new datasets (https://github.com/jjgiorgio/cognitive_harmonisation)

We used non-parametric k-NN to impute differentially sampled neuropsychological data between cohorts, replicating disease-related covariance patterns across all cohorts while achieving good predictions when imputing held out data. The result of this is a full set of 125 cognitive variables for > 18,000 neuropsychological assessments. While previous work has used parametric imputation approaches on similar cognitive data to impute missing data,[24,26] these approaches are limited by the structure of missing data and may have poor performance due to differential sampling across cohorts. A recent non-parametric approach showed good performance harmonizing some variables between AIBL and ADNI data;[35] our work differs as we have integrated multiple cohorts with varying degrees of overlap resulting in a 3-fold increase in variables available for subsequent imputation. In addition, we extend previous work[35] providing a benchmarked approach to use the resultant imputed data to derive the novel CC-ACC with all required code and data readily available in our standalone toolbox. Here, we used k-NN to overcome constraints on parametric approaches as it is less biased by the structure of the missing data (i.e., missing at random).[27] Further, we performed a stepwise approach imputing data between cohorts that have maximal overlap at each stage, aiming to mitigate possible issues when large proportions of data are missing.[36] To maintain a more accurate representation of the data structure of the imputed data, we optimized the number of neighbors (k) ensuring that a lower k was preferred.[37] This reliably reproduced covariance patterns between item-level variables, suggesting the harmonized data can be used in the empirical derivation of cognitive composites that leverage the covariance between variables. Together, we show k-NN provides a simple, robust, and computationally efficient approach to harmonize item-level neuropsychological data from independent cohorts regardless of size.

Using our global measure, the CC-ACC, we integrate multiple cognitive domains showing benchmarked sensitivity to Aβ-related decline across clinical and tau pathological stages in different cohorts. In addition, we show the component domains of the CC-ACC (i.e., executive function and memory) are also sensitive to Aβ-related change. Our approach provides a substantial advantage over previous harmonization approaches, which assume a single driving latent factor of cognition and use this as the common scale to harmonize cohorts.[38] Further, our strategy of deriving component scores from a combination of tests may also improve the ability to detect relationships between cognitive domains and their biological underpinnings, which can be affected by the degree of missingness when using item-level analysis.[39] Recent approaches have included multiple domains in their harmonization; however, the selection of tests and anchor points for these domains requires expert assessment and restricts the freedom to interrogate cognitive variables on a more granular level.[40]

The CC-ACC calls on methods used to derive the PACC[10,11] but differs in three main regards. First, the CC-ACC takes all possible item-level variables (real and imputed) and derives a single factor for either memory (2 vs. 25 scores), executive function (1 vs. 21 scores), and general cognition (1 vs. 4 scores). Second, the CC-ACC does not apply a double weight to the memory domain. Previous work has shown that cognitive composites that increase the weighting of executive[7,12,13] and general[41] domains increases sensitivity to AD-related change. Such findings come with the caveat that poor performance on executive function tests may be initially driven by the pathological short-term memory loss.[42] Finally, the CC-ACC does not choose different item-level variables across different cohorts. This variability in variable selection makes the PACC difficult to interoperate and may introduce artificial cohort differences.[11]

Our harmonization approach has several strengths. First, we have developed and benchmarked an approach for researchers to interoperate their cognitive data and thus greatly improve their statistical power. Using our approach to derive a set of cognitive variables covering multiple cognitive domains researchers can validate, derive, and benchmark different cognitive composites. This alleviates a major restriction in clinical research as cohorts are now able to be merged for neuroimaging data and cognitive outcomes. Although standardization in neuropsychological testing protocols is being promoted,[43,44] our approach offers flexible harmonization between pre-existing and future research cohorts. Second, our harmonization software provides a way to test the efficacy of imputation, by (1) assessing the variance explained in a holdout variable and (2) calculating the covariance between real and imputed tests, and comparing these for the target (i.e., a new sample) and reference cohort. Knowing the minimum amount of data required to impute across cohorts is difficult as the efficacy of imputation will not only be determined by the absolute number of variables acquired but also the domains and ranges of available data.[37] As such we have not attempted to empirically derive the optimal or minimum data required to effectively impute data, but suggest researchers confirm the fidelity of imputation. Finally, we have shown using a simple derivation procedure based on existing literature that our harmonized data are sensitive to AD-related change.[10] However, any number of targeted data-driven approaches can be applied following the derivation of our harmonized data set. These may include the derivation of single or multiple domain cognitive composites, or mining of paired cognitive and biological (i.e., imaging, biofluid, or genetic) data to assess relationships in healthy aging or AD,[45,46] as well as modeling the temporal course of the disease.[47,48]

