Received: 2 November 2022 | Revised: 23 March 2023 | Accepted: 29 March 2023

DOI: 10.1002/dad2.12433

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Development of a latent dementia index in the aging, demographics, and memory study: Validation and measurement invariance by sex

Joseph Saenz[1] | **Christopher R. Beam**[2,3] | **Alice J. Kim**[2,3]

[1]Edson College of Nursing and Health Innovation, Arizona State University, Phoenix, Arizona, USA

[2]Davis School of Gerontology, University of Southern California, Los Angeles, California, USA

[3]Department of Psychology, University of Southern California, Los Angeles, California, USA

Correspondence
Joseph Saenz, Ph.D. Edson College of Nursing and Health Innovation, Arizona State University, 500 N 3rd Street, Phoenix, AZ 85004, USA.
Email: joseph.l.saenz@asu.edu

Funding information
National Institute on Aging, Grant/Award Numbers: T32 AG000037, R00AG058799, R01AG060470, R01AG063949, U01AG09740; Alzheimer's Association, Grant/Award Number: AARF-17-505302

Abstract

Latent variable models can create a latent dementia index (LDI) using cognitive and functional ability to approximate dementia likelihood. The LDI approach has been applied across diverse cohorts. It is unclear whether sex affects its measurement properties. We use Wave A (2001–2003) of the Aging, Demographics, and Memory Study (n = 856). Multiple group confirmatory factor analysis (CFA) was used to test measurement invariance (MI) using informant-reported functional ability and cognitive performance tasks, which we group into verbal, nonverbal, and memory. Partial scalar invariance was found, allowing for testing sex differences in LDI means (M_{Diff} = 0.38). The LDI correlated with consensus panel dementia diagnosis, Mini-Mental State Examination (MMSE), and dementia risk factors (low education, advanced age, and apolipoprotein ε4 [APOE-ε4] status) for men and women. The LDI validly captures dementia likelihood to permit estimation of sex differences. LDI sex differences indicate higher dementia likelihood in women, potentially due to social, environmental, and biological factors.

KEYWORDS
dementia, HRS, latent variable, measurement, sex differences

1 | BACKGROUND

Women represent two-thirds of Alzheimer's disease cases, with lifetime risk doubling that of men.[1,2] Sex differences may be explained by environmental and biological factors including genetic and endocrinological differences like APOE ε4,[3] hormonal differences,[4] and longevity.[5] Sex differences in Alzheimer's disease and related dementias (ADRD) lead to questions of whether cognitive/functional manifestations of ADRD are sex invariant. For example, measurement non-invariance of functional ability was found across sex.[6] Women's higher verbal ability may also reflect cognitive reserve withstanding neuropathology.[7,8]

Latent variable models quantify dementia likelihood into a latent dementia index (LDI) using cognitive and functional ability assessments that are reliable and valid measures of dementia likelihood.[9–11] Comparing mean differences in dementia likelihood in an LDI requires measurement of functional/cognitive ability to be identical across sex, known as measurement invariance (MI). Configural invariance tests whether model structure is equivalent across sex. Metric invariance tests whether the latent variable capturing dementia likelihood correlates with cognitive/functional ability assessments similarly across sex. Scalar invariance determines whether expected values of assessments are identical across sex and is required to test latent mean differences.

We know of no studies testing MI across sex in latent dementia models. Some studies observed strict MI for memory measures across sex[12] whereas others found partial scalar invariance across sex and ethnicity.[13] Similarly, the Colorado Cognitive Assessment did not show sex biases.[14] Yet, women's verbal ability has been found to be higher for longer periods of older adulthood than men's, possibly reflecting cognitive reserve advantages.[7] Among older adults with mild cognitive impairment, women had higher verbal memory than equally impaired men.[8] These studies, along with measurement non-invariance of functional ability across sex,[6] suggest that MI of cognitive/functional items related with dementia needs to be examined to ensure measurement equivalence.

1.1 | The health and retirement study

The Health and Retirement Study (HRS) and the Aging, Demographics, and Memory Study (ADAMS) subsample are instrumental in dementia research in the United States, including estimating dementia prevalence rates.[15] The HRS is a biennial nationally representative study of individuals aged 50+.[16] The HRS sample is large, with approximately 43,000 individuals interviewed from 1992–2016,[17] making dementia ascertainment challenging. Dementia diagnosis relies on neuropsychological and functional ability assessment, which can take several hours and require clinical assay of cerebrospinal fluid or time-intensive neuroimaging. This makes true clinical diagnosis of dementia cost prohibitive. Instead, the HRS collects cognitive performance data through immediate/delayed recall of 10-word lists, serial 7 subtraction, and backwards counting, and various other assessments described elsewhere.[18] Functional data are assessed through activities of daily living (ADL) and instrumental ADL (IADL) limitations.

Algorithms have been developed to differentiate cognitive impairment and dementia from normal cognitive function using available HRS measures. Langa-Weir[19] assigned cut-points to a cognitive score summing participants' performance across several cognitive tasks. Classification of proxies is based on proxy assessment of memory, IADLs, and interviewer assessment of difficulty completing interviews due to cognitive limitation.[20] Existing algorithmic approaches to HRS dementia risk scoring[21] often use only cognitive tasks. Latent dementia approaches incorporate the day-to-day functional ability to better characterize disease processes[22] that, also, are unbiased by measurement error.

1.2 | The current study

Using the HRS ADAMS subsample, we fit a multigroup single latent variable model to participants' cognitive/functional ability assessments. We first test MI across sex expecting that a partial scalar invariance model would fit the data the best. Second, we test the LDI's validity using ADAMS clinical diagnoses to ensure that it is an equal measure across sex, correlates with diagnosis, and can reproduce sex differences in dementia observed in prior studies.[1,5]

RESEARCH IN CONTEXT

1. **Systematic review**: Authors reviewed literature on latent variable measures of dementia, sex differences in dementia, and sex differences in measurement of cognitive/functional ability. Studies have either validated latent variable measures of dementia or evaluated how sex affects measurement of cognitive and functional ability. We did not find studies testing measurement invariance (MI) of latent variable measures of dementia across sex.
2. **Interpretation**: Our research shows support for partial scalar invariance of a latent dementia index (LDI) that can capture dementia likelihood equivalently for men and women.
3. **Future directions**: The manuscript lays groundwork for constructing LDI measures to test group differences in other studies. This study uses a subsample of a large nationally representative study of aging (the Health and Retirement Study), which is part of an international network of aging studies. Our model, thus, will facilitate cross-cultural comparisons of dementia and sex differences in dementia across the globe.

