Title

Predictors of Retest Effects in a Longitudinal Study of Cognitive Aging in a Diverse Community-Based Sample

Permalink

https://escholarship.org/uc/item/2716t264

Journal

Journal of the International Neuropsychological Society, 21(7)

ISSN

1355-6177

Authors

Gross, Alden L
Benitez, Andreana
Shih, Regina
et al.

Publication Date

2015-08-01

DOI

10.1017/s1355617715000508

Peer reviewed

Predictors of retest effects in a longitudinal study of cognitive aging in a diverse community-based sample

Alden L. Gross[1], **Andreana Benitez**[2], **Regina Shih**[3], **Katherine J. Bangen**[4], **M Maria M Glymour**[5], **Bonnie Sachs**[6], **Shannon Sisco**[7], **Jeannine Skinner**[8], **Brooke C. Schneider**[9], and **Jennifer J. Manly**[10]

[1]Departments of Epidemiology and Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; and Center on Aging and Health, Johns Hopkins University, Baltimore, Maryland [2]Department of Radiology and Radiological Sciences, Center for Biomedical Imaging, Medical University of South Carolina, Charleston, South Carolina [3]RAND Corporation, Arlington, Virginia [4]Department of Psychiatry, University of California, San Diego, La Jolla, California [5]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California [6]Department of Neurology, Wake Forest Baptist Medical Center, Winston Salem, North Carolina [7]North Florida/South Georgia Veterans Health System, Department of Veterans Affairs, Gainesville, Florida [8]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington [9]Department of Psychiatry and Psychotherapie, University Hospital Hamburg-Eppendorf, Hamburg, Germany [10]Taub Institute for Research on Alzheimer's disease and the Aging Brain, Columbia University, New York, New York; Gertrude H. Sergievsky Center, Columbia University, New York, New York; and Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, New York

Abstract

OBJECTIVE—Better performance due to repeated testing can bias long-term trajectories of cognitive aging and correlates of change. We examined whether retest effects differ as a function of individual differences pertinent to cognitive aging: race/ethnicity, age, sex, language, years of education, and dementia risk factors including APOE ε4 status, baseline cognitive performance, and cardiovascular risk.

METHOD—We used data from the Washington Heights-Inwood Columbia Aging Project, a community-based cohort of older adults (n=4,073). We modeled cognitive change and retest effects in summary factors for general cognitive performance, memory, executive functioning, and language using multilevel models. Retest effects were parameterized in two ways, as improvement between the first and subsequent testings, and as the square root of the number of prior testings. We evaluated whether the retest effect differed by individual characteristics.

RESULTS—The mean retest effect for general cognitive performance was 0.60 standard deviations (95%CI: 0.46, 0.74), and was similar for memory, executive functioning, and language. Retest effects were greater for participants in the lowest quartile of cognitive performance, consistent with regression to the mean. Retest did not differ by other characteristics.

CONCLUSIONS—Retest effects are large in this community-based sample, but do not vary by demographic or dementia-related characteristics. Differential retest effects may not limit the generalizability of inferences across different groups in longitudinal research.

Keywords

practice effect; retest effect; neuropsychological testing; older adults

Introduction

Estimation of the pace of cognitive decline throughout the lifecourse is central to research on cognitive aging and dementia (Salthouse, 2010a). Cognitive decline is a more compelling marker of Alzheimer's disease (AD) dementia than impairment at one testing session because it is less affected by historical factors such as years of education that precede the onset of AD (Glymour et al., 2005). However, design and analysis of longitudinal studies, wherein cognitive testing is repeatedly conducted on the same person over time, can be complicated because, in addition to normal aging or maturation, factors such as selective attrition, period and cohort effects, statistical artifacts (e.g., regression to the mean), and retest or practice effects contribute to changes in cognitive test performance (Dodge et al., 2011; Salthouse, 2010a; 2010b).

Retest or practice effects refer to the extent to which repeated cognitive testing results in improved performance due to familiarity with the testing materials and setting (Horton, 1992; Zehnder, Blasi, Berres, Spiegel, & Monsch, 2007). These effects are well-documented in longitudinal studies of cognitive aging (Abner et al., 2012; Basso et al., 1999; Calamia et al., 2012; Collie et al., 2003; Cooper, Lacritz, Weiner, Rosenberg, & Cullum, 2004; Duff et al., 2011; Ferrer et al., 2004, 2005; Frank et al., 1996; Horton et al., 1992; Howieson et al., 2008; Ivnik et al., 1999; Jacqmin-Gadda et al., 1997; Machulda et al., 2013; Mitrushina et al., 1991; Rabbitt et al., 2001, 2004; Salthouse, 2009; Wilson, Leurgans, Boyle, & Bennett, 2011; Wilson et al., 2006; Zehnder et al., 2007). A consensus conference for clinical neuropsychology has called for research on ramifications of repeated cognitive testing (Heilbronner et al., 2010). Van der Elst and colleagues (2008) found a robust increase of between 0.2 and 0.6 standard deviations (SD) in verbal list-learning performance three years after the first testing occasion in a large sample of cognitively normal older adults, while Bartels and colleagues (2010) found medium to large retest effects between 0.36 and 1.19 SD after approximately 3 months. Although both of these studies conceptualize retest effects as a one-time boost between the first and subsequent occasions, retest effects may also exist at each visit with diminishing returns (Collie et al., 2003; Sliwinski et al., 2011).

Failure to account for retest effects obscures the estimated rate of cognitive decline. If retest effects are correlated with risk factors of interest, ignoring them may lead to biased estimates of their effects on the rate of cognitive change. Retest effects may differ by the type of cognitive task. Tests that measure different cognitive abilities (e.g., memory, language) (Cooper et al., 2004) or that use different administration or response modalities (e.g., oral vs written) might show different patterns of retest effects. In this study, we

examined retest effects at the level of constructs rather than individual cognitive tests to avoid detecting differences in modality.

In addition to the type of test, retest effects may be attributable to participant characteristics related to proficiency in test-taking via test-taking strategies and less test anxiety, in which case persons with less testing experience might show larger retest effects (Thorndike, 1922). Retest effects may also be attributed to episodic memory, or the successful learning and retention of test content such that subsequent improved performance is facilitated by recollection of the content. This is a motivation behind the use of alternate forms for tests of episodic memory (e.g., Benedict et al., 1998; Delis et al., 2000). Thus, testing for differential retest effects by factors related to test experience and episodic memory provide a way to better understand retest effects.

Socio-demographic factors related to test experience

Because educational attainment is a strong predictor of cognitive performance in later life, retest effects may differ by number of years of education (Cagney & Lauderdale, 2002; Stern et al., 1994). Individuals with less education or lower quality education have less prior experience with test-taking and strategies for maximizing test performance. Such individuals have the most to gain from practice with the test. Similarly, given differences in early educational experiences for older adults by race and ethnicity due to persistent educational inequalities (Glymour & Manly, 2008), we hypothesized that Hispanic older adults, most of whom in the present sample are immigrants to the US, may be less familiar on average with testing and therefore experience greater retest effects (Gould, 1996).

Age, sex, and language spoken at home may also moderate retest effects. Previous research suggests that, with the exception of measures of word list recall, retest effects are inversely related to age (Mitrushina & Satz, 1991; Rabbitt et al., 2008). Sex differences in cognitive performance have been documented for a range of cognitive abilities, suggesting differential retest effects may also occur. Women tend to do better on memory tests (leaving men with more room to improve upon retest), while men tend to do better on visuospatial tasks (Mann et al., 1990; Salthouse, 2010a; Voyer et al., 1995). Primary language may also be important to retest; one study found Spanish-speakers demonstrated (greater?) retest effects than English-speakers (Mungas et al., 2000).

Dementia risk factors

The ability to learn and retain new information may also facilitate retest effects. Previous studies suggest that the absence of retest effects may reflect amnestic Mild Cognitive Impairment (MCI) or AD (Darby et al., 2002; Duff et al., 2011; Frank et al., 1996; Schrijnemaekers et al., 2006). However, at least one recent study reported retest effects for memory in participants with MCI and dementia (Machulda et al., 2013). The apolipoprotein E (APOE) ε4 allele predicts earlier onset of Alzheimer's disease among older Whites (Baxter et al., 2003; Blair et al., 2005; Haan et al., 1999), but the association seems to be attenuated among Blacks (Borenstein et al., 2006; Tang et al., 1998). A previous study found APOE e4 carriers did not exhibit a retest effect (Machulda et al., 2013). Further, cardiovascular burden is an established risk factor for poorer cognition and

neurodegenerative disease, especially among minority older adults (Flicker, 2010; Luchsinger et al., 2005). Thus, it is possible that greater cardiovascular risk burden may affect the magnitude of retest effects.

The present study

We examined whether retest effects vary by demographic factors such as race/ethnicity, age, language spoken at home, sex, years of education, and dementia risk factors including APOE ε4 status, baseline cognitive status, and cardiovascular burden. We estimated multilevel random effects models of change in general cognitive performance, memory, executive function, and language. The mean retest effect was allowed to differ by the characteristic of interest. We hypothesized that Hispanic racial/ethnic group membership and fewer years of education predict larger retest effects, while dementia risk factors such as possession of the APOE e4 allele, lower cognitive performance at baseline, and greater cerebrovascular risk burden predict smaller retest effects.

Methods

Participants and Procedures

We used data on N=4,073 participants from the Washington Heights-Inwood Columbia Aging Project (WHICAP), an ongoing epidemiologic cohort of community-living Medicare-eligible older adults recruited from northern Manhattan (Tang et al., 2001). Participants were residents of three contiguous US census tracts in Northern Manhattan, New York. Individuals were invited to participate in an in-person survey in 1992, with follow-up visits every two to three years. Recruitment re-opened in 1999 to replenish the cohort. At each interview, participants answered extensive questionnaires about their early life education, health, and cognitive performance. The present study used data from 4,073 participants who participated in neuropsychological assessments. Details of the sampling strategies, recruitment outcomes (Luchsinger, Tang, Shea & Mayeux, 2001; Manly et al., 2005), and study design and procedures have been published previously (Tang et al., 2001). The study was approved by Institutional Review Boards at Columbia Presbyterian Medical Center, Columbia University Health Sciences, and the New York State Psychiatric Institute.

Measures

Racial/ethnic group—Participants self-reported their race by selecting membership from categories of American Indian/Alaska Native, Asian, Native Hawaiian or other Pacific Islander, Black or African American, or White. Participants were then asked whether they were Hispanic. We grouped participants into categories of non-Hispanic White, non-Hispanic Black, and Hispanic.