We consider the following potential limitations to our approach. First, we combined data for each domain using a variance-normalized mean, deriving a single factor score. This approach was not psychometrically driven and ignores potential subdomain factors and

differential scaling of individual variables that could be accounted for using latent modeling approaches.[49,50] However, our harmonization approach replicates item-to-item covariance, suggesting the harmonized data are well suited for deriving psychometrically derived composites which may track changes in cognitive domains due to aging[51] or AD pathology.[52] Second, prior to imputation we scaled several similar but not identical scores. This approach has been used in previous harmonization studies;[53] however, the scaled values may not exactly replicate the underlying score.[54] Therefore, a more intensive pre-harmonization approach may improve performance.[55] Our strategy of deriving factor scores based on the covariance between tests may ameliorate some of this random variability in single variable alignment. Third, our derivation of the cognitive composite, analogous to the PACC, did not take into account potential practice effects, which are known to show differences in Aβ-related cognitive decline.[56] Future work using psychometric approaches such as item response theory applied to the harmonized cognitive data could explicitly model such practice effects and capture more fine-grained item-level weighting that may more closely track AD-related cognitive decline.[57–60] Fourth, despite their widespread use, our reference benchmark tests have intrinsic limitations themselves. We note that all cohorts in our analysis included the MMSE, suggesting that the test remains prevalent in research use; however, the MMSE has substantial limitations such as ceiling effects and proprietary restrictions.[61] Notwithstanding, it has remained a comparator benchmark for the performance of other composite scores such as ADNI-Mem[15] and PACC.[11] Further, we note although the PACC is a current standard cognitive composite it is a non-psychometrically derived composite score.[10] Finally, test and subtest variables that are commonly used in AD research may be missing from our harmonized cohort. Our approach is easily built on as more data are made available, allowing us to expand the library of neuropsychological variables as well as increase sample heterogeneity.

Composite cognitive scores across cohorts with diverse neuropsychological source variables are an important step forward. We hope that composites such as CC-ACC will facilitate the generation and testing of hypotheses in independent datasets, improve trials design through validation of *in silico* models and stratification tools, and enable pooling across cohorts of data from under-represented groups.

AUTHOR CONTRIBUTIONS

Joseph Giorgio: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review & editing. Ankeet Tanna: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review & editing. Maura Malpetti: data curation, formal analysis, investigation, writing—review & editing. Simon R. White: formal analysis, investigation, writing—review & editing. Jingshen Wang: formal analysis, investigation, writing—review & editing. Suzanne Baker: data curation, formal analysis, investigation, writing—review & editing. Susan Landau: data curation, formal analysis, investigation, writing—review & editing. Tomotaka Tanaka: data curation, formal analysis, investigation, writing—review & editing. Christopher Chen: data curation, formal analysis, investigation, writing—review & editing. James B Rowe: data curation, formal analysis, investigation, writing—review & editing. John O'Brien: data curation, formal analysis, investigation, writing—review & editing. Jurgen Fripp: data curation, formal analysis, investigation, writing—review & editing. Michael Breakspear: data curation, formal analysis, investigation, writing—review & editing. William Jagust: conceptualization, data curation, investigation, writing—original draft, writing—review & editing. Zoe Kourtzi: conceptualization, investigation, methodology, writing—original draft, writing—review & editing.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This work was supported by grants to Z.K. from the Biotechnology and Biological Sciences Research Council (H012508 and BB/P021255/1), Alan Turing Institute (TU/B/000095), Wellcome Trust (205067/Z/16/Z and 221633/Z/20/Z) and to Z.K. and W.J. from the Global Alliance. J.B.R. receives funding from the Wellcome Trust (220258) and Medical Research Council (SUAG 092 G116768). M.M. receives funding from Race Against Dementia Alzheimer's Research UK (ARUK-RADF2021A-010). J.G. receives funding from the Alzheimer's Association (23AARF-1026883). For the purpose of open access, the author has applied for a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests. Author disclosures are available in the supporting information.