2 | METHODS

2.1 | Participants

We use data from Wave A of the ADAMS, a subsample of 856 adults aged 70+ years from the 2000 and 2002 waves of the HRS selected based on their cognitive scores. Between 2001 and 2003, ADAMS assessments were conducted in-person in respondents' homes by a nurse and neuropsychology technician. Assessments included neuropsychological testing of respondents and informant-reported information including functional ability, history of cognitive symptoms, history of medical conditions, and current medications, among other domains.[23] Although additional follow-ups of the ADAMS sample have been conducted, we use the baseline (Wave A) due to larger sample size and less vulnerability to biases from loss to follow-up and selective survival in later follow-up waves of the Wave A cohort. Furthermore, follow-up waves (Waves B–D) were conducted on select subsamples of the Wave A cohort such as those with cognitive impairment not dementia (CIND) and those without a prior diagnosis of dementia.[24] Wave A, thus, provides the most complete picture of dementia likelihood in the population.

2.2 | Cognitive assessment

ADAMS included an extensive battery of neuropsychological tests to ascertain cognitive function across cognitive domains. We use the

following as cognitive measures in our latent measure of dementia (Figure 1).

2.2.1 | Nonverbal subtests

Constructional Praxis (CPT), Delayed Constructional Praxis (DCPT), Recognition Constructional Praxis (RCPT), Trail Making Test A and B (TMTA, TMTB), Digit Span (DS), and Digital Symbol Modality (DSM).

2.2.2 | Verbal subtests

Animal Fluency (AF), Boston Naming Test (BNT), Controlled Oral Word Association Test (COWAT), Wide Range Achievement Test (WRAT), and Shipley Vocabulary Test (SHIP).

2.2.3 | Memory subtests

Benton Visual Retention (BENT), Immediate/Delayed/Recognition Word Recall (WLI, WLD, WLR), Wechsler Memory Scale (WMS) Logical Memory I (LM1), and WMS Logical Memory II (LM2). The Fuld Object Memory Evaluation (FULD) is based on recall of ten common household objects, which the subject first identifies by touch.

2.3 | Functional ability

Functional ability was reported by knowledgeable informants. Functional ability included ADL and IADL limitations. ADL limitations included difficulty getting across a room, dressing, bathing, eating, getting out of bed, and using the toilet. IADL limitations included difficulty preparing meals, shopping for groceries, making telephone calls, taking medication, and managing money. We created two latent variables for ADL and IADL limitations.

2.4 | Dementia diagnosis

Respondents were classified as having normal cognitive function, CIND, and dementia. Preliminary diagnoses were completed by a panel including a Duke University geropsychiatrist, neurologist, neuropsychologist, and cognitive neuroscientist using all information from the ADAMS in-home assessment. The geropsychiatrist then reviewed medical records and revised the preliminary diagnosis where appropriate based on the additional medical information. Specific dementia etiologies were identified, and diagnostic procedures were based on standard diagnostic procedures used at the time, that is, the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R & DSM-IV).[25,26] CIND was based on the judgment of an expert consensus panel and considered broadly using mild functional or cognitive impairment not meeting criteria for dementia or neuropsychological test performance below expectation and more than 1.5 standard deviations below published norms. The final diagnoses were decided through an expert consensus panel consisting of neurologists, neuropsychologists, geropsychologists, and internists. More detail on the diagnostic process is available in other work.[23] In validation analyses, respondents were categorized into those diagnosed with any dementia and those not diagnosed with dementia (i.e., normal cognitive function and CIND respondents).

2.5 | Data analysis

We first present the means and standard deviations of cognitive and functional ability measures across sex. In a preliminary step, we estimated a two-factor model of functional ability using M*plus* 8.8[27] to test metric invariance of the two factors. This model was imported into the confirmatory factor model, so the LDI was fit to the normally distributed latent variables (Figure 1).

Multigroup confirmatory factor analysis (CFA) was used to estimate and compare LDI scores across men and women using M*plus*. The configural model served as the baseline model and is specified in the following expression:

$$\sum_{pp}^{k} = \Lambda_p^k \Phi^k \Lambda_p'^k + \Theta_{pp}^k \quad (1)$$

$$\mu_p^k = \tau_p^k + \Lambda_p^k k^k \quad (2)$$

Covariance among items, Σ_{pp}, is a linear function of the weighted effect, λ, of the latent variable, Φ, and the unique factor variance, Θ for k groups, here male and female. The latent variances of the LDI factor, Φ, are fixed to one in the configural model. In Φ, the covariances among the items are independent in except for the nonmemory cognitive items. This portion of the matrix is saturated to account for covariances among this subset of items due to a general ability factor. The mean structure of the model, Equation 2, is estimated so that the intercepts, τ_p, are free across items whereas the latent means, κ, are fixed to zero. Figure 1 presents the model fit to the male and female samples. Lower LDI scores indicate a higher likelihood of dementia. Several cognitive ability variables were rescaled to facilitate model estimation, which is described in the Appendix.

We next tested for full metric invariance followed by scalar invariance. If the difference between configural and metric invariance models (or metric invariance and scalar invariance models) were statistically different, we tested whether parcels of items were invariant (i.e., all verbal items, all nonverbal items, all memory items, and both functional ability variables). Of the various approaches used when some but not all measured variables are unbiased by group membership, this approach is compromissory.[28] As prior literature suggests sex differences in verbal ability and memory measures more often than in nonverbal ability, we tested nonverbal ability clusters first, verbal ability clusters second, memory clusters third, and functional ability items last. Distributions of male and female LDI scores are presented from the best-fitting model.