Cardiovascular burden—We used a summary of cardiovascular burden based on presence of diabetes, hypertension, heart disease, stroke, central obesity, and current smoking status (Schneider et al., 2014).

Cognitive Performance—WHICAP administered a neuropsychological test battery at each study visit (Tang et al., 2001). Tests were designed for administration in Spanish or

English (Dugbartey et al., 2000; Jacobs et al., 1995). The tests are described in the Appendix. We constructed factor scores for general cognitive performance, memory, executive function, and language using confirmatory factor analysis models for each domain. We used immediate recall, delayed recall, and delayed recognition from the Buschke Selective Reminding Test to construct the memory factor. The executive functioning factor was derived using the Color Trail-Making Test (A and B), WAIS Similarities, Identities/Oddities, shape time, time to detect a consonant trigram, and semantic fluency for animals. Language was derived using phonemic fluency, 15-item Boston Naming, repetition, and comprehension. All of the above variables contributed to the general cognitive factor. The assignment of tests to factors is largely consistent with a previously published factor analysis of the neuropsychological test battery in the WHICAP cohort (Siedlecki et al., 2010), except that we dropped the speed factor derived by Siedlecki et al. and added an executive functioning factor. The executive functioning factor has more indicators that represent broader fluid ability, and is more reliable than two separate factors for speed and reasoning that are each based on fewer measures.

Each factor was scaled to have a mean of 50 and standard deviation (SD) of 10 in the US population of adults aged 70 years and older to facilitate comparison of magnitudes of effects across domains and with future studies. Details are provided elsewhere (Gross et al., 2014a, 2014b). Briefly, we calibrated the factors using a nationally representative sample of adults aged 70 and older from the Aging, Demographics, and Memory Study (ADAMS), a sub-study of the Health and Retirement Study (Juster & Suzman, 1995; Langa et al., 2005). The ADAMS battery included Trails A and B, Digits Forward and Backward, semantic and phonemic fluency, Boston Naming Test, Symbol Digit Modalities, and a 10-noun word recall task. Items common to ADAMS and WHICAP served as links to calibrate cognitive factors. The factor analysis was performed in a longitudinal dataset with multiple records per participant. We fixed item discrimination and difficulty parameters for common items in the factor analysis including both WHICAP and ADAMS to the values estimated in an ADAMS-only factor analysis. This scaling approach does not make the WHICAP sample nationally representative, but it allows future analysts, using other datasets with items overlapping with ADAMS, to derive directly comparable scores. This approach assumes measurement invariance of factors with respect to time: an assumption previously verified in other samples of older adults and which we evaluated in WHICAP through formal tests described earlier (Hayden et al., 2011; Johnson et al., 2012). We additionally tested longitudinal measurement invariance of the factors among participants assessed at baseline and whose second study visit was between 1.5 and 2.5 years later (median: 2.1 years) using multiple group confirmatory factor analysis models. Details are provided in the Appendix.

Analyses

To test hypotheses, we used multilevel models with random effects for people and time alongside fixed effects for retest in general cognition, memory, executive functioning, and language (Johnson et al., 2012; Laird & Ware, 1982; Muthen et al., 1997; Raudenbush & Bryk, 2002). Time since enrollment into the study was the timescale of interest. The system of equations below describes the basic model:

$$\text{Level 1:} \quad Y_{ij} = \beta_{0i} + \beta_{1i} * \text{time}_{ij} + \beta_2 * \text{retest}_{ij} + \sum_{p=3}^{5} \beta_p * X_i + \varepsilon_{ij} \quad (1)$$

$$\text{Level 2:} \quad \beta_{0i} = \gamma_{00} + U_{0i} \quad (2)$$

$$\beta_{1i} = \gamma_{10} + U_{1i} \quad (3)$$

Y_{ij} is a cognitive outcome (general cognitive performance, memory, executive functioning, or language) for participant i at time j. The level 1 model describes within-person change over time based on random (U_{0i}) and fixed (γ_{00}) effects for participants, random (U_{1i}) and fixed (γ_{10}) effects for time, a fixed effect (β_2) for the retest effect, adjustment variables β_p, and residual error ε_{ij} for each participant at each time. Level 2 equations describe the random and fixed effects for participants and time. Distributions of ε_{ij}, U_{0i}, and U_{1i} are assumed to be normal with mean 0 and variance 1.

We coded retest in two ways to acknowledge different conceptualizations of how they come about. First, as our primary analysis, the retest variable was coded 0 at each participant's first study visit in which they were administered a neuropsychological battery, and 1 otherwise. Retest effects here are interpretable as the difference or jump in performance from the first assessment to the predicted performance based on the level and slope of change at the second and later assessments. This characterization is consistent with previous studies examining retest effects (Ivnik et al., 1999; Rabbitt, Diggle, Holland, & McInnes, 2004; Salthouse, Schroeder, & Ferrer, 2004; Salthouse & Tucker-Drob, 2008). Previous studies have suggested that subsequent gains after the second testing occasion are negligible (but see the Discussion) (Kausler et al., 1994; Rabbitt et al., 1993, 2004). Second, to allow for the possibility that participants learn more at each test occasion with diminishing returns over time (Abner et al., 2012; Collie et al., 2003; Sliwinski et al., 2011), we also calculated retest as the square root of the number of prior test occasions. We adjusted all models for sex, baseline age, and recruitment cohort (1992 or 1999).

To determine whether retest effects vary by individual characteristics, or effect modification, we extended the model described above to a series of multiple group models in a structural equation modeling framework, in which groups were defined based on the characteristic of interest. Groups were defined by race/ethnicity (non-Hispanic Black, non-Hispanic White, Hispanic), age (<75 years, 75–80 years, 80+ years), sex, years of education (less than 8 years, 8 or more years), APOE ε4 status (carrier, noncarrier), quartile of baseline general cognitive performance, and number of cardiovascular risk factors (0, 1, 2, or 3). We conducted analyses by baseline quartiles of cognitive performance instead of adjudicated dementia diagnosis because, in WHICAP, neuropsychological test performance was considered during the adjudication procedure. Differences in mean retest effects between these groupings are estimated in a manner analogous to using an interaction between the characteristic and retest indicator, as follows:

$$\text{Level 1: } Y_{ij} = \beta_{0i} + \beta_{1i} * time_{ij} + \beta_2 * retest_i + \beta_3 * moderator_i + \beta_4 * (moderator_i * retest_{ij}) + \varepsilon_{ij} \quad (4)$$

The interaction of the moderator and the retest effect, β_4, is the parameter of interest.

In planned sensitivity analyses, we examined retest effects for all component tests in the WHICAP neuropsychological battery. Analyses were conducted with Mplus statistical software (version 7.11, Muthen & Muthen, 1998–2012) using robust maximum likelihood estimation that assumed outcome observations were missing at random, conditional on covariates (Little & Rubin, 1987). Fit of modeled trajectories to data was assessed with a pseudo-R^2 statistic. The pseudo-R^2 represents the proportion of variability in observed data explained by the model (Singer & Willet, 2003). It is calculated by squaring the correlation between observed and model-estimated (including random effects terms) outcome scores. We adjusted models for potential selective survival using inverse probability weights (Hernan & Robins, 2006) calculated from a logistic regression of death on age, sex, baseline general cognitive performance, APOE e4 status, education in years, recruitment cohort, and cardiovascular risk measured at baseline.

Results

The study sample was mostly female (68.5%), had eight or more years of education (53.6%), and the average age at the first visit was 77 (range 63, 103)(Table 1). The sample was ethnically diverse, with 33.7% non-Hispanic Black, 24.9% non-Hispanic White, and 41.4% Hispanic. The percentage of participants with at least one APOE ε4 allele was 22.6%.

Overall retest effect

The median number of study visits was three (interquartile interval: 2, 4) and median follow-up time was 3.9 years (interquartile interval: 1, 7.8). The second study visit took place on average 1.9 years (interquartile interval: 1.3, 2.2 years) after the first study visit. For each cognitive outcome, a 1-point difference is analogous to a 0.1 standard deviation difference. As expected, the overall retest effect was considerable for all domains. For general cognitive performance, the retest effect was 0.60 SD, while the annual rate of general cognitive decline was only –0.047 SD (Table 2). Thus, the absolute value of the retest effect is the same magnitude as 12.8 years of cognitive decline (Table 2). Retest effects were also large for memory (retest=0.57 SD, 95% CI: 0.42, 0.72 SD), executive functioning (retest=0.45 SD, 95% CI: 0.32, 0.58 SD), and language (retest=0.64 SD, 95% CI: 0.47, 0.81 SD)(Table 2).

Effect modification of retest effects by participant characteristics

Models fit well to the data, with pseudo-R^2 values above 0.79 for each cognitive outcome (Table 3). Visual inspection of model residuals confirmed adequate fit to the data. The magnitude of the retest effect, parameterized as the jump from the first to subsequent test occasions, was statistically significant and positive for general cognitive performance in nearly every subgroup (Table 3). Inferences were similar for memory and for language. For executive functioning, average retest effects tended to be smaller but were mostly

statistically significant (Table 3). This pattern of results was identical when we parameterized retest effects as the square root of the number of prior test occasions (Appendix Table 2).

The magnitude of retest effects did not differ significantly by race/ethnicity, age, language, sex, education, APOE status, or cardiovascular burden (Table 3). Participants in the lowest quartile of baseline general cognitive performance demonstrated greater retest effects compared to participants in the middle two quartiles of general cognitive performance, for whom retest effects were not significant (Table 3). Figure 1 shows the model-estimated cognitive trajectory for participants at these quartiles of cognitive function. Although we did not exclude participants who had an adjudicated diagnosis of dementia in WHICAP, we observed that 645 of 679 (94.9%) of participants with dementia were in the lowest quartile of baseline cognitive performance (sensitivity), and 3,006 of 3,369 (89.2%) of non-demented participants had a score above the lowest quartile (specificity).

We examined the magnitude of retest effects for each component test in the WHICAP battery. Results of this sensitivity analysis were consistent with findings using the factor scores. Retest effects were generally greater in magnitude for memory tests than for executive functioning tests.

Discussion

In this large, diverse community-based sample of older adults, we examined differences in retest effects by racial/ethnic group, age, language spoken at home, sex, years of education, APOE ε4 status, baseline cognitive function, and cardiovascular burden. Despite the relatively long two year interval between testing occasions, the overall magnitude of retest was on average more than ten times the annual rate of subsequent cognitive decline, and greatest for language. The magnitude of retest effects did not differ by any characteristic examined other than baseline cognitive status: on average, participants performing in the lowest quartile at baseline experienced the greatest boost from repeated testing. This finding is probably attributable to regression to the mean. Overall, the results suggest retest effects do not differ greatly across observable demographic and dementia-related factors.