COLLABORATORS

ADNI Acknowledgement:

https://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

REFERENCES

1. Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. *N Engl J Med*. 2013;368:1169-1171. doi:10.1056/nejmp1302513
2. Albert M, Soldan A, Gottesman R, et al. Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to ApoE genotype. *Curr Alzheimer Res*. 2014;11:773-784. doi:10.2174/1567205011081409101219 20
3. Ellis KA, Bush AI, Darby D, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatrics*. 2009;21:672-687. doi:10.1017/S1041610209009405
4. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201. doi:10.1212/WNL.0B013E3181CB3E25
5. Sager MA, Hermann B, La Rue A. Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's prevention. *J Geriatr Psychiatry Neurol*. 2005;18:245-249. doi:10.1177/0891988705281882
6. Storandt M, Balota DA, Aschenbrenner AJ, Morris JC. Clinical and psychological characteristics of the initial cohort of the Dominantly Inherited Alzheimer Network (DIAN). *Neuropsychology*. 2014;28:19-29. doi:10.1037/NEU0000030
7. Jonaitis EM, Koscik RL, Clark LR, et al. Measuring longitudinal cognition: individual tests versus composites. *Alzheimers Dement (Amst)*. 2019;11:74-84. doi:10.1016/j.dadm.2018.11.006
8. Langbaum JB, Hendrix SB, Ayutyanont N, et al. An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimers Dement*. 2014;10:666-674. doi:10.1016/J.JALZ.2014.02.002
9. Papp K V., Buckley R, Mormino E, et al. Clinical meaningfulness of subtle cognitive decline on longitudinal testing in preclinical AD. *Alzheimers Dement*. 2020;16:552-560. doi:10.1016/j.jalz.2019.09.074
10. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol*. 2014;71:961-970. doi:10.1001/jamaneurol.2014.803
11. Insel PS, Weiner M, Scott MacKin R, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. *Neurology*. 2019;93:E322-E333. doi:10.1212/WNL.0000000000007831
12. Lim YY, Snyder PJ, Pietrzak RH, et al. Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: introducing the Z-scores of attention, verbal fluency, and episodic memory for Nondemented older adults composite score. *Alzheimers Dement (Amst)*. 2016;2:19-26. doi:10.1016/j.dadm.2015.11.003
13. Papp KV, Rentz DM, Orlovsky I, Sperling RA, Mormino EC. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5. *Alzheimers Dement (N Y)*. 2017;3:668-677. doi:10.1016/j.trci.2017.10.004
14. Petersen RC, Wiste HJ, Weigand SD, et al. Association of elevated amyloid levels with cognition and biomarkers in cognitively normal people from the community. *JAMA Neurol*. 2016;73:85-92. doi:10.1001/jamaneurol.2015.3098
15. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516. doi:10.1007/s11682-012-9186-z
16. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527. doi:10.1007/s11682-012-9176-1
17. Baker JE, Lim YY, Pietrzak RH, et al. Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: a meta-analysis. *Alzheimers Dement (Amst)*. 2017;6:108-121. doi:10.1016/j.dadm.2016.09.002
18. Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. *Annu Rev Clin Psychol*. 2013;9:61-89. doi:10.1146/ANNUREV-CLINPSY-050212-185522
19. Siddique J, Reiter JP, Brincks A, Gibbons RD, Crespi CM, Brown CH. Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis. *Stat Med*. 2015;34:3399-3414. doi:10.1002/SIM.6562