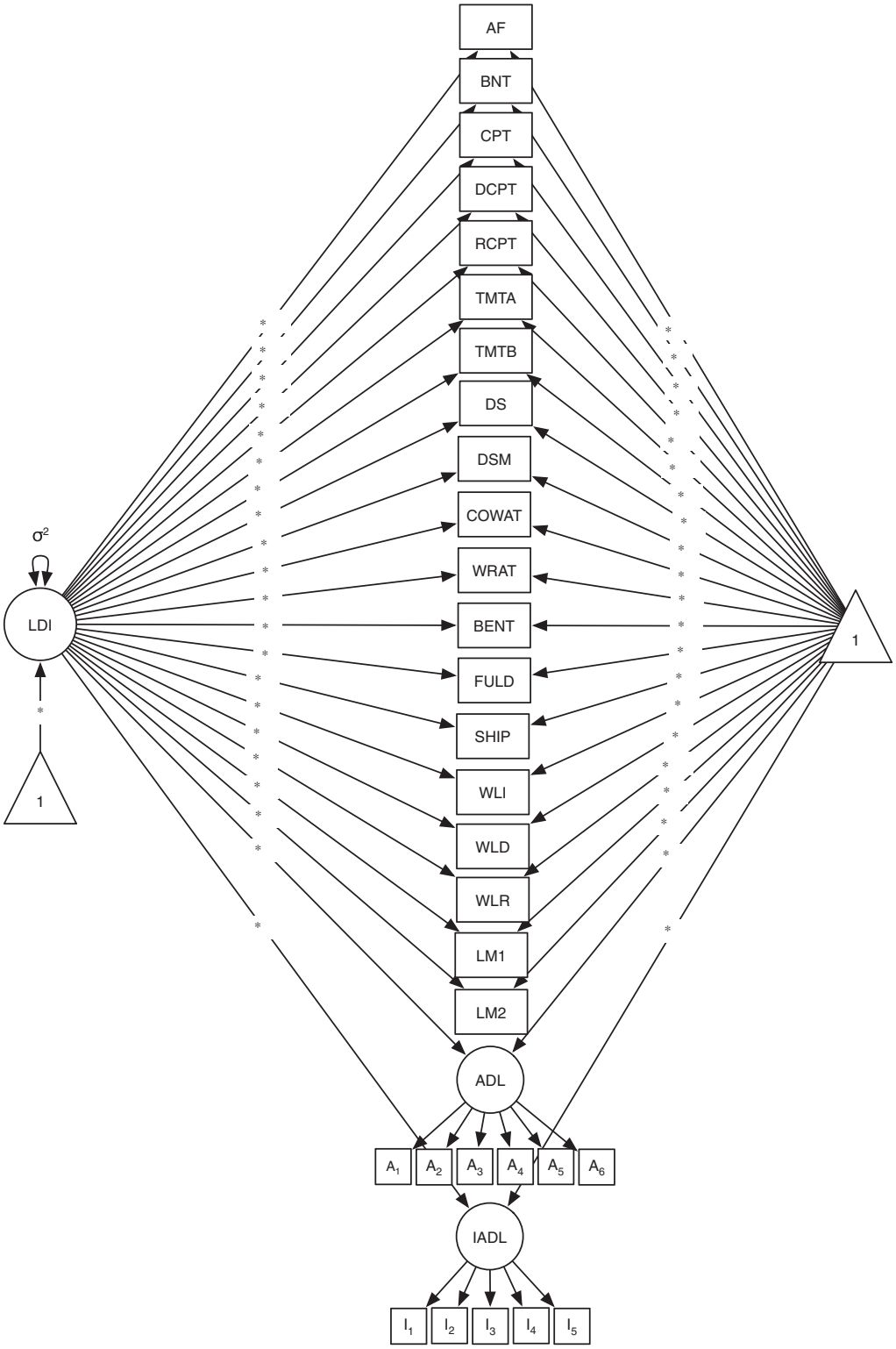

FIGURE 1 Confirmatory factor model for the latent dementia index. AF, Animal Fluency; BENT, Benton Visual Retention; BNT, Boston Naming Test; COWAT, Controlled Oral Word Association Test; CPT, Constructional Praxis Total; DS, Digit Span; DSM, Digit Symbol Modality; DCPT, Delayed Constructional Praxis Total; FULD, Fuld Object Memory Evaluation; LDI, Latent Dementia Index; LM1, Wechsler Memory Scale Logical Memory 1; LM2, Wechsler Memory Scale Logical Memory 2; RCPT, Recognition Constructional Praxis Total; TMTA, Trail Making Test (A); TMTB, Trail Making Test (B); SHIP, Shipley Vocabulary Test; WLD, Word List Delayed; WLI, Word List Immediate; WLR, Word List Recognition; WRAT, Wide Range Achievement Test; ADL, Activities of Daily Living measured using items $A_{1\text{-}6}$, which are difficulty getting across a room, dressing, bathing, eating, getting out of bed, and using the toilet; IADL, Instrumental Activities of Daily Living measured using items $I_{1\text{-}5}$, which are difficulty preparing meals, shopping for groceries, making telephone calls, taking medication, and managing money. Circles represent unobserved (latent) variables, rectangles indicate observed (manifest) variables, triangles represent means, and * indicates freely estimated parameters.

Finally, we estimated and tested differences in correlations among the LDI, dementia diagnosis, and Mini-Mental Status Exam (MMSE) scores. For all CFA and validation analyses, we used maximum likelihood with robust standard errors to account for violations of multivariate normality and missing data. Nonresponse was assumed to be missing at random or missing not at random; in both cases, full information maximum likelihood is the recommended approach for obtaining the least biased parameter estimates and standard errors, even in cases where nonresponse is missing not at random.[29] Models were compared using the Satorra–Bentler chi-squared difference test of nested models. More parsimonious models that did not statistically significantly differ from more complex models were accepted. Bonferroni-corrected α cut-off values were used to maintain a Type I error rate of 0.05. As we present in the Results section, we conducted 36 total tests with this data set, so we used a corrected α cut-off of 0.001. We additionally used the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to compare models, as each penalizes model overfitting differently – the BIC incurs a greater penalty per parameter[30]; with both methods, lower values suggest better model fit.

3 | RESULTS

3.1 | Descriptive analyses

Table 1 presents means and standard deviations of cognitive and functional ability measures for men and women in the ADAMS. The sample included 355 men and 501 women. A higher proportion of female participants (42.5%) were considered to have dementia compared to male participants (26.8%). The mean age for women (82.6) was slightly higher than for men (80.2).

3.2 | Confirmatory factor analyses and measurement invariance of latent dementia index

We first used multigroup CFA to test whether the ADL and IADL factors were invariant across sex. A metric invariant model fit equally as well as a configural model (S-Bχ^2 = 3.75, Δdf = 9, p = 0.927), and a scalar invariant model fit equally as well as a metric invariant model (S-Bχ^2 = 19.25, Δdf = 9, p = 0.023). The latent ADL and IADL means significantly differed between men and women, with women scoring 0.52 and 0.47 units lower than men, respectively, We, thus, proceeded with testing MI of the LDI across sex.