Previous research indicates that the magnitude of retest effects varies widely across different tests (Calamia et al., 2012; Frank et al., 1996), with effects typically but not always largest for visual memory and smallest for visuospatial ability (Calamia et al., 2012, but see also Dodrill et al., 1975; Ferrer et al., 2004; Frank et al., 1996; McCaffrey et al, 1992). In our study, we built on prior research by considering cognitive domains instead of individual tests in an attempt to draw conclusions at the level of constructs, and mitigate the potential for spurious findings from multiple tests. A further implicit advantage of our study was the choice of scaling to an external standard, the ADAMS HRS. This scaling made no difference in the results compared to factors scores that were scaled internally. Scale choice is in many cases arbitrary. However, we believe that future scientific progress in the area of cognitive aging will be accelerated if findings are presented on a common scale across studies. Resources are available that describe how other studies can be linked to an external

metric such as the nationally representative sample used here (e.g., Gross et al., 2014a, 2014b; Jones et al., 2010).

Some previous studies have not found a retest effect in cognitively impaired people (Cooper et al., 2004; Howieson et al., 2008; Wilson et al., 2011). A recent meta-analysis suggested that older age and absence of dementia predict smaller retest effects (Calamia et al., 2012). Our study did not replicate these differences. At least two studies, which like ours featured test-retest intervals between one and two years, found no difference in the magnitude of retest effects between cognitively normal and impaired older adults (Frank et al., 1996; Machulda et al., 2013).

Findings from this study indicated that retest effects do not differ by race/ethnicity or years of education, which were intended to be proxies for testing experience. However, years of education only captures testing experience from early life, and does not reflect experiences accumulated throughout life. Admittedly, race and ethnicity are imperfect markers of test experience, and thus our results cannot conclusively disprove the hypothesis that test experience plays a role in retest effects. Further, most Hispanic participants in WHICAP were immigrants, whose years of education are systematically lower and do not easily translate to years of education in the US (Hoffmeyer-Zlotnik et al., 2005).

Our finding of differential retest effects by baseline cognitive status is likely attributable to regression to the mean (Barnett et al., 2005). This presents unusual implications for the study of cognitive decline in dementia, because most participants performing in the lowest quartile of cognitive performance had a study diagnosis of dementia. Persons with dementia have impaired learning and memory, and thus one might expect they should exhibit smaller retest effects assuming that retest is attributable largely to episodic memory. Previous studies have reported no retest effects in persons with MCI and dementia (Darby et al., 2002; Schrijnemaekers et al., 2006) or minimal (Duff et al., 2011). Incipient dementia may not attenuate retest effects if procedural memory accounts for improvement on repeated test administration (Mitrushina et al., 2005). Procedural memory, or long-term, unconscious recollection of previous experiences important for retaining skills (e.g., typing on a keyboard or riding a bicycle), is relatively well-preserved in people with dementia (Meyer & Schvaneveldt, 1971; Perani et al., 1993; Sabe et al., 1995; Schaie, 2005; Tulving & Markowitsch, 1998).

The limitations of our study must be noted. First, we defined retest effects in two ways based on the discontinuity between first and second assessments and on the square root of the number of prior test occasions. The former approach imposes the assumption that the retest benefit is constant across the second and subsequent assessments. The latter approach assumes accumulating retest effects at each successive assessment, with diminishing additional benefit at each successive assessment. Although modest violations of these assumptions are plausible, such violations are unlikely to substantively alter our findings. There are other plausible specifications of retest effects. For example, if each successive test occasion were to hypothetically confer a slightly larger retest benefit, our effect estimates would be a weighted average of these effects. This phenomenon would obscure subgroup differences in the magnitude of retest effects only if such differences occurred for some, but

not all, waves of assessment. We think such a complex pattern of retest effects is unlikely. Second, regardless of how we parameterize them, retest effects are difficult to disentangle from aging in studies that have roughly equally spaced assessment intervals because the number of prior assessments is nearly collinear with time since baseline (Hoffman et al., 2011). In the absence of random assignment of timing of the first assessment, simplifying assumptions are necessary to identify retest effects (Hoffman et al., 2011; Sliwinski et al., 2010). We did not attempt to estimate retest differences as a function of the amount of time elapsed between successive tests because such variability is relatively small in WHICAP, by design, and any variance that is observed may be due to other variables such as respondents' enthusiasm for participating in cognitive assessments.

A third limitation is that in our study, we cannot know for certain whether we are capturing retest effects between the first and second visits, or change in cognitive performance. Improvement in cognitive performance is unlikely given that many who showed larger retest effects had dementia, and cognition is not expected to improve over time in people with dementia. The retest effects in our regressions are based either on a contrast between cognitive performance at the first assessment and cognitive performance at subsequent assessments, or an accumulating benefit with diminishing returns. Thus, a further limitation of our approach is that, to the extent age-related change is incorrectly estimated, the estimated retest effect will also be incorrect (Hoffman et al., 2011). However, in a typical cohort study design, we believe this approach is the best available strategy to estimate retest effects. A final limitation is that the present analysis was restricted to cognitive domains tested in WHICAP. Measures of spatial ability, processing speed, and higher-level task-switching, for example, were not available. The mechanisms by which retest effects operate, and thus predictors of differential retest effects, may differ for different domains.

Retest effects cannot be ignored in longitudinal research in cognitive aging because they may mask age-related cognitive decline (Ronnlund & Nilsson, 2006; Ronnlund, Nyberg, Backman, & Nilsson, 2005). The present study addresses the need for research on effects of repeated neuropsychological testing by evaluating differential patterns of retest effects for several cognitive domains in a diverse sample of community-living older adults (Heilbronner et al., 2010). Overall, our findings with respect to differential retest effects may alleviate concerns from clinicians and researchers about differential retest effects among observable groups in cognitive aging research. Although the findings suggest differential retest effects may not limit the generalizability of inferences across groups in longitudinal research, replication in other cohorts with different participant characteristics and retest intervals is warranted.

References

Abner EL, Dennis BC, Mathews MJ, Mendiondo MS, Caban-Holt A, Kryscio RJ, Schmitt FA, Crowley JJ. PREADViSE Investigators, SELECT Investigators. Practice effects in a longitudinal, multi-center Alzheimer's disease prevention clinical trial. Trials. 2012; 13:217.10.1186/1745-6215-13-217 [PubMed: 23171483]

Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol. 2005; 34(1):215–20. [PubMed: 15333621]

Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. BMC Neurosci. 2010; 11:118.10.1186/1471-2202-11-118 [PubMed: 20846444]

Basso MR, Bornstein RA, Lang JM. Practice effects on commonly used measures of executive function across twelve months. Clin Neuropsychol. 1999; 13(3):283–292.10.1076/clin. 13.3.283.1743 [PubMed: 10726600]

Baxter LC, Caselli RJ, Johnson SC, Reiman E, Osborne D. Apolipoprotein E epsilon 4 affects new learning in cognitively normal individuals at risk for Alzheimer's disease. Neurobiol Aging. 2003; 24(7):947–952.10.1016/S0197-4580(03)00006-X [PubMed: 12928055]

Ben-Yishay Y, Diller L, Mandleberg I, Gordon W, Gerstman LJ. Differences in matching persistence behavior during block design performance between older normal and brain-damaged persons: a process analysis. Cortex. 1974; 10(2):121–32.10.1016/S0010-9452(74)80003-1 [PubMed: 4844465]

Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. J Clin Exp Neuropsychol. 1998; 20(3):339–52.10.1076/jcen.20.3.339.822 [PubMed: 9845161]

Bentler PM. Comparative Fit Indexes in Structural Models. Psychological Bulletin. 1990; 107:238–246.10.1037/0033-2909.107.2.238 [PubMed: 2320703]

Blair CK, Folsom AR, Knopman DS, Bray MS, Mosley TH, Boerwinkle E. Atherosclerosis Risk in Communities (ARIC) Study Investigators. APOE genotype and cognitive decline in a middle-aged cohort. Neurology. 2005; 64(2):268–276.10.1212/01.WNL.0000149643.91367.8A [PubMed: 15668424]

Bontempo, DE.; Hofer, SM. Assessing factorial invariance in cross-sectional and longitudinal studies. In: Ong, AD.; van Dulmen, M., editors. Handbook of methods in positive psychology. New York: Oxford University Press; 2007. p. 153-175.

Borenstein AR, Copenhaver CI, Mortimer JA. Early-life risk factors for Alzheimer disease. Alzheimer Dis Assoc Disord. 2006; 20(1):63–72.10.1097/01.wad.0000201854.62116.d7 [PubMed: 16493239]

Borsboom D, Romeijn JW, Wicherts JM. Measurement invariance versus selection invariance: Is fair selection possible? Psychological Methods. 2008; 13:75–98.10.1037/1082-989X.13.2.75 [PubMed: 18557679]

Buschke H. Selective reminding for analysis of memory and learning. Journal of Verbal Learning and Verbal Behavior. 1973; 12:543–550.10.1016/S0022-5371(73)80034-9

Cagney KA, Lauderdale DS. Education, wealth, and cognitive functioning in later life. Journal of Gerontology: Psychological Sciences. 2002; 2:163–172.10.1093/geronb/57.2.P163

Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. Clin Neuropsychol. 2012; 26(4):543–570.10.1080/13854046.2012.680913 [PubMed: 22540222]

Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. J Int Neuropsychol Soc. 2003; 9(3):419–28. [PubMed: 12666766]

Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment: Reduced effect of practice in test–retest conditions. Alzheimer Disease and Associated Disorders. 2004; 18:120–122. [PubMed: 15494616]

Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. Neurology. 2002; 59:1042–1046.10.1212/WNL.59.7.1042 [PubMed: 12370459]

Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. California Verbal Learning Test: Second Edition. San Antonio, TX: Psychological Corporation; 2000.