20. Griffith LE, Van Den Heuvel E, Fortier I, et al. Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported. *J Clin Epidemiol*. 2015;68:154-162. doi:10.1016/J.JCLINEPI.2014.09.003
21. Roth PL. Missing data: a conceptual review for applied psychologists. *Pers Psychol*. 1994;47:537-560. doi:10.1111/J.1744-6570.1994.TB01736.X
22. Dai S. Handling missing responses in psychometrics: methods and software. *Psych*. 2021;3:673-693. doi:10.3390/PSYCH3040043
23. Harel O, Zhou XH. Multiple imputation for the comparison of two screening tests in two-phase Alzheimer studies. *Stat Med*. 2007;26:2370-2388. doi:10.1002/SIM.2715
24. Burns RA, Butterworth P, Kiely KM, et al. Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. *J Clin Epidemiol*. 2011;64:787-793. doi:10.1016/j.jclinepi.2010.10.011
25. Godin J, Keefe J, Andrew MK. Handling missing Mini-Mental State Examination (MMSE) values: results from a cross-sectional long-term-care study. *J Epidemiol*. 2017;27:163. doi:10.1016/J.JE.2016.05.001
26. Rawlings AM, Sang Y, Sharrett AR, et al. Multiple imputation of cognitive performance as a repeatedly measured outcome. *Eur J Epidemiol*. 2017;32:55-66. doi:10.1007/s10654-016-0197-8
27. Choi HS, Choe JY, Kim H, et al. Deep learning based low-cost high-accuracy diagnostic framework for dementia using comprehensive neuropsychological assessment profiles. *BMC Geriatr*. 2018;18:234. doi:10.1186/s12877-018-0915-z
28. Dodge HH, Zhu J, Lee CW, Chang CCH, Ganguli M. Cohort effects in age-associated cognitive trajectories. *J Gerontol A Biol Sci Med Sci*. 2014;69:687-694. doi:10.1093/GERONA/GLT181
29. Chai YL, Chong JR, Raquib AR, et al. Plasma osteopontin as a biomarker of Alzheimer's disease and vascular cognitive impairment. *Sci Reports*. 2021;11:1-11. doi:10.1038/s41598-021-83601-6
30. Bevan-Jones WR, Surendranathan A, Passamonti L, et al. Neuroimaging of Inflammation in Memory and Related Other Disorders (NIMROD) study protocol: a deep phenotyping cohort study of the role of brain inflammation in dementia, depression and other neurological illnesses. *BMJ Open*. 2017;7:13187. doi:10.1136/bmjopen-2016-013187
31. Zhuang K, Chen X, Cassady KE, Baker SL, Jagust WJ. Metacognition, cortical thickness, and tauopathy in aging. *Neurobiol Aging*. 2022;118:44-54. doi:10.1016/J.NEUROBIOLAGING.2022.06.007
32. Farrell ME, Jiang S, Schultz AP, et al. Defining the lowest threshold for amyloid-PET to predict future cognitive decline and amyloid accumulation. *Neurology*. 2021;96:e619-e631. doi:10.1212/wnl.0000000000011214
33. Maass A, Landau S, Baker SL, et al. Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *Neuroimage*. 2017;157:448-463. doi:10.1016/J.NEUROIMAGE.2017.05.058
34. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1-67. doi:10.18637/jss.v045.i03
35. Shishegar R, Cox T, Rolls D, et al. Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI. *Sci Reports*. 2021;11:1-11. doi:10.1038/s41598-021-02827-6
36. Khoshgoftaar TM, Van Hulse J. Imputation techniques for multivariate missingness in software measurement data. *Softw Qual J*. 2008;16:563-600. doi:10.1007/s11219-008-9054-7

37. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak*. 2016;16:197-208. doi:10.1186/S12911-016-0318-Z/TABLES/5
38. Gross AL, Sherva R, Mukherjee S, et al. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. *Neuroepidemiology*. 2014;43:194-205. doi:10.1159/000367970
39. Nichols E, Deal JA, Swenor BK, et al. The effect of missing data and imputation on the detection of bias in cognitive testing using differential item functioning methods. *BMC Med Res Methodol*. 2022;22:1-12. doi:10.1186/S12874-022-01572-2/FIGURES/6
40. Mukherjee S, Choi S-E, Lee ML, et al. Cognitive domain harmonization and cocalibration in studies of older adults. *Neuropsychology*. 2023;37(4):409-423. doi:10.1037/NEU0000835
41. Knopman DS, Beiser A, Machulda MM, et al. Spectrum of cognition short of dementia: framingham heart study and mayo clinic study of aging. *Neurology*. 2015;85:1712-1721. doi:10.1212/WNL.0000000000002100
42. Jahn H. Memory loss in Alzheimer's disease. *Dialogues Clin Neurosci*. 2013;15:445-454. doi:10.31887/DCNS.2013.15.4/HJAHN
43. Costa A, Bak T, Caffarra P, et al. The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: consensus document of the joint Program for neurodegenerative diseases working group. *Alzheimers Res Ther*. 2017;9:27. doi:10.1186/S13195-017-0254-X
44. Boccardi M, Monsch AU, Ferrari C, et al. Harmonizing neuropsychological assessment for mild neurocognitive disorders in Europe. *Alzheimers Dement*. 2022;18:29-42. doi:10.1002/ALZ.12365
45. Oschwald J, Guye S, Liem F, et al. Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev Neurosci*. 2019;31:1-57. doi:10.1515/REVNEURO-2018-0096/ASSET/GRAPHIC/J_REVNEURO-2018-0096_FIG_003.JPG
46. Hu Y, Zhang Y, Zhang H, et al. Cognitive performance protects against Alzheimer's disease independently of educational attainment and intelligence. *Mol Psychiatry*. 2022;27:1-10. doi:10.1038/s41380-022-01695-4
47. Ashford JW, Shan M, Butler S, Rajasekar A, Schmitt FA. Temporal quantification of Alzheimer's disease severity: "time index" model. *Dementia* 1995;6:269-280. doi:10.1159/000106958
48. Ashford JW, Schmitt FA. Modeling the time-course of Alzheimer dementia. *Curr Psychiatry Rep*. 2001;3:20-28. doi:10.1007/s11920-001-0067-1
49. Loehlin JC. Latent variable models: an introduction to factor, path, and structural equation analysis. 4th ed. Psychology Press; 2003:1-290. doi:10.4324/9781410609823/LATENT-VARIABLE-MODELS-JOHN-LOEHLIN
50. McNeish D, Wolf MG. Thinking twice about sum scores. *Behav Res Methods*. 2020;52:2287-2305. doi:10.3758/S13428-020-01398-0/FIGURES/12
51. Glisky EL. Changes in cognitive function in human aging. *Brain Aging*. 2007:3-20. doi:10.1201/9781420005523-1
52. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. *Alzheimers Dement (N Y)*. 2018;4:64-75. doi:10.1016/J.TRCI.2018.01.003
53. Bath PA, Deeg D, Poppelaars J. The harmonisation of longitudinal data: a case study using data from cohort studies in The Netherlands and the United Kingdom. *Ageing Soc*. 2010;30:1419-1437. doi:10.1017/S0144686X1000070X
54. Gross AL, Inouye SK, Rebok GW, et al. Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. *J Clin Exp Neuropsychol*. 2012;34:758. doi:10.1080/13803395.2012.681628
55. Briceño EM, Gross AL, Giordani BJ, et al. Pre-statistical considerations for harmonization of cognitive instruments: harmonization of ARIC, CARDIA, CHS, FHS, MESA, and NOMAS HHS public access. *J Alzheimers Dis*. 2021;83:1803-1813. doi:10.3233/JAD-210459
56. Hassenstab J, Ruvolo D, Jasielec M, Xiong C, Grant E, Morris JC. Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*. 2015;29:940-948. doi:10.1037/NEU0000208
57. Ashford JW, Kolm P, Colliver JA, Bekian C, Hsu LN. Alzheimer patient evaluation and the Mini-Mental State: Item characteristic curve analysis. *Journals Gerontol*. 1989;44:P139-P146. doi:10.1093/geronj/44.5.P139
58. Calderón C, Beyle C, Véliz-Garciá O, Bekios-Calfa J. Psychometric properties of Addenbrooke's Cognitive Examination III (ACE-III): an item response theory approach. *PLoS One*. 2021;16:e0251137. doi:10.1371/journal.pone.0251137
59. Balsis S, Unger AA, Benge JF, Geraci L, Doody RS. Gaining precision on the Alzheimer's disease assessment scale-cognitive: a comparison of item response theory-based scores and total scores. *Alzheimers Dement*. 2012;8:288-294. doi:10.1016/j.jalz.2011.05.2409
60. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. *Alzheimer's Res Ther*. 2015;7:64. doi:10.1186/s13195-015-0151-0
61. Ashford JW. Screening for memory disorders, dementia and Alzheimer's disease. *Aging Health*. 2008;4:399-432. doi:10.2217/1745509X.4.4.399

SUPPORTING INFORMATION

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

How to cite this article: Giorgio J, Tanna A, Malpetti M. A robust harmonization approach for cognitive data from multiple aging and dementia cohorts. *Alzheimer's Dement*. 2023;15:e12453. https://doi.org/10.1002/dad2.12453