Table 2 presents the model-fitting results of the factorial invariance tests across sex. A metric invariant model did not fit equally as well as the configural model (Model 2 vs. 1: S-Bχ^2 = 45.24, Δdf = 20, p = 0.001) so we proceeded with testing for partial metric and scalar invariance. We found that all nonverbal item loadings could be constrained as equal across men and women (Model 2a vs. 1: S-Bχ^2 = 7.80, Δdf = 6, p = 0.253). Constraining all verbal item loadings across sex resulted in statistically significantly different model fit (Model 2b vs. 2a: S-Bχ^2 = 22.22, Δdf = 5, p < 0.001), but only the WRAT loading was significantly different (S-Bχ^2 = 14.42, Δdf = 1, p < 0.001) (see Table S1, which contains tests of individual items). All memory item loadings could be constrained across sex (Model 2c vs. 2a: S-Bχ^2 = 23.33, Δdf = 7, p = 0.001) as well as both the latent ADL and IADL factors (Model 2d vs. 2a: S-Bχ^2 = 0.97, Δdf = 2, p = 0.614). Model 2e in the LDI section of Table 2, thus, was accepted as the final partial metric invariant model in which only the WRAT statistically significantly differed across sex.

Next, we tested for partial scalar invariance of all items except the WRAT and the latent ADL and IADL factors, as the WRAT loading was not invariant across sex and the means (now intercepts) of the ADL and IADL factors were found to differ in the analysis of the ADL and IADL items in our initial analytic step (i.e., Model FA3 in Table 2). Whereas all intercepts that could be tested for scalar invariance could not be constrained across sex (Model 3 vs. 2e: S-Bχ^2 = 95.44, Δdf = 17, p < 0.001), we proceeded to test ability clusters as in the partial metric invariance step. The intercepts of all nonverbal items could be constrained as equal across sex (Model 3a vs. 2e: S-Bχ^2 = 11.49, Δdf = 6, p = 0.074). Although a model that constrained all verbal item intercepts to be the same across sex was rejected (Model 3b vs. 3a: S-Bχ^2 = 24.16, Δdf = 4, p < 0.001), no item was found to differ significantly when evaluated individually (see Table S1). We, thus, constrained all verbal item intercepts except the WRAT to be the same across sex. All memory intercepts could not be constrained across sex (Model 3c vs. 3a: S-Bχ^2 = 45.37, Δdf = 7, p < 0.001). When we tested individual memory items (see Table S1), immediate recall and delayed recall intercepts were found to differ between men and women whereas all other memory item intercepts did not. We thus selected Model 3d in Table 2 as the best fitting model that allowed the WRAT loading and intercept, the immediate and delayed recall intercepts, and the latent ADL and IADL intercepts to differ between men and women.

Parameter estimates from Model 3d are presented in Table 3. All nonverbal items were constrained across sex with the highest loadings among DCPT, DS, CPT, and digit symbol modality. Among the verbal items, AF and the BNT had the highest factor loadings common across sex whereas the factor loading for the WRAT was greater in men than women. All memory items loaded highly on the LDI with logical memory tests having the highest loadings. For functional ability, the latent IADL factor was more strongly associated with the LDI than the latent ADL factor.

3.3 | Validation and evaluation of the latent dementia index

The receiver operator characteristic curve (ROC) (Figure S1), which we constructed using LDI factor scores from Model 3d, suggests excellent sensitivity and specificity of the LDI (area under the curve value = 0.97) with a cut-off score of −0.65 that maximizes true positive rate (sensitivity) and minimizes false positive rate. The LDI cutoff score of −0.65 suggests that scores below this value are associated with probable diagnosis of dementia whereas scores above it are associated with no diagnosis. In support of the LDI's validity, its correlation with MMSE

TABLE 1 Descriptive characteristics of baseline aging, demographics, and memory study participants by gender ($n = 856$).

			Men ($n = 355$)		Women ($n = 501$)	
	n	Range				
Demographics and dementia						
Age (mean, SD)	856	70–110	80.2	6.8	82.6	7.2
Dementia diagnosis (*n*, %)	856	0–1	95	26.8	213	42.5
			Mean	**SD**	**Mean**	**SD**
Cognitive ability						
Word List Immediate	788	0–14.5	6.6	2.7	6.5	3.2
Word List Delayed	783	0–10	3.8	2.5	3.6	2.9
Word List Recognition	774	0–10	8.4	2.1	8.1	2.4
WMS Logical Memory 1	748	0–18.5	7.5	4.4	6.5	4.7
WMS Logical Memory 2	728	0–18.5	5.4	4.3	4.5	4.4
Animal Fluency	790	0–16.5	6.4	2.7	5.5	2.8
Boston Naming Test	801	0–7.5	6.2	1.5	5.6	1.7
Constructional Praxis Total	737	0–11	8.9	1.9	8.3	2.1
Constructional Praxis Delayed	732	0–11	5.9	3.5	4.6	3.4
Constructional Praxis Recognition	725	0–4	3.0	1.1	2.8	1.1
Trails A	688	−3.73 to −0.05	−0.8	0.6	−0.9	0.7
Trails B	370	−7.27 to −0.31	−1.7	0.9	−1.8	1.1
Shipley Vocabulary Test	529	0–4	2.8	0.8	2.7	0.8
Digit Span	730	0–14	6.1	2.1	5.7	2.2
Digit Symbol Modality	531	0–6.3	2.5	1.2	2.4	1.3
Word Association	673	0–6.6	2.4	1.2	2.3	1.3
Wide Range Achievement Test	414	0.5–5.7	4.0	1.1	4.2	0.9
Benton	644	0–9	3.5	2.4	3.0	2.2
Fuld Object Memory	701	0–15	9.9	3.4	9.2	4.4
MMSE	814	0–30	22.3	6.3	21.1	7.6
			n	**%**	***n***	**%**
Activities of Daily Living						
Difficulty getting across a room	731	0–1	62	20.3	152	35.7
Difficulty dressing	728	0–1	59	19.4	151	35.6
Difficulty bathing	731	0–1	66	21.6	180	42.3
Difficulty eating	725	0–1	34	11.2	77	18.3
Difficulty getting out of bed	729	0–1	43	14.1	129	30.4
Difficulty using the toilet	724	0–1	38	12.7	113	26.7
			n	**%**	***n***	**%**
Instrumental Activities of Daily Living						
Difficulty preparing a meal	702	0–1	76	26.3	185	44.8
Difficulty shopping for groceries	707	0–1	80	27.2	200	48.4
Difficulty making telephone calls	722	0–1	80	26.4	153	36.5
Difficulty taking medication	724	0–1	66	21.8	157	37.3
Difficulty managing money	718	0–1	62	20.7	164	39.2

Notes: Authors' own calculations using Wave A of the Aging, Demographics, and Memory Study. "Dementia diagnosis" made by consensus panels, see Langa et al.[23] for more details.

Abbreviations: MMSE, Mini-Mental State Examination; WMS, Wechsler Memory Scale.

TABLE 2 Model fit results.