Dodge HH, Wang CN, Chang CC, Ganguli M. Terminal decline and practice effects in older adults without dementia: the MoVIES project. Neurology. 2011; 77(8):722–730.10.1212/WNL. 0b013e31822b0068 [PubMed: 21832224]

Dodrill CB, Troupin AS. Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. J Nerv Ment Dis. 1975; 161(3):185–190.10.1097/00005053-197509000-00006 [PubMed: 1176975]

Duff K, Lyketsos CG, Beglinger LJ, Chelune G, Moser DJ, Arndt S, … McCaffrey RJ. Practice effects predict cognitive outcome in amnestic mild cognitive impairment. Am J Geriatr Psychiatry. 2011; 19(11):932–939.10.1097/JGP.0b013e318209dd3a [PubMed: 22024617]

Dugbartey AT, Townes BD, Mahurin RK. Equivalence of the Color Trails Test and Trail Making Test in nonnative English-speakers. Arch Clin Neuropsychol. 2000; 15(5):425–431.10.1016/S0887-6177(99)00034-7 [PubMed: 14590218]

Ferrer E, Salthouse TA, McArdle JJ, Stewart WF, Schwartz BS. Multivariate modeling of age and retest in longitudinal studies of cognitive abilities. Psychol Aging. 2005; 20(3):412–22.10.1037/0882-7974.20.3.412 [PubMed: 16248701]

Ferrer E, Salthouse TA, Stewart WF, Schwartz BS. Modeling age and retest processes in longitudinal studies of cognitive abilities. Psychol Aging. 2004; 19(2):243–259.10.1037/0882-7974.19.2.243 [PubMed: 15222818]

Flicker L. Cardiovascular risk factors, cerebrovascular disease burden, and healthy brain aging. Clin Geriatr Med. 2010; 26(1):17–27.10.1016/j.cger.2009.12.005 [PubMed: 20176290]

Frank R, Wiederholt WC, Kritz-Silverstein DK, Salmon DP, Barrett-Connor E. Effects of sequential neuropsychological testing of an elderly community-based sample. Neuroepidemiology. 1996; 15(5):257–68.10.1159/000109915 [PubMed: 8878078]

Glymour MM, Manly JJ. Lifecourse social conditions and racial and ethnic patterns of cognitive aging. Neuropsychol Rev. 2008; 18:223–254.10.1007/s11065-008-9064-z [PubMed: 18815889]

Gould, SJ. The Mismeasure of Man. New York: W.W. Norton & Company; 1996.

Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol. 2005; 162(3):267–78. [PubMed: 15987729]

Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. Neuroepidemiology. 2014a; 42:144–153. [PubMed: 24481241]

Gross AL, Sherva R, Mukherjee S, Newhouse S, Kauwe JSK, Munsie LM, Waterston LB, Bennett DA, Jones RN, Green RC, Crane PK. for the Alzheimer's Disease Neuroimaging Initiative, GENAROAD Consortium, and AD Genetics Consortium. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. Neuroepidemiology. 2014b In press.

Haan MN, Shemanski L, Jagust WJ, Manolio TA, Kuller L. The role of APOE epsilon4 in modulating effects of other risk factors for cognitive decline in elderly persons. JAMA. 1999; 282:40–46.10.1001/jama.282.1.40 [PubMed: 10404910]

Harrison JE, Buxton P, Husain M, Wise R. Short test of semantic and phonological fluency: normal performance, validity and test-retest reliability. British Journal of Clinical Psychology. 2000; 39(Pt 2):181–191.10.1348/014466500163202 [PubMed: 10895361]

Hayden KM, Jones RN, Zimmer C, Plassman BL, Browndyke JN, Pieper C, et al. Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: an evaluation of invariance between and within groups over time. Alzheimer Disease and Associated Disorders. 2011; 25(2):128–137.10.1097/WAD.0b013e3181ffa76d [PubMed: 21606904]

Heilbronner RL, Sweet JJ, Attix DK, Krull KR, Henry GK, Hart RP. Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. The Clinical Neuropsychologist. 2010; 24:1267–1278.10.1080/13854046.2010.526785 [PubMed: 21108148]

Hernan MA, Robins JM. Estimating Causal Effects from Epidemiological Data. J Epi Comm. 2006; 60:578–596.10.1136/jech.2004.029496

Hoffman L, Hofer SM, Sliwinski MJ. On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. Psychology and Aging. 2011; 26:778–791.10.1037/a0023910 [PubMed: 21639642]

Hoffmeyer-Zlotnik, JHP.; Warner, U. How to Measure Education in Cross-National Comparison: Hoffmeyer-Zlotnik/Warner-Matrix of Education as a New Instrument. In: Hoffmeyer-Zlotnik, JHP.; Harkness, JA., editors. Methodological Aspects in Cross-National Research. Mannheim: ZUMA; 2005. p. 223-240.ZUMA Nachrichten Special 11

Horton AM Jr. Neuropsychological practice effects x age: A brief note. Percept Mot Skills. 1992; 75(1):257–258. not available. [PubMed: 1528677]

Howieson D, Carlson N, Moore M, Wasserman D, Abendroth C, Payne-Murphy J, Kaye J. Trajectory of mild cognitive impairment onset. Journal of the International Neuropsychological Society. 2008; 14:192–198. [PubMed: 18282317]

Ivnik RJ, Smith GE, Lucas JA, Petersen RC, Boeve BF, Kokmen E, Tangalos EG. Testing normal older people three or four times at 1- to 2-year intervals: Defining normal variance. Neuropsychology. 1999; 13:121–127. [PubMed: 10067783]

Jacqmin-Gadda H, Fabrigoule C, Commenges D, Dartigues JF. A five-year longitudinal study of Mini-Mental State Examination in normal aging. Am J Epidemiol. 1997; 145:498–506. [PubMed: 9063339]

Jacobs DM, Sano M, Dooneief G, Marder K, Bell KL, Stern Y. Neuropsychological detection and characterization of preclinical Alzheimer's disease. Neurology. 1995; (5):957–62.10.1212/WNL. 45.5.957 [PubMed: 7746414]

Johnson JK, Gross AL, Pa J, McLaren DG, Park LQ, Manly JJ. for the Alzheimer's Disease Neuroimaging Initiative. Longitudinal change in neuropsychological performance using latent growth models: A study of mild cognitive impairment. Brain Imaging and Behavior. 2012; 6(4): 540–550.10.1007/s11682-012-9161-8 [PubMed: 22562439]

Jones RN, Rudolph JL, Inouye SK, Yang FM, Fong TG, Milberg WP, et al. Development of a unidimensional composite measure of neuropsychological functioning in older cardiac surgery patients with good measurement precision. Journal of Clinical and Experimental Neuropsychology. 2010; 32:1041–1049. [PubMed: 20446144]

Juster FT, Suzman R. An overview of the Health and Retirement Study. Journal of Human Resources. 1995; 30(Suppl):7–56.10.2307/146277

Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Philidelphia: Lea and Febiger; 1983.

Kausler, DH. Learning and memory in normal aging. San Diego, CA: Academic Press; 1994.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38(4):963–974.10.2307/2529876 [PubMed: 7168798]

Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The Aging, Demographics, and Memory Study: Study design and methods. Neuroepidemiology. 2005; 25:181–191.10.1159/000087448 [PubMed: 16103729]

Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: John Wiley & Sons; 1987.

Luchsinger JA, Tang MX, Shea S, Mayeux R. Diabetes mellitus and risk of Alzheimer's disease and dementia with stroke in a multiethnic cohort. American Journal of Epidemiology. 2001; 154:635–641.10.1093/aje/154.7.635 [PubMed: 11581097]

Luchsinger JA, Reitz C, Honig LS, Tang MX, Shea S, Mayeux R. Aggregation of vascular risk factors and risk of incident Alzheimer's disease. Neurology. 2005; 65:545–551.10.1212/01.wnl. 0000172914.08967.dc [PubMed: 16116114]

Machulda MM, Pankratz VS, Christianson TJ, Ivnik RJ, Mielke MM, Roberts RO, Knopman DS, Boeve BF, Petersen RC. Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. Clin Neuropsychol. 2013; 27(8):1247–64. Epub 2013 Sep 17. 10.1080/13854046.2013.836567 [PubMed: 24041121]

Manly JJ, Schupf N, Tang MX, Stern Y. Cognitive decline and literacy among ethnically diverse elders. Journal of Geriatric Psychiatry and Neurology. 2005; 18:213–217.10.1177/0891988705281868 [PubMed: 16306242]

Mann VA, Sasanuma S, Sakuma N, Masaki S. Sex differences in cognitive abilities: a cross-cultural perspective. Neuropsychologia. 1990; 28(10):1063–77.10.1016/0028-3932(90)90141-A [PubMed: 2267058]

Mattis, S. Dementia Rating Scale: Professional manual. Odessa, FL: Psychological Assessment Resources; 1988.

McCaffrey RJ, Onega A, Orsillo SM, Nelles WB, Haase RF. Practice effects in repeated neuropsychological assessments. The Clinical Neuropsychologist. 1992; 6:32–42.10.1080/13854049208404115

Meyer DE, Schvaneveldt RW. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. Journal of Experimental Psychology. 1971; 90(2):227–234.10.1037/h0031564 [PubMed: 5134329]

Mitrushina, M.; Boone, KB.; Razani, J.; D'Elia, LF. Handbook of normative data for neuropsychological assessment. 2. New York, NY: Oxford University Press; 2005.

Mitrushina M, Satz P. Effect of repeated administration of a neuropsychological battery in the elderly. J Clin Psychol. 1991; 47(6):790–801.10.1002/1097-4679(199111)47:6<790::AID-JCLP2270470610>3.0.CO;2-C [PubMed: 1757583]

Mungas D, Reed BR, Marshall SC, González HM. Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. Neuropsychology. 2000; 14(2):209–23. [PubMed: 10791861]

Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. Psychological Methods. 1997; 2:371–402.10.1037//1082-989X.2.4.371

Muthén, LK.; Muthén, BO. Mplus User's Guide. 7. Los Angeles, CA: Muthén & Muthén; 1998–2012.

Perani D, Bressi S, Cappa SF, Vallar G, Alberini M, … Fazio F. Evidence of multiple memory systems in the human brain: A [18F] FDG PET metabolic study. Brain. 1993; 116:903–919.10.1093/brain/116.4.903 [PubMed: 8353715]

Rabbitt PM. Does it all go together when it goes? Quarterly Journal of Experimental Psychology. 1993; 46(A):385–433.10.1080/14640749308401055 [PubMed: 8378549]

Rabbitt P, Diggle P, Smith D, Holland F, McInnes L. identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. Neuropsychologia. 2001; (39):532–543.10.1016/S0028-3932(00)00099-3 [PubMed: 11254936]

Rabbitt P, Diggle P, Holland F, McInnes L. Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. J Gerontol B Psychol Sci Soc Sci. 2004; (2):84–97.10.1093/geronb/59.2.P84

Rabbitt P, Lunn M, Wong D, Cobain M. Age and ability affect practice gains in longitudinal studies of cognitive change. J Gerontol B Psychol Sci Soc Sci. 2008; 63(4):P235–P240.10.1093/geronb/63.4.P235 [PubMed: 18689765]

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2. Thousand Oaks, CA: Sage; 2002.