Model	LL	Parameters	Comparison	S-Bχ^2	Δdf	p	AIC	BIC
Functional Ability								
FA1. Configural	−3002.52	47					6099.03	6315.36
FA2. Metric	−3004.24	38	FA2 vs. FA1	3.75	9	0.927	6084.48	6259.38
FA3. Scalar	−3013.88	29	FA3 vs. FA2	19.25	9	0.023	6085.76	6219.23
LDI								
1. Configural	−25275.26	327	–	–	–	–	51204.53	52755.06
2. Metric	−25301.05	307	2 vs. 1	45.24	20	0.001	51216.11	52671.81
Partial metric invariance								
2a. All nonverbal items constrained	−25280.22	321	2a vs. 1	7.80	6	0.253	51202.45	52724.53
2b. All verbal items constrained	−25291.25	316	2b vs. 2a	22.22	5	< 0.001	51214.51	52712.88
2c. All memory items constrained	−25293.26	314	2c vs. 2a	23.33	7	0.001	51214.51	52703.41
2d. All latent functional ability variables constrained	−25280.75	319	2d vs. 2a	0.97	2	0.614	51199.50	52712.10
2e. All items *except* WRAT	−25296.98	308	2e vs. 2a	29.94	13	0.005	51209.96	52670.40
3. Scalar	−25345.56	291	3 vs. 2e	95.44	17	< 0.001	51273.13	52652.96
Partial scalar invariance								
3a. All nonverbal intercepts constrained	−25303.12	302	3a vs. 2e	11.49	6	0.074	51210.24	52642.24
3b. All verbal intercepts constrained (except WRAT)	−25315.54	298	3b vs. 3a	24.16	4	< 0.001	51227.07	52640.10
3c. All memory intercepts constrained	−25325.23	295	3c vs. 3a	45.37	7	< 0.001	51240.46	52639.26
3d. All intercepts *except* WLI, WLD, ADL, IADL, & WRAT	−25324.65	293	3d vs. 3a	42.85	9	< 0.001	51235.31	52624.63

Abbreviations: LL, log-likelihood of model; ΔS-Bχ^2, Satorra–Bentler corrected χ^2 difference; Δdf, difference in degrees of freedom; ADL, Activities of Daily Living; AIC, Akaike information criterion; BIC, Bayesian information criterion; FA, Functional Ability; IADL, Instrumental Activities of Daily Living; LDI, Latent Dementia Index; WLD, Word List Delayed; WLI, Word List Immediate; WRAT, Wide Range Achievement Test.

is 0.84. The correlation between the LDI and a dichotomous dementia diagnosis (estimated through maximum likelihood estimation) is −0.79 (SE = 0.02) whereas the correlation between MMSE and diagnosis is −0.73 (SE = 0.02).

Under a partial scalar invariant model, the LDI mean was 0.38 units lower in the female group compared to males; the distribution of LDI scores is shifted slightly to the left in the female group (in blue) compared to the male group (in red; Figure 2). Associations between LDI scores and ADAMS-diagnosed cognitive status (Figure 3) shows that mean LDI scores are generally highest in the cognitively normal (CN) group, followed by the cognitively impaired not dementia group, and then lowest in the dementia group, regardless of sex. However, there are sex differences. The location of distributions for CN and cognitively impaired not demented male and female groups are reasonably comparable. Visual inspection of the distributions of the demented groups suggests women have higher dementia likelihoods than men, on average.

Figures S2A–S2C present sex differences of LDI distributions as a function of age, APOE ε4 allele count, and education as an additional validity check given expected differences in dementia across these domains.[31–33] LDI score distributions appear to be similar across age (Panel A) in males as females; those in the highest age groups had the lowest LDI scores, although the middle age group was bimodal for women but not men. Whereas the distributions of male participants with < high school education, a high school education, and > high school education are fairly symmetric, the distribution of female participants are bimodal or platykurtic suggesting dementia values are not well discriminated (Panel B). Finally, LDI scores were distributed differently across APOE ε4 allele counts for men and women (Panel C). The distributions were nearly identical for zero and one alleles in the men whereas two alleles conferred greater likelihood of dementia. For women, likelihood of dementia increased with each additional allele.

4 | DISCUSSION

Establishing partial scalar invariance across sex suggests that LDI scores lay on the same distribution for women as men; mean differences, thus, represent estimates of true population differences in dementia likelihood. Our establishment of metric invariance also enables meaningful comparison of associations between dementia risk factors and LDI scores across sex.

Although dementia diagnoses are available in ADAMS, there are many studies without diagnoses because they were never considered or were too cumbersome given the sample of study like the HRS. The LDI approach may be extended to the core HRS and the more recent Harmonized Cognitive Assessment Protocol (HCAP) subsample.[34] Furthermore, multigroup latent variable models can test

TABLE 3 Parameter estimates from best fitting partial scalar invariance model (Model 3d).

	Men		Women	
	Estimate	**0.95 CI**	**Estimate**	**0.95 CI**
Loadings				
Verbal				
Animal Fluency	2.09	[1.88, 2.30]	2.09	[1.88, 2.30]
Boston Naming	1.07	[0.95, 1.20]	1.07	[0.95, 1.20]
COWAT	0.88	[0.78, 0.99]	0.88	[0.78, 0.99]
Shipley	0.56	[0.48, 0.65]	0.56	[0.48, 0.65]
WRAT	0.67	[0.53, 0.81]	0.45	[0.35, 0.54]
Nonverbal				
Constructional Praxis	1.31	[1.11, 1.50]	1.31	[1.11, 1.50]
Constructional Praxis—Delayed	2.73	[2.46, 2.99]	2.73	[2.46, 2.99]
Constructional Praxis—Recognition	0.72	[0.64, 0.81]	0.72	[0.64, 0.81]
TMT A	0.28	[0.21, 0.35]	0.28	[0.21, 0.35]
TMT B	0.95	[0.73, 1.17]	0.95	[0.73, 1.17]
Digit Span	1.38	[1.21, 1.55]	1.38	[1.21, 1.55]
Digit Symbol	1.10	[0.97, 1.23]	1.10	[0.97, 1.23]
Memory				
Benton	1.77	[1.57, 1.98]	1.77	[1.57, 1.98]
Fuld	3.21	[2.90, 3.52]	3.21	[2.90, 3.52]
Word List—Immediate	2.57	[2.34, 2.80]	2.57	[2.34, 2.80]
Word List—Delayed	2.35	[2.15, 2.56]	2.35	[2.15, 2.56]
Word List—Recognition	1.27	[1.08, 1.46]	1.27	[1.08, 1.46]
Logical Memory A	4.09	[3.70, 4.48]	4.09	[3.70, 4.48]
Logical Memory B	3.81	[3.41, 4.21]	3.81	[3.41, 4.21]
Functional Ability				
ADL	−0.61	[−0.79, −0.44]	−0.61	[−0.79, −0.44]
IADL	−1.02	[−1.24, −0.81]	−1.02	[−1.24, −0.81]
Means/Intercepts				
LDI	0.00	–	−0.38	–
Animal Fluency	6.07	[5.82, 6.33]	6.07	[5.82, 6.33]
Boston Naming	6.07	[5.93, 6.21]	6.07	[5.93, 6.21]
COWAT	2.20	[2.08, 2.33]	2.20	[2.08, 2.33]
Shipley	2.43	[2.34, 2.52]	2.43	[2.34, 2.52]
WRAT	3.86	[3.72, 4.01]	4.07	[3.97, 4.17]
Constructional Praxis	8.58	[8.39, 8.78]	8.58	[8.39, 8.78]
Constructional Praxis—Delayed	5.23	[4.89, 5.56]	5.23	[4.89, 5.56]
Constructional Praxis—Recognition	2.88	[2.77, 2.98]	2.88	[2.77, 2.98]
TMTA	−0.85	[−0.0.91, −0.79]	−0.85	[−.0.91, −0.79]
TMTB	−2.60	[−2.83, −2.36]	−2.60	[−2.83, −2.36]
Digit Span	5.91	[5.72, 6.10]	5.91	[5.72, 6.10]
Digit Symbol	2.01	[1.87, 2.16]	2.01	[1.87, 2.16]
Benton	2.99	[2.75, 3.23]	2.99	[2.75, 3.23]
Fuld	9.39	[9.03, 9.76]	9.39	[9.03, 9.76]

(Continues)

TABLE 3 (Continued)

	Men		Women	
	Estimate	**0.95 CI**	**Estimate**	**0.95 CI**
Word List—Immediate	6.43	[6.15, 6.72]	7.10	[6.76, 7.43]
Word List—Delayed	3.63	[3.36, 3.90]	4.11	[3.81, 4.42]
Word List—Recognition	8.34	[8.16, 8.52]	8.34	[8.16, 8.52]
Logical Memory A	7.18	[6.72, 7.64]	7.18	[6.72, 7.64]
Logical Memory B	5.06	[4.62, 5.50]	5.06	[4.62, 5.50]
ADL	0.00	–	0.42	–
IADL	0.00	–	0.34	–

Abbreviations: ADL, Activities of Daily Living; COWAT, Controlled Oral Word Association Test; IADL, Instrumental Activities of Daily Living; LDI, Latent Dementia index; TMTA, Trail Making Test (A); TMTB, Trail Making Test (B); WRAT, Wide Range Achievement Test.

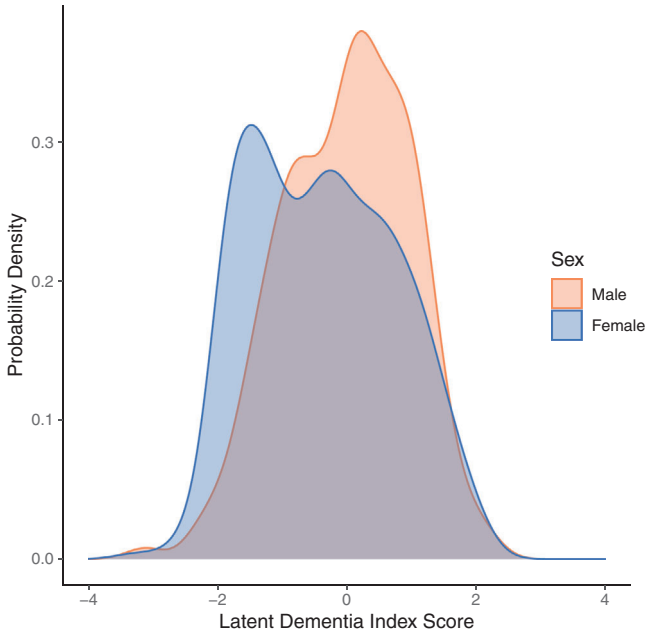

FIGURE 2 LDI distribution for male and female ADAMS participants.

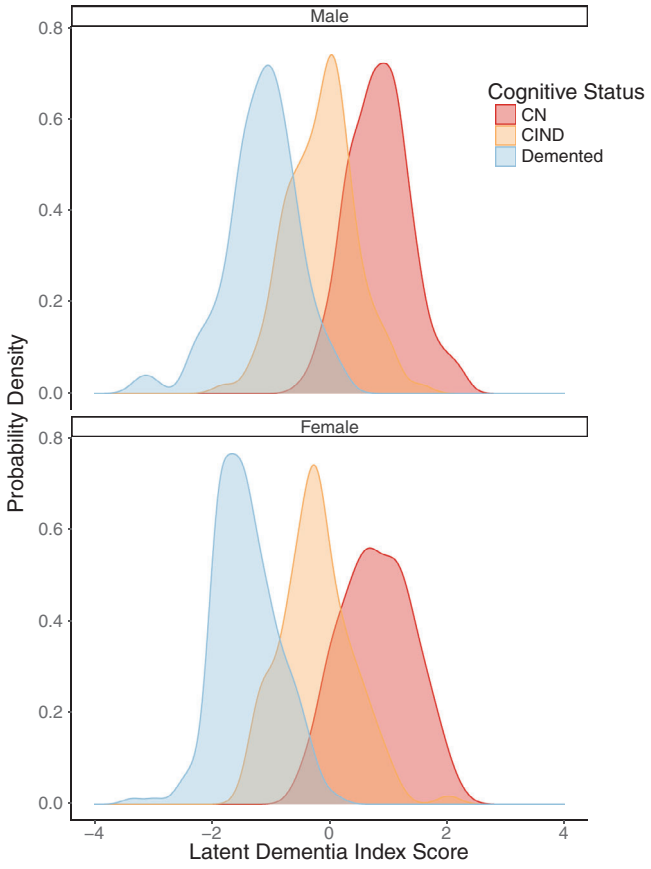

FIGURE 3 LDI distribution for dementia cases, CIND, and cognitive normal ADAMS participants by sex. CIND, cognitive impairment not dementia; CN, cognitively normal.

whether common cognitive protocols deployed across cultures – as in HRS, the ELSA (England), LASI (India), and the MHAS (Mexico) – enable meaningful cross-cultural comparison of dementia likelihood. Sex differences in dementia likelihood can be explored both within and between countries. Cross-nationally comparative work on sex differences in dementia allows evaluation of consistency of sex differences in dementia across countries with varying structures and norms surrounding gender, such as access to cognition-promoting resources like education.[35]