Reitan R. Validity of the trail making test as an indicator of organic brain damage. Perceptual and Motor Skills. 1958; 8:271–276.10.2466/PMS.8.7.271-276

Ronnlund M, Nilsson LG. Adult life-span patterns in WAIS-R Block Design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. Intelligence. 2006; 34:63–78.10.1016/j.intell.2005.06.004

Ronnlund M, Nyberg L, Backman L, Nilsson L. Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. Psychology and Aging. 2005; 20:3–18.10.1037/0882-7974.20.1.3 [PubMed: 15769210]

Sabe L, Jason L, Juejati M, Leiguarda R, Starkstein SE. Dissociation between declarative and procedural learning in dementia and depression. Journal of Clinical and Experimental Neuropsychology. 1995; 17:841–848.10.1080/01688639508402433 [PubMed: 8847390]

Salthouse TA. When does age-related cognitive decline begin? Neurobiol Aging. 2009; 30(4):507–14.10.1016/j.neurobiolaging.2008.09.023 [PubMed: 19231028]

Salthouse, TA. Major Issues in Cognitive Aging. New York: Oxford University Press; 2010a.

Salthouse TA. Influence of age on practice effects in longitudinal neurocognitive change. Neuropsychology. 2010b; 24:563–572.10.1037/a0019026 [PubMed: 20804244]

Salthouse T, Schroeder D, Ferrer E. Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. Developmental Psychology. 2004; 40(5):813–822.10.1037/0012-1649.40.5.813 [PubMed: 15355168]

Salthouse TA, Tucker-Drob EM. Implications of short-term retest effects for the interpretation of longitudinal change. Neuropsychology. 2008; 22:800–811.10.1037/a0013091 [PubMed: 18999354]

Schaie, KW. Developmental influences on adult intelligence: The Seattle Longitudinal Study. New York: Oxford University Press; 2005.

Schneider BC, Gross AL, Bangen KJ, Skinner JC, Benitez A, Glymour MM, Sachs BC, Shihh R, Sisco S, Manly JJ, Luchsinger JA. Association of Vascular Risk Factors with Cognition in a Multiethnic Sample. Journal of Gerontology: Series B. 2014

Schrijnemaekers AM, de Jager CA, Hogervorst E, Budge MM. Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls. J Clin Exp Neuropsychol. 2006; 28:438–455.10.1080/13803390590935462 [PubMed: 16618630]

Siedlecki KL, Manly JJ, Brickman AM, Schupf N, Tang MX, Stern Y. Do neuropsychological tests have the same meaning in Spanish speakers as they do in English speakers? Neuropsychology. 2010; 24(3):402–11.10.1037/a0017515 [PubMed: 20438217]

Singer, JD.; Willet, J. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York, NY: Oxford University Press; 2003.

Sliwinski, M.; Hoffman, L.; Hofer, S. Modeling retest and aging effects in a measurement burst design. In: Molenaar, PCM.; Newell, KM., editors. Individual pathways of change: Statistical models for analyzing learning and development. Washington DC: American Psychological Association; 2010. p. 37-50.

Spreen, O.; Benton, A. Neurosensory Centre Comprehensive Examination for Aphasia. Victoria, British Columbia, Canada: University of Victoria; 1969.

Steiger, JH.; Lind, JC. Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society; Iowa, City, IA. 1980.

Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. JAMA. 1994; 271:1004–1010. doi:jama. 271.13.1004. [PubMed: 8139057]

Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, … Mayeux R. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology. 2001; 56:49–56.10.1212/WNL.56.1.49 [PubMed: 11148235]

Tang MX, Stern Y, Marder K, Bell K, Gurland B, … Mayeux R. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. JAMA. 1998; 279(10):751–755.10.1001/jama.279.10.751 [PubMed: 9508150]

Thorndike EL. Practice effects in intelligence tests. Journal of Experimental Psychology. 1922; 5:101–107.10.1037/h0074568

Tulving E, Markowitsch HJ. Episodic and declarative memory: role of the hippocampus. Hippocampus. 1998; 8(3):198–204.10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.3.CO;2-J [PubMed: 9662134]

Van der Elst W, Van Boxtel MPJ, Van Breukelen GJP, Jolles J, Aartsen M, Martin M, et al. Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's Verbal Learning Test, and the Letter Digit Substitution Test: The regression-based change approach. Journal of the International Neuropsychological Society. 2008; 14:71–80. doi:10.10170S1355617708080028. [PubMed: 18078533]

Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol Bull. 1995; 117(2):250–270.10.1037//0033-2909.117.2.250 [PubMed: 7724690]

Wechsler, D. Wechsler Adult Intelligence Scale-Revised. New York: The Psychological Corporation; 1981.

Wilson RS, Li Y, Bienias JL, Bennett DA. Cognitive decline in old age: Separating retest effects from the effects of growing older. Psychol Aging. 2006; (4):774–789.10.1037/0882-7974.21.4.774 [PubMed: 17201497]

Wilson R, Leurgans S, Boyle P, Bennett B. Cognitive decline in prodromal Alzheimer disease and mild cognitive impairment. Archives of Neurology. 2011; 68:351–356. [PubMed: 21403020]

Zehnder A, Blasi S, Berres M, Spiegel R, Monsch A. Lack of practice effects on neuropsychological tests as early cognitive markers of Alzheimer disease? American Journal of Alzheimer's Disease and Other Dementias. 2007; 22:416–426.

Appendix material: Description of the WHICAP neuropsychological battery

The Buschke Selective Reminding Test tests memory and learning; participants are asked to remember a list of semantically unrelated words (Buschke, 1973). During subsequent recall trials, participants are reminded only of the words they did not recall previously. There is a delayed recall trial, followed by a recognition trial with target and non-target words not in the original list. The Color Trail Making Test, designed as a substitute to the Trail Making Test, assesses visual attention and executive processing using letters and numbers (Reitan, 1958). WAIS-R Similarities is a test of abstract reasoning in which participants describe how sets of nouns are alike (Wechsler, 1981). The Identities/Oddities subtest of the Mattis Dementia Rating Scale also tests abstract reasoning by asking participants how two objects are similar and how a third differs from the others (Mattis, 1988). Semantic fluency for animals and phonemic fluency for F, A, and S words assess participants' ability to spontaneously generate semantically related words (animals) and words starting with certain letters, respectively, within 60 seconds for each trial (Harrison et al. 2000). The 15-item Boston Naming Test was used to measure confrontation naming and language (Kaplan, Goodglass, & Weintraub, 1983). In a repetition task, participants were asked to repeat eight phrases. Comprehension was assessed by asking participants eight general knowledge questions (e.g., Will a stone sink in water?) and four questions from a short four-sentence story from the Boston Diagnostic Aphasia Examination (Spreen & Benton, 1969). During visits early in the study, two cancellation tasks involving detection of a shape form and a consonant trigram (TMX) from arrays of shape and phonemic distractions, respectively, were used to assess attention (Ben-Yishai et al., 1974).

Appendix Material: Measurement invariance of cognitive performance over time

Method

We tested longitudinal measurement invariance of the factors among participants assessed at baseline and whose second study visit was between 1.5 and 2.5 years later (median: 2.1 years) using multiple group confirmatory factor analysis models. This set of analyses was used to confirm the factors were measuring cognitive performance in the same way over time (Bontempo & Hofer, 2007; Borsboom et al., 2008). We tested configural, metric, scalar, and strict measurement invariance of the cognitive factors. Good fit of a model with configural invariance, in which factor loadings, intercepts, and residual variances of cognitive tests are free to vary over time, indicates the factors exist in both groups. Metric invariance tests whether each cognitive test measures a factor in the same way in each group by constraining factor loadings to be equal. Scalar invariance further tests whether intercepts of cognitive tests are equal across group, from which one would infer that each cognitive test measures the cognitive factor in the same location in latent variable space across group. Finally, in strict invariance factor loadings, intercepts, and residual variances of tests are

fixed to be equal across groups. We compared different levels of invariance with likelihood ratio tests and by examining absolute fit with the root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) and comparative fit index (CFI; Bentler, 1990).

Results

We examined configural, metric, scalar, and strict invariance of the factor structure over time. Configural invariance was met based on excellent model fit (RMSEA=0.057, CFI=0.969). Although the model for metric invariance fit significantly worse than for configural invariance (χ^2=44.8, df=12, p<0.01), absolute fit was still excellent (RMSEA=0.054, CFI=0.968). The same was true for scalar invariance tested against metric invariance (χ^2=209.9, df=12, p<0.01; RMSEA=0.057, CFI=0.961). Strict invariance was met (compared to scalar invariance: χ^2=11.9, df=9, p=0.31; RMSEA=0.057, CFI=0.961). Model modification indices showed most of the model misfit was attributable to factor loadings on the language factor across time. Memory and executive functioning demonstrated stronger evidence of metric invariance than language. Given acceptable absolute model fit through strict measurement invariance over time, we concluded the factors were invariant over time. The method by which we constructed factor scores imposed strict measurement invariance, thus assuring that retest effects are not related to longitudinal measurement non-invariance of the factor structure.