We now discuss how our approach fits in with other approaches to identifying dementia cases in the ADAMS and HRS data sets.[21] First, our ROC curve suggests strong sensitivity and specificity. Second, the LDI partitions true score variance from measurement variance so mean comparisons of the LDI across sex are not confounded by error variance. Third, given the richness of the neuropsychological and functional ability variables used to estimate dementia likelihood, the current results extend algorithmic approaches[21] using only cognitive and proxy measures. We use respondents' cognitive performance in addition to functional ability. For example, a comparison of algorithmic approaches in HRS show inconsistent invariance in memory-related measures.[36] The LDI approach, however, is a comprehensive test of invariance at the level of the association between dementia likelihood and all measures of cognitive and functional ability. Fourth, the LDI

provides a continuous measure of dementia likelihood, capturing the dementia process,[11] instead of relying on an ordered categorical variable decided by the placement of cut-points. Fifth, compared to prior algorithmic approaches, we explicitly tested whether the LDI scores are comparable across sex. Simple sum scores of manifest variables cannot determine whether the scores measure the same underlying construct across sex. Finally, the LDI approach accommodates missing data by using full information maximum likelihood estimation to reduce biases associated with missing data even when data are not missing at random.[29] This is often the case in population-based studies of cognition where respondents may not be able to complete cognitive assessments due to cognitive decline itself.

Substantively, there are notable sex differences in LDI scores worthy of mention. First, several of the LDI distributions in female subsamples were bimodal whereas the distributions were generally unimodal in male subsamples. For example, the LDI distributions of each education group in the female participants are either bimodal (< high school and > high school) or somewhat uniform (high school). These distributions could represent heterogeneous subgroups within each education group, that is, dementia likelihood is not equally distributed in more highly educated women whereas it is more equally distributed in less educated (≤ high school) women. Second, LDI distributions as a function of age differed between men and women. As with education, the bimodal LDI distribution in women aged 80–90 suggests high-risk and low-risk groups. As expected, women in the oldest age group had the highest overall expected dementia likelihood with scores piling at the low end of the distribution compared to men. Third, LDI distributions based on APOE allele counts followed expected sex differences.[3,37] Expected dementia likelihood increased with each additional allele in women whereas only men with two APOE $\varepsilon 4$ had higher risk compared to men with one or no alleles.

There are several limitations. First, ADAMS had lower participation rates than anticipated,[23] leading to potential selection biases. Second, the ADAMS sample only includes individuals aged 70+. Although most dementia cases occur in this group, results regarding patterns of (non)invariance across sex may differ in other age groups. Third, ADAMS data were collected 2001–2003. More recent older adult cohorts show higher education and more gender parity in education, which may affect cognitive function and dementia.[38,39] Additionally, the age of the ADAMS data means that standards and criteria for the diagnosis of dementia and cognitively impaired not demented have changed during this time. A strength of the LDI is that its estimation does not depend on the evolution of diagnostic criteria for major neurocognitive disorders or mild cognitive impairment. We nevertheless used the diagnostic criteria in the ADAMS sample for the reason the LDI should be able to order individuals in a population based on their likelihood of dementia. Last, although full scalar invariance of cognitive assessments across education level has been reported,[40] prior work has largely tested MI of measures of general cognitive ability or cognitive domains instead of latent dementia measures.

Future work should assess MI across sex in more recent cohorts of older adults with potentially different characteristic profiles, such as higher educational attainment. Additionally, future work should estimate comparable LDI scores across different subpopulations of the HRS and its cousin studies (e.g., the Mexican Health and Aging Study, the Longitudinal Study of Aging in India). Notably, one advantage of the LDI is that it can be estimated using any combination of cognitive ability (both memory and nonmemory) and functional ability items.[9] Thus, harmonization of LDI scores across populations should be feasible as long as studies have a subset of items in common.

ACKNOWLEDGMENTS

The Health and Retirement Study and the Aging, Demographics, and Memory Study are supported by the National Institute on Aging (U01 AG09740). Joseph Saenz acknowledges support from the National Institute on Aging (R00 AG058799). Christopher R. Beam acknowledges support from the Alzheimer's Association (AARF-17-505302). Alice J. Kim acknowledges support from the National Institute on Aging (T32 AG000037).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

Informed consent was obtained from all subjects involved in the study.

REFERENCES

1. Mazure CM, Swendsen J. Sex differences in Alzheimer's disease and other dementias. *Lancet Neurol*. 2016;15:451-452. doi: 10.1016/S1474-4422(16)00067-3
2. Wiley J. Alzheimer's disease facts and figures. 2021;17.
3. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and alzheimer disease: a meta-analysis. *JAMA*. 1997;278:1349-1356. doi: 10.1001/jama.1997.03550160069041
4. Luo J, Beam CR, Gatz M. Is stress an overlooked risk factor for dementia? A systematic review from a lifespan developmental perspective. *Prev Sci*. 2022;1-14. doi: 10.1007/s11121-022-01385-1
5. Beam CR, Kaneshiro C, Jang JY, Reynolds CA, Pedersen NL, Gatz M. Differences between women and men in incidence rates of dementia and Alzheimer's disease. *J Alzheimers Dis*. 2018;64:1077-1083. doi: 10.3233/JAD-180141
6. Sheehan CM, Tucker-Drob EM. Gendered expectations distort male – female differences in instrumental activities of daily living in later adulthood. *J Gerontol Ser B*. 2019;74:715-723. doi: 10.1093/geronb/gbw209
7. Sundermann EE, Maki PM, Rubin LH, et al. Female advantage in verbal memory: evidence of sex-specific cognitive reserve. *Neurology*. 2016;87:1916-1924. doi: 10.1212/WNL.0000000000003288
8. Sundermann EE, Biegon A, Rubin LH, et al. Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. *Neurology*. 2016;86:1368-1376. doi: 10.1212/WNL.0000000000002570
9. Beam CR, Luczak SE, Panizzon MS, et al. Estimating likelihood of dementia in the absence of diagnostic data: a latent dementia index in 10 genetically informed studies. *J Alzheimers Dis*. 2022;90:1187-1201. doi: 10.3233/JAD-220472
10. Gavett BE, Vudy V, Jeffrey M, John SE, Gurnani AS, Adams JW. The δ latent dementia phenotype in the uniform data set: cross-validation and extension. *Neuropsychology*. 2015;29:344-352. doi: 10.1037/neu0000128