Appendix Table 1

Means and standard errors of random effects for level and change in cognitive performance in overall and multiple group models: Results from WHICAP (N=4,073)

Sample	Cognitive outcome	Level (Age 75)			Annual rate of change		
		Mean	Standard error	Mean/ Standard error	Mean	Standard error	Mean/ Standard error
Overall sample							
	General cognitive performance	45.98	0.18	261.57	−0.47	0.01	−69.15
	Memory	50.33	0.18	275.41	−0.54	0.01	−55.67
	Executive functioning	45.41	0.16	281.23	−0.42	0.01	−56.15
	Language	45.46	0.20	225.36	−0.38	0.01	−43.36
Race/ethnicity							
Non-Hispanic White (n=1013)							
	General cognitive performance	51.08	0.58	87.93	−0.66	0.02	−28.11
	Memory	53.98	0.80	67.77	−0.73	0.05	−16.31
	Executive functioning	50.44	0.48	104.58	−0.64	0.03	−23.27
	Language	50.14	0.86	58.24	−0.53	0.05	−11.65
Non-Hispanic Black (n=1372)							
	General cognitive performance	45.39	0.42	107.71	−0.60	0.02	−32.39
	Memory	49.19	0.45	108.56	−0.69	0.03	−20.60
	Executive functioning	44.81	0.38	118.02	−0.51	0.03	−20.67
	Language	45.53	0.51	89.92	−0.48	0.03	−15.91

Sample	Cognitive outcome	Level (Age 75)			Annual rate of change		
		Mean	Standard error	Mean/ Standard error	Mean	Standard error	Mean/ Standard error
Hispanic (n=1688)							
	General cognitive performance	43.42	0.34	129.48	−0.54	0.01	−38.78
	Memory	48.87	0.35	140.10	−0.60	0.02	−26.25
	Executive functioning	43.02	0.30	143.82	−0.46	0.02	−24.42
	Language	42.49	0.38	111.69	−0.45	0.02	−20.44
Sex							
Male (N=1283)							
	General cognitive performance	46.08	0.42	109.73	−0.60	0.02	−33.68
	Memory	49.35	0.46	106.49	−0.67	0.04	−18.63
	Executive functioning	46.16	0.38	122.55	−0.58	0.03	−22.15
	Language	46.19	0.53	87.59	−0.48	0.03	−14.54
Female (N=2790)							
	General cognitive performance	45.86	0.35	130.47	−0.58	0.01	−54.15
	Memory	50.66	0.37	136.63	−0.65	0.02	−33.44
	Executive functioning	45.07	0.32	139.99	−0.50	0.01	−37.12
	Language	45.00	0.44	102.47	−0.49	0.02	−27.57
Age							
Under 75 (n=1667)							
	General cognitive performance	50.31	0.52	97.68	−0.56	0.02	−35.83
	Memory	54.86	0.57	95.64	−0.64	0.03	−20.74
	Executive functioning	48.64	0.47	104.47	−0.52	0.02	−24.42
	Language	49.27	0.67	73.07	−0.45	0.03	−16.05
Age 75 to 80 (n=1158)							
	General cognitive performance	45.82	0.43	106.50	−0.62	0.02	−34.67
	Memory	49.93	0.45	110.65	−0.68	0.03	−23.31
	Executive functioning	45.10	0.38	120.41	−0.53	0.02	−24.08
	Language	45.38	0.51	88.26	−0.50	0.03	−16.28
80 and over (n=1248)							
	General cognitive performance	39.47	0.37	107.29	−0.85	0.03	−32.92
	Memory	44.00	0.36	122.47	−0.85	0.04	−20.52
	Executive functioning	40.88	0.33	123.47	−0.60	0.03	−18.92
	Language	39.31	0.40	98.37	−0.73	0.04	−20.63
Years of education							
7 years or less (n=1883)							
	General cognitive performance	41.63	0.28	148.08	−0.53	0.01	−41.25
	Memory	47.29	0.30	158.67	−0.59	0.02	−27.75
	Executive functioning	41.28	0.25	168.32	−0.42	0.02	−22.73
	Language	40.45	0.30	137.00	−0.41	0.02	−19.92
8 or more years (n=2179)							
	General cognitive performance	50.19	0.36	138.16	−0.65	0.02	−44.29

Sample	Cognitive outcome	Level (Age 75)			Annual rate of change		
		Mean	Standard error	Mean/ Standard error	Mean	Standard error	Mean/ Standard error
	Memory	53.32	0.45	118.60	−0.73	0.03	−26.71
	Executive functioning	49.46	0.33	151.87	−0.62	0.02	−34.84
	Language	50.06	0.50	100.96	−0.55	0.03	−21.40
APOE ε4 status							
No ε4 allele (n=3151)							
	General cognitive performance	45.74	0.31	147.42	−0.54	0.01	−51.44
	Memory	50.15	0.34	147.20	−0.60	0.02	−29.77
	Executive functioning	45.22	0.28	160.39	−0.49	0.01	−35.13
	Language	45.05	0.39	116.27	−0.43	0.02	−22.25
Possesses ε4 allele (n=922)							
	General cognitive performance	46.70	0.63	73.85	−0.72	0.02	−32.52
	Memory	50.88	0.63	80.48	−0.79	0.03	−23.22
	Executive functioning	46.10	0.56	82.71	−0.63	0.03	−22.49
	Language	46.52	0.77	60.14	−0.63	0.03	−21.08
Baseline cognitive status							
Quartile 1 (n=1036)							
	General cognitive performance	32.56	0.22	147.40	−0.30	0.03	−9.83
	Memory	38.14	0.22	173.33	−0.19	0.04	−4.72
	Executive functioning	35.54	0.21	170.22	−0.22	0.04	−5.41
	Language	33.11	0.25	132.54	−0.27	0.04	−6.92
Quartile 2 (n=1013)							
	General cognitive performance	43.28	0.27	161.40	−0.49	0.02	−20.18
	Memory	47.47	0.33	142.64	−0.48	0.04	−12.13
	Executive functioning	42.77	0.27	158.56	−0.40	0.03	−13.83
	Language	42.42	0.33	128.20	−0.42	0.04	−12.08
Quartile 3 (n=1013)							
	General cognitive performance	49.56	0.30	167.47	−0.59	0.03	−22.86
	Memory	53.37	0.39	135.31	−0.65	0.04	−16.65
	Executive functioning	47.65	0.31	156.26	−0.51	0.03	−17.00
	Language	48.54	0.44	110.17	−0.47	0.04	−10.67
Quartile 4 (n=1011)							
	General cognitive performance	58.27	0.41	142.98	−0.71	0.02	−37.36
	Memory	62.22	0.64	96.75	−0.83	0.04	−20.31
	Executive functioning	55.44	0.35	156.67	−0.66	0.02	−26.98
	Language	57.06	0.61	93.11	−0.54	0.03	−16.61
Cardiovascular risk burden							
Zero (n=884)							
	General cognitive performance	47.53	0.56	84.47	−0.62	0.02	−32.90
	Memory	51.25	0.54	95.88	−0.69	0.03	−23.21
	Executive functioning	47.30	0.52	90.38	−0.55	0.02	−23.89

Sample	Cognitive outcome	Level (Age 75)			Annual rate of change		
		Mean	Standard error	Mean/Standard error	Mean	Standard error	Mean/Standard error
	Language	47.21	0.67	70.21	−0.50	0.03	−14.64
One (n=1414)							
	General cognitive performance	45.95	0.53	86.24	−0.58	0.02	−26.40
	Memory	50.73	0.61	83.40	−0.67	0.04	−15.32
	Executive functioning	45.03	0.47	95.34	−0.50	0.03	−18.96
	Language	45.11	0.63	71.75	−0.45	0.03	−14.64
Two (n=1167)							
	General cognitive performance	44.29	0.64	69.65	−0.55	0.02	−26.47
	Memory	48.56	0.67	72.52	−0.58	0.04	−13.45
	Executive functioning	44.29	0.57	77.43	−0.52	0.03	−16.63
	Language	43.86	0.83	52.59	−0.50	0.04	−13.34
Three or more (n=608)							
	General cognitive performance	43.39	0.94	46.38	−0.48	0.05	−10.49
	Memory	47.76	1.04	46.03	−0.51	0.09	−5.52
	Executive functioning	43.29	0.78	55.45	−0.49	0.05	−9.44
	Language	43.31	1.12	38.64	−0.38	0.08	−5.09

Appendix Table 2

Retest effects for general and domain-specific cognitive performance: Results from WHICAP (N=4,073)

Parameter	General cognitive performance Estimate (95% CI)	Memory Estimate (95% CI)	Executive functioning Estimate (95% CI)	Language Estimate (95% CI)
Race/ethnicity				
Non-Hispanic White (n=1013)	4.67* (2.69, 6.65)	5.48* (2.70, 8.26)	3.14* (1.26, 5.02)	5.39* (3.41, 7.37)
Non-Hispanic Black (n=1372)	4.09* (2.48, 5.70)	3.71* (2.00, 5.42)	3.80* (2.33, 5.27)	3.74* (1.62, 5.86)
Hispanic (n=1688)	3.26* (2.20, 4.32)	3.16* (2.02, 4.30)	2.10* (1.00, 3.20)	3.36* (2.05, 4.67)
Group differences				
White - Black	0.58 (−1.97, 3.13)	1.77 (−1.50, 5.04)	−0.66 (−3.03, 1.71)	1.65 (−1.25, 4.55)
White - Hispanic	1.41 (−0.82, 3.64)	2.33 (−0.69, 5.35)	1.04 (−1.14, 3.22)	2.03 (−0.34, 4.40)
Black - Hispanic	0.84 (−1.10, 2.78)	0.56 (−1.50, 2.62)	1.70 (−0.12, 3.52)	0.39 (−2.10, 2.88)
Pseudo-R^2	0.896	0.803	0.814	0.856
Age				
Under 75 (n=1667)	2.84* (1.19, 4.49)	2.74* (0.84, 4.64)	2.61* (1.02, 4.20)	3.34* (1.36, 5.32)
Age 75 to 80 (n=1158)	3.39* (1.72, 5.06)	3.41* (1.55, 5.27)	2.77* (1.16, 4.38)	3.04* (0.55, 5.53)
80 and over (n=1248)	3.79* (2.65, 4.93)	3.22* (2.12, 4.32)	1.93* (0.91, 2.95)	3.99* (2.77, 5.21)
Group differences				
Under 75 – (75 to 80)	−0.55 (−2.88, 1.78)	−0.67 (−3.34, 2.00)	−0.16 (−2.41, 2.09)	0.30 (−2.88, 3.48)
Under 75 – Over 80	−0.95 (−2.95, 1.05)	−0.48 (−2.68, 1.72)	0.68 (−1.20, 2.56)	−0.65 (−2.98, 1.68)

Parameter	General cognitive performance Estimate (95% CI)	Memory Estimate (95% CI)	Executive functioning Estimate (95% CI)	Language Estimate (95% CI)
(75 to 80) – Over 80	−0.40 (−2.42, 1.62)	0.19 (−1.99, 2.37)	0.84 (−1.06, 2.74)	−0.95 (−3.73, 1.83)
Pseudo-R²	0.898	0.803	0.814	0.858
Sex				
Male (N=1283)	3.67* (2.04, 5.30)	3.21* (1.33, 5.09)	3.27* (1.78, 4.76)	4.43* (2.47, 6.39)
Female (N=2790)	4.54* (3.38, 5.70)	4.61* (3.39, 5.83)	3.35* (2.17, 4.53)	4.49* (3.14, 5.84)
Group differences				
Female - Male	0.87 (−1.13, 2.87)	1.40 (−0.83, 3.63)	0.08 (−1.82, 1.98)	0.06 (−2.31, 2.43)
Pseudo-R²	0.896	0.803	0.813	0.856
Predominant language spoken at home				
English (n=1572)	3.53* (2.10, 4.96)	3.21* (1.70, 4.72)	2.47* (0.90, 4.04)	3.94* (2.20, 5.68)
Non-English (N=1009)	5.53* (3.94, 7.12)	5.76* (3.78, 7.74)	4.76* (3.19, 6.33)	5.66* (3.68, 7.64)
Group differences				
English - Non-English	2.01 (−0.11, 4.13)	2.55 (0.06, 5.04)	2.28 (0.05, 4.51)	1.72 (−0.93, 4.37)
Pseudo-R²	0.88	0.781	0.807	0.846
Years of education				
7 years or less (n=1883)	3.24* (2.36, 4.12)	3.01* (2.05, 3.97)	2.24* (1.36, 3.12)	3.23* (2.23, 4.23)
8 or more years (n=2179)	4.36* (3.20, 5.52)	4.56* (3.07, 6.05)	3.41* (2.29, 4.53)	4.51* (3.22, 5.80)
Group differences				
(8 or more years) - (7 years or less)	1.12 (−0.33, 2.57)	1.55 (−0.23, 3.33)	1.17 (−0.24, 2.58)	1.29 (−0.34, 2.92)
Pseudo-R²	0.896	0.803	0.813	0.856
APOE ε4 status				
No ε4 allele (n=3151)	4.32* (3.22, 5.42)	4.22* (3.00, 5.44)	3.34* (2.24, 4.44)	4.55* (3.33, 5.77)
Possesses ε4 allele (n=922)	3.99 (1.95, 6.03)	4.00* (2.04, 5.96)	3.06* (1.20, 4.92)	3.93* (1.28, 6.58)
Group differences				
(Possesses ε4) - (no ε4)	−0.33 (−2.64, 1.98)	−0.23 (−2.54, 2.08)	−0.27 (−2.43, 1.89)	−0.63 (−3.55, 2.29)
Pseudo-R²	0.897	0.804	0.815	0.857
Baseline cognitive status				
Quartile 1 (n=1036)	2.80* (2.17, 3.43)	2.62* (1.95, 3.29)	1.25* (0.62, 1.88)	1.87* (1.11, 2.63)
Quartile 2 (n=1013)	0.36 (−0.25, 0.97)	0.32 (−0.64, 1.28)	0.42 (−0.38, 1.22)	0.89 (−0.29, 2.07)
Quartile 3 (n=1013)	−0.21 (−0.84, 0.42)	0.06 (−0.80, 0.92)	−0.11 (−1.25, 1.03)	−0.06 (−1.29, 1.17)
Quartile 4 (n=1011)	0.52 (−1.69, 2.73)	0.40 (−2.97, 3.77)	0.42 (−1.38, 2.22)	1.11 (−2.16, 4.38)
Group differences				
Quartile1 – Quartile2	2.43* (1.57, 3.29)	2.30* (1.14, 3.46)	0.83 (−0.19, 1.85)	0.98 (−0.43, 2.39)
Quartile1 – Quartile3	3.01* (2.13, 3.89)	2.55* (1.47, 3.63)	1.37* (0.08, 2.66)	1.93* (0.48, 3.38)
Quartile1 – Quartile4	2.28 (−0.01, 4.57)	2.21 (−1.24, 5.66)	0.84 (−1.06, 2.74)	0.75 (−2.60, 4.10)
Quartile2 – Quartile3	0.57 (−0.29, 1.43)	0.26 (−1.01, 1.53)	0.54 (−0.85, 1.93)	0.95 (−0.76, 2.66)
Quartile2 – Quartile4	−0.16 (−2.45, 2.13)	−0.08 (−3.59, 3.43)	0.00 (−1.98, 1.98)	−0.22 (−3.69, 3.25)
Quartile3 – Quartile4	−0.73 (−3.02, 1.56)	−0.34 (−3.83, 3.15)	−0.53 (−2.67, 1.61)	−1.17 (−4.66, 2.32)
Pseudo-R²	0.889	0.799	0.804	0.85
Cardiovascular risk burden				

Parameter	General cognitive performance Estimate (95% CI)	Memory Estimate (95% CI)	Executive functioning Estimate (95% CI)	Language Estimate (95% CI)
Zero (n=884)	4.06* (1.75, 6.37)	4.71* (2.40, 7.02)	2.42* (0.09, 4.75)	4.17* (1.33, 7.01)
One (n=1414)	4.69* (3.08, 6.30)	4.05* (2.21, 5.89)	4.03* (2.50, 5.56)	4.97* (3.07, 6.87)
Two (n=1167)	3.92* (2.14, 5.70)	3.88* (1.94, 5.82)	2.93* (0.95, 4.91)	4.09* (1.95, 6.23)
Three or more (n=608)	3.40* (0.42, 6.38)	3.35 (−0.57, 7.27)	2.95* (0.52, 5.38)	3.45 (−0.57, 7.47)
Group differences				
Zero – One	−0.64 (−3.46, 2.18)	0.67 (−2.29, 3.63)	−1.61 (−4.41, 1.19)	−0.80 (−4.21, 2.61)
Zero – Two	0.14 (−2.78, 3.06)	0.84 (−2.18, 3.86)	−0.51 (−3.59, 2.57)	0.08 (−3.47, 3.63)
Zero – Three	0.66 (−3.12, 4.44)	1.36 (−3.19, 5.91)	−0.52 (−3.89, 2.85)	0.72 (−4.20, 5.64)
One – Two	0.77 (−1.62, 3.16)	0.17 (−2.50, 2.84)	1.10 (−1.41, 3.61)	0.88 (−1.98, 3.74)
One – Three	1.29 (−2.10, 4.68)	0.70 (−3.63, 5.03)	1.08 (−1.80, 3.96)	1.52 (−2.91, 5.95)
Two – Three	0.52 (−2.95, 3.99)	0.52 (−3.85, 4.89)	−0.01 (−3.15, 3.13)	0.64 (−3.91, 5.19)
Pseudo-R^2	0.897	0.803	0.815	0.857

Retest here is parameterized as the square root of the number of prior test occasions to accommodate a diminishing returns conceptualization of practice. Multilevel models of changes for general cognitive performance, memory, executive functioning, and language using time in study as the timescale.

The retest parameters correspond to β_2 parameters in equation 4, and group differences correspond to parameter β_4.
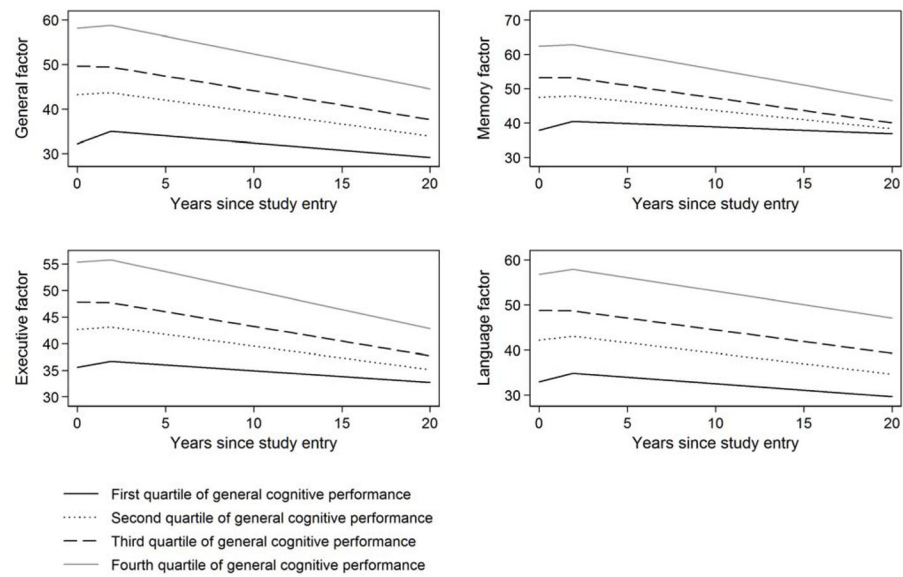
Figure 1.
Model-estimated trajectories of general and domain-specific cognitive functioning by
baseline cognitive performance level: Results from WHICAP (N=4,073)
Graphical depiction of model-estimated trajectories of general and domain-specific
cognitive function to illustrate effect of practice effects in cognitive aging. The second study
visit occurred on average 1.9 years after the first study visit, so the practice effect depicted is
1.9 years after study entry in figures.

Table 1

Demographic characteristics: Results from WHICAP (N=4,073)

Variable	WHICAP sample (N=4,073)	Observed range
Age, mean (SD)	77.3 (7.0)	(63.0, 103.0)
Male, n (%)	1283 (31.5)	
Recruitment cohort, n (%)		
1992	1932 (47.4)	
1999	2141 (52.6)	
Years of follow-up, median (interquartile interval)	3.9	(1.0, 7.8)
Years between first and second testing, median (interquartile interval)	1.9	(1.3, 2.2)
Number of study visits, median (interquartile interval)	3.0	(2.0, 4.0)
Race/ethnicity, n (%)		
Non-Hispanic White	1013 (24.9)	
Non-Hispanic Black	1372 (33.7)	
Hispanic	1688 (41.4)	
Years of education, n (%)		
7 years or less	1883 (46.4)	
8 or more years	2179 (53.6)	
APOE ε4 status, n (%)		
Possesses ε4 allele	3151 (77.4)	
No ε4 allele	922 (22.6)	
Vascular risk factors, n (%)		
None	884 (21.7)	
One	1414 (34.7)	
Two	1167 (28.7)	
Three or more	608 (14.9)	
Cognitive factor scores, mean (SD)		
General cognitive performance	46.3 (12.0)	(−11.4, 87.5)
Memory	55.2 (13.4)	(13.8, 96.0)
Executive functioning	44.8 (11.0)	(11.7, 84.8)
Language	44.6 (10.7)	(5.7, 80.4)
Predominant language spoken at home, n English (%)	1626 (59.6)	
Quartile of general cognitive performance, n (%)		
First (lowest) quartile	1027 (25.2)	
Second quartile	1018 (25.0)	
Third quartile	1015 (24.9)	
Fourth (highest) quartile	1013 (24.9)	

SD: standard deviation

Table 2

Retest effects and slopes for general and domain-specific cognitive performance: Results from WHICAP (N=4,073)

Cognitive outcome	Retest effect (95% CI)	Mean annual rate of decline (95% CI)	Ratio of retest to annual rate of decline
General cognitive performance	6.01 (4.58, 7.43)	−0.47 (−0.48, −0.46)	12.78
Memory	5.72 (4.21, 7.22)	−0.54 (−0.56, −0.52)	10.57
Executive functioning	4.46 (3.15, 5.78)	−0.42 (−0.44, −0.41)	10.55
Language	6.44 (4.74, 8.13)	−0.38 (−0.40, −0.36)	16.93

Parallel process latent growth models of changes in global cognition, memory, and executive functioning score changes using time in study as the timescale. Each cognitive score was scaled to have a mean of 50 and standard deviation of 10 at the baseline study visit. The annual rate of decline is the mean of the random slope in the model. The ratio of retest and slope reflects the relative magnitude of the retest effect compared to subsequent annual cognitive decline. The retest parameters correspond to β_2 parameters in equation 1. The model-estimated proportion of total variance attributable to between-persons differences was 86%, 74%, 78%, and 81% for general cognitive performance, memory, executive functioning, and language, respectively.