11. Royall DR, Palmer RF, O'Bryant SE. Consortium for the TAR and C. Validation of a latent variable representing the dementing process. *J Alzheimers Dis*. 2012;30:639-649. doi: 10.3233/JAD-2012-120055
12. Blankson AN, McArdle JJ. Measurement invariance of cognitive abilities across ethnicity, gender, and time among older Americans. *J Gerontol Ser B*. 2015;70:386-397. doi: 10.1093/geronb/gbt106
13. Avila JF, Rentería MA, Witkiewitz K, Verney SP, Vonk JMJ, Manly JJ. Measurement invariance of neuropsychological measures of cognitive aging across race/ethnicity by sex/gender groups. *Neuropsychology*. 2020;34:3-14. doi: 10.1037/neu0000584
14. Gurnani AS, Lin SS-H, Gavett BE. The Colorado Cognitive Assessment (CoCA): development of an advanced neuropsychological screening tool. *Arch Clin Neuropsychol*. 2020;35:176-187. doi: 10.1093/arclin/acz066
15. Plassman BL, Langa KM, Fisher GG, et al. Prevalence of dementia in the United States: the Aging, Demographics, and Memory Study. *Neuroepidemiology*. 2007;29:125-132. doi: 10.1159/000109998
16. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43:576-585. doi: 10.1093/ije/dyu067
17. Fisher GG, Ryan LH. Overview of the Health and Retirement Study and introduction to the special issue. *Work Aging Retire*. 2018;4:1-9. doi: 10.1093/workar/wax032
18. Ofstedal MB, Fisher GG, Herzog AR. Documentation of cognitive functioning measures in the Health and Retirement Study. *Ann Arbor MI Univ Mich*. 2005;10. Retrieved from https://hrs.isr.umich.edu/publications/biblio/5620
19. Langa KM, Kabeto M, Weir D. Report on race and cognitive impairment using HRS, *2010 Alzheimer's disease facts and figures*. 2009. Retrieved from http://www.alz.org/documents_custom/report_alzfactsfigures2010.pdf
20. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and memory Study. *J Gerontol Ser B*. 2011:i162-71. doi: 10.1093/geronb/gbr048
21. Gianattasio KZ, Wu Q, Glymour MM, Power MC. Comparison of methods for algorithmic classification of dementia status in the Health and Retirement Study. *Epidemiol Camb Mass*. 2019;30:291-302. doi: 10.1097/EDE.0000000000000945
22. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263-269. doi: 10.1016/j.jalz.2011.03.005
23. Langa KM, Plassman BL, Wallace RB, et al. The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*. 2005;25:181-191. doi: 10.1159/000087448
24. HRS. Health and Retirement Study: Aging, Demographics, and Memory Study (ADAMS) Supplement. Data Description and Usage 2013.
25. American Psychological Association. Diagnostic and Statistical Manual of Mental Disorders (Rev. 3rd ed.). 3rd ed. Washington DC: American Psychological Association; 1987.
26. American Psychological Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed.). 4th ed. Washington DC: American Psychological Association; 1994.
27. Muthén L, Muthén B. *Mplus User's Guide*. 8th ed. Muthén & Muthén; 1998.
28. Millsap RE. *Statistical Approaches to Measurement Invariance*. Routledge; 2011. doi: 10.4324/9780203821961
29. Enders CK. *Applied Missing Data Analysis*. Guilford Press; 2010.
30. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res*. 2004;33:261-304. doi: 10.1177/0049124104268644
31. Meng X, D'Arcy C. Education and dementia in the context of the cognitive reserve hypothesis: a systematic review with meta-analyses and qualitative analyses. *PLOS ONE*. 2012;7:e38268. doi: 10.1371/journal.pone.0038268
32. Verghese PB, Castellano JM, Holtzman DM. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *Lancet Neurol*. 2011;10:241-252. doi: 10.1016/S1474-4422(10)70325-2
33. Ziegler-Graham K, Brookmeyer R, Johnson E, Arrighi HM. Worldwide variation in the doubling time of Alzheimer's disease incidence rates. *Alzheimers Dement J Alzheimers Assoc*. 2008;4:316-323. doi: 10.1016/j.jalz.2008.05.2479
34. Langa KM, Ryan LH, McCammon RJ, et al. The Health and Retirement Study harmonized cognitive assessment protocol project: study design and methods. *Neuroepidemiology*. 2020;54:64-74. doi: 10.1159/000503004
35. Seedat S, Scott KM, Angermeyer MC, et al. Cross-National Associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Arch Gen Psychiatry*. 2009;66:785-795. doi: 10.1001/archgenpsychiatry.2009.36
36. Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Maria Glymour M. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Dis Assoc Disord*. 2013;27:207-212. doi: 10.1097/WAD.0b013e31826cfe90
37. Beam CR, Kaneshiro C, Jang JY, Reynolds CA, Pedersen NL, Gatz M. A twin study of sex differences in genetic risk for all dementia, Alzheimer's disease (AD), and non-AD dementia. *J Alzheimers Dis*. 2020;76:539-551. doi: 10.3233/JAD-191192
38. Angrisani M, Lee J, Meijer E. The gender gap in education and late-life cognition: evidence from multiple countries and birth cohorts. *J Econ Ageing*. 2020;16:100232. doi: 10.1016/j.jeoa.2019.100232
39. Langa KM, Larson EB, Crimmins EM, et al. A comparison of the prevalence of dementia in the United States in 2000 and 2012. *JAMA Intern Med*. 2017;177:51-58. doi: 10.1001/jamainternmed.2016.6807
40. Ma Y, Carlsson CM, Wahoske ML, et al>. Latent factor structure and measurement invariance of the NIH toolbox cognition battery in an Alzheimer's disease research sample. *J Int Neuropsychol Soc*. 2021;27:412-425.

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: Saenz J, Beam CR, Kim AJ. Development of a latent dementia index in the aging, demographics, and memory study: Validation and measurement invariance by sex. *Alzheimer's Dement*. 2023;15:e12433. https://doi.org/10.1002/dad2.12433

APPENDIX: RESCALING OF VARIABLES FOR CONFIRMATORY FACTOR ANALYSES

Several variables were rescaled to facilitate model estimation. Mini-Mental State Examination (MMSE), animal fluency, Boston Naming, digit span, Immediate Word Recall, WMS Logical Memory 1 & 2, and Fuld Object Memory were divided by two. Shipley, Digit Symbol Modalities, Word Association, and Wide Range Achievement Test were divided by ten. Trails making tests were divided by one hundred and multiplied by negative one so higher values represent faster completion time (that is, higher cognitive ability).