Table 3

Retest effects for general and domain-specific cognitive performance: Results from WHICAP (N=4,073)

Parameter	General cognitive performance Estimate (95% CI)	Memory Estimate (95% CI)	Executive functioning Estimate (95% CI)	Language Estimate (95% CI)
Race/ethnicity				
Non-Hispanic White (n=1013)	4.95* (1.87, 8.03)	5.81* (1.38, 10.24)	3.34* (0.71, 5.97)	5.65* (2.26, 9.04)
Non-Hispanic Black (n=1372)	4.28* (1.69, 6.87)	3.85* (1.26, 6.44)	3.98* (1.61, 6.35)	3.92* (0.69, 7.15)
Hispanic (n=1688)	3.49* (1.88, 5.10)	3.35* (1.66, 5.04)	2.19* (0.52, 3.86)	3.56* (1.54, 5.58)
Group differences				
White - Black	0.67 (-3.35, 4.69)	1.97 (-3.15, 7.09)	-0.64 (-4.19, 2.91)	1.74 (-2.94, 6.42)
White - Hispanic	1.46 (-2.01, 4.93)	2.46 (-2.28, 7.20)	1.14 (-1.98, 4.26)	2.09 (-1.85, 6.03)
Black - Hispanic	0.79 (-2.27, 3.85)	0.50 (-2.58, 3.58)	1.78 (-1.12, 4.68)	0.36 (-3.46, 4.18)
Pseudo-R^2	0.90	0.80	0.81	0.86
Age				
Under 75 (n=1667)	2.93* (0.09, 5.77)	2.84 (-0.02, 5.70)	2.64 (-0.05, 5.33)	3.39 (-0.24, 7.02)
Age 75 to 80 (n=1158)	3.66* (1.31, 6.01)	3.59* (0.92, 6.26)	2.92* (0.65, 5.19)	3.36 (-0.15, 6.87)
80 and over (n=1248)	4.02* (2.41, 5.63)	3.37* (1.78, 4.96)	2.10* (0.71, 3.49)	4.21* (2.43, 5.99)
Group differences				
Under 75 – (75 to 80)	-0.73 (-4.41, 2.95)	-0.74 (-4.64, 3.16)	-0.28 (-3.81, 3.25)	0.03 (-5.03, 5.09)
Under 75 – Over 80	-1.09 (-4.36, 2.18)	-0.53 (-3.80, 2.74)	0.54 (-2.50, 3.58)	-0.82 (-4.86, 3.22)
(75 to 80) – Over 80	-0.36 (-3.22, 2.50)	0.22 (-2.88, 3.32)	0.82 (-1.85, 3.49)	-0.85 (-4.79, 3.09)
Pseudo-R^2	0.90	0.80	0.81	0.86
Sex				
Male (N=1283)	3.92* (1.41, 6.43)	3.45* (0.84, 6.06)	3.42* (1.15, 5.69)	4.65* (1.61, 7.69)
Female (N=2790)	4.77* (3.01, 6.53)	4.80* (2.94, 6.66)	3.50* (1.78, 5.22)	4.72* (2.54, 6.90)
Group differences				
Female - Male	0.85 (-2.21, 3.91)	1.34 (-1.85, 4.53)	0.07 (-2.77, 2.91)	0.07 (-3.67, 3.81)
Pseudo-R^2	0.88	0.79	0.82	0.85
Predominant language spoken at home				
English (n=1572)	3.69* (1.48, 5.90)	3.37* (1.10, 5.64)	2.51* (0.12, 4.90)	4.04* (1.14, 6.94)

Parameter	General cognitive performance Estimate (95% CI)	Memory Estimate (95% CI)	Executive functioning Estimate (95% CI)	Language Estimate (95% CI)
Non-English (N=1009)	5.69* (2.81, 8.57)	5.90* (2.61, 9.19)	4.88* (2.21, 7.55)	5.80* (2.31, 9.29)
Group differences				
English - Non-English	2.00 (−1.63, 5.63)	2.53 (−1.47, 6.53)	2.38 (−1.21, 5.97)	1.76 (−2.79, 6.31)
Pseudo-R^2	0.88	0.781	0.807	0.846
Years of education				
7 years or less (n=1883)	3.41* (2.10, 4.72)	3.16* (1.81, 4.51)	2.30* (1.01, 3.59)	3.37* (1.78, 4.96)
8 or more years (n=2179)	4.58* (2.76, 6.40)	4.76* (2.53, 6.99)	3.57* (1.85, 5.29)	4.71* (2.46, 6.96)
Group differences				
(8 or more years) - (7 years or less)	1.17 (−1.08, 3.42)	1.60 (−1.01, 4.21)	1.28 (−0.88, 3.44)	1.34 (−1.42, 4.10)
Pseudo-R^2	0.90	0.80	0.81	0.86
APOE ε4 status				
No ε4 allele (n=3151)	4.57* (2.92, 6.22)	4.43* (2.67, 6.19)	3.49* (1.92, 5.06)	4.80* (2.84, 6.76)
Possesses ε4 allele (n=922)	4.22* (1.20, 7.24)	4.23* (1.09, 7.37)	3.21* (0.52, 5.90)	4.07* (0.15, 7.99)
Group differences				
(Possesses ε4) - (no ε4)	−0.19 (−3.80, 3.42)	−0.28 (−3.40, 2.84)	−0.73 (−5.12, 3.66)	
Pseudo-R^2	0.90	0.80	0.82	0.86
Baseline cognitive status				
Quartile 1 (n=1036)	2.75* (1.75, 3.75)	2.54* (1.44, 3.64)	1.17* (0.21, 2.13)	1.85* (0.65, 3.05)
Quartile 2 (n=1013)	0.42 (−0.50, 1.34)	0.37 (−1.14, 1.88)	0.47 (−0.75, 1.69)	0.89 (−0.91, 2.69)
Quartile 3 (n=1013)	−0.23 (−1.27, 0.81)	0.04 (−1.19, 1.27)	−0.17 (−1.99, 1.65)	−0.10 (−2.10, 1.90)
Quartile 4 (n=1011)	0.58 (−1.85, 3.01)	0.46 (−3.36, 4.28)	0.45 (−1.65, 2.55)	1.16 (−2.43, 4.75)
Group differences				
Quartile1 – Quartile2	2.33* (0.96, 3.70)	2.17* (0.29, 4.05)	0.70 (−0.85, 2.25)	0.95 (−1.23, 3.13)
Quartile1 – Quartile3	2.98* (1.53, 4.43)	2.51* (0.86, 4.16)	1.35 (−0.71, 3.41)	1.95 (−0.38, 4.28)
Quartile1 – Quartile4	2.18 (−0.47, 4.83)	2.08 (−1.90, 6.06)	0.72 (−1.59, 3.03)	0.69 (−3.09, 4.47)
Quartile2 – Quartile3	0.65 (−0.74, 2.04)	0.33 (−1.61, 2.27)	0.65 (−1.53, 2.83)	0.99 (−1.70, 3.68)
Quartile2 – Quartile4	−0.16 (−2.77, 2.45)	−0.09 (−4.21, 4.03)	0.02 (−2.39, 2.43)	−0.26 (−4.28, 3.76)
Quartile3 – Quartile4	−0.80 (−3.45, 1.85)	−0.42 (−4.44, 3.60)	−0.62 (−3.38, 2.14)	−1.26 (−5.38, 2.86)
Pseudo-R^2	0.89	0.80	0.80	0.85

Parameter	General cognitive performance	Memory	Executive functioning	Language
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Cardiovascular risk burden				
Zero (n=884)	4.30* (1.12, 7.48)	5.02* (1.86, 8.18)	2.46 (−0.62, 5.54)	4.33* (0.23, 8.43)
One (n=1414)	4.97* (2.38, 7.56)	4.27* (1.47, 7.07)	4.24* (1.79, 6.69)	5.22* (2.14, 8.30)
Two (n=1167)	4.15* (1.39, 6.91)	4.02* (0.84, 7.20)	3.14* (0.36, 5.92)	4.36* (0.99, 7.73)
Three or more (n=608)	3.48 (−0.91, 7.87)	3.42 (−2.50, 9.34)	2.97 (−0.36, 6.30)	3.51 (−1.84, 8.86)
Group differences				
Zero – One	−0.68 (−4.78, 3.42)	0.75 (−3.46, 4.96)	−1.78 (−5.72, 2.16)	−0.88 (−6.02, 4.26)
Zero – Two	0.15 (−4.04, 4.34)	1.00 (−3.49, 5.49)	−0.68 (−4.82, 3.46)	−0.03 (−5.34, 5.28)
Zero – Three	0.82 (−4.59, 6.23)	1.60 (−5.10, 8.30)	−0.51 (−5.06, 4.04)	0.82 (−5.92, 7.56)
One – Two	0.83 (−2.95, 4.61)	0.25 (−3.98, 4.48)	1.10 (−2.60, 4.80)	0.86 (−3.71, 5.43)
One – Three	1.49 (−3.61, 6.59)	0.85 (−5.70, 7.40)	1.26 (−2.88, 5.40)	1.71 (−4.46, 7.88)
Two – Three	0.67 (−4.50, 5.84)	0.60 (−6.12, 7.32)	0.17 (−4.16, 4.50)	0.85 (−5.46, 7.16)
Pseudo-R^2	0.90	0.80	0.82	0.86

*
$p < 0.05$

Multilevel models of changes for general cognitive performance, memory, executive functioning, and language using time in study as the timescale. The retest parameters correspond to β2 parameters in equation 4, and group differences correspond to parameter β4. Retest effects are parameterized here as the jump in performance between the first and subsequent testing occasions.