



Published in final edited form as:

*Alzheimers Dement.* 2023 September ; 19(9): 4028–4036. doi:10.1002/alz.13067.

## Pragmatic approaches to handling practice effects in longitudinal cognitive aging research

Ruijia Chen, ScD<sup>1</sup>, Camilla Calmasini, MS<sup>1</sup>, Kaitlin Swinnerton, MS<sup>1</sup>, Jingxuan Wang, Mphil<sup>1</sup>, Sebastien Haneuse, PhD<sup>2</sup>, Sarah F Ackley, PhD<sup>1</sup>, Andrew K Hirst, MS<sup>3</sup>, Eleanor Hayes-Larson, PhD<sup>4</sup>, Kristen M George, PhD<sup>5</sup>, Rachel Peterson, PhD<sup>6</sup>, Yenee Soh, ScD<sup>3</sup>, Lisa L Barnes, PhD<sup>7</sup>, Elizabeth Rose Mayeda, PhD<sup>4</sup>, Paola Gilsanz, ScD<sup>3</sup>, Dan M Mungas, PhD<sup>8</sup>, Rachel A Whitmer, PhD<sup>5</sup>, Maria M Corrada, ScD<sup>9</sup>, M Maria Glymour, ScD<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158, United States

<sup>2</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, United States

<sup>3</sup>Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612, United States

<sup>4</sup>Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095, United States

<sup>5</sup>Department of Public Health Sciences, University of California, Davis, Davis, CA 95616, United States

<sup>6</sup>School of Public Health and Community Health Sciences, University of Montana, Missoula, MT 59812, United States

<sup>7</sup>Department of Neurological Sciences and Rush Alzheimer's Disease Center, Rush Medical College, Chicago, IL 60612, United States

<sup>8</sup>Department of Neurology, University of California, Davis, Davis, CA 95616, United States

<sup>9</sup>Department of Neurology, University of California, Irvine, Irvine, CA 92697, United States

### Abstract

**INTRODUCTION:** The challenge of accounting for practice effects (PEs) when modeling cognitive change was amplified by the COVID-19 pandemic, which introduced period and mode effects that may bias the estimation of cognitive trajectory.

**METHODS:** In three Kaiser Permanente Northern California prospective cohorts, we compared predicted cognitive trajectories and the association of grip strength with cognitive decline using

---

**Corresponding Author:** Dr. Maria Glymour, Department of Epidemiology and Biostatistics, University of California, San Francisco, UCSF Box 0560, 550 16<sup>th</sup> Street, 2<sup>nd</sup> Floor, San Francisco, CA 94158, tel : 415/514-8014, maria.glymour@ucsf.edu.

**Conflict of Interests:** The authors declare no conflicts of interest.

Consent Statement

All human subjects provided informed consent

three approaches: 1) no acknowledgement of PE; 2) inclusion of a wave indicator; and 3) constraining the PE based on a preliminary model (APM) fit using a subset of the data.

**RESULTS:** APM-based correction for PEs based on balanced, pre-pandemic data and using current age as the timescale produced the smallest discrepancy between within-person and between-person estimated age effects. Estimated associations between grip strength and cognitive decline were not sensitive to the approach used.

**DISCUSSION:** Constraining PEs based on a preliminary model is a flexible, pragmatic approach allowing meaningful interpretation of cognitive change.

### Keywords

Practice effects; cognitive function; aging; period effects; pandemic

---

### Background

Modeling longitudinal trajectories of cognition is essential in cognitive aging and dementia research but is often complicated by practice effects (PEs).<sup>1</sup> PEs are improvements in test performance from repeated exposure to test materials rather than changes in the value of the underlying cognitive construct.<sup>2</sup> PEs are common, but the magnitude may vary with time between tests, test characteristics, and individual characteristics, including age, education, and comorbidities.<sup>3-7</sup>

Failure to account for PEs can lead to biased results and incorrect interpretation of longitudinal models.<sup>8-11</sup> PEs may offset true cognitive decline, leading to an underestimation of the rate of cognitive decline. Without correcting for PEs, estimated treatment effects, especially without an appropriate comparison group, may be biased.<sup>5,12</sup> Several approaches have been proposed to minimize PEs and many involve distinct study designs, including the dual-baseline method approach, the alternative test form approach, or randomizing timing of the first assessment.<sup>13,14</sup> Epidemiological research uses regression-based approaches to correct for PEs typically include an indicator variable for the first assessment.<sup>9,15,16</sup> Due to high correlation between number of prior assessments and time since enrollment, however, these methods may yield imprecise PE estimates.

In addition to PEs, studies of cognitive decline during the COVID-19 pandemic may be subject to mode effects (phone vs. in-person interview) and period effects (differences in scores due to external events or methodologic changes that influence test performance at a specific time).<sup>17</sup> Mode and period effects are important factors to consider in longitudinal cognitive aging research. Recent research has shown that cognitive scores differed by mode of administration and that the association between daily living difficulty and cognitive function differed significantly by mode of administration, suggesting the importance of accounting for mode effects.<sup>18</sup> The pandemic may also introduce period effects, due to adverse effects on social isolation, physical activity, sleep quality, and depression.<sup>20,21</sup> PEs can coexist with period or mode effects, and it may not be possible to simultaneously identify all three without additional, randomization-based assessments.

Simple, feasible approaches are needed to estimate longitudinal cognitive changes using the data immediately available, especially in the context of the pandemic.

The primary goal of this study is to compare and contrast three pragmatic PE approaches in the presence of COVID-19-related period and mode effects. To illustrate, we applied the three approaches to assess the effect of baseline hand grip strength on cognitive trajectory using data from three racially and ethnically diverse cohorts with heterogeneous design features. By separately leveraging data from three cohorts with wide age variations, we were able to evaluate the performance of the three approaches across different age groups.

## Methods

### Samples

We used data from three harmonized prospective cohorts of Kaiser Permanente Northern California (KPNC) members (Figure 1). The Kaiser Healthy Aging and Diverse Life Experiences (KHANDLE) study is comprised of community-dwelling older adults aged 65 years and older residing in the San Francisco Bay and Sacramento Areas of California. The Study of Healthy Aging in African Americans (STAR) cohort includes community-dwelling Black older adults aged 50 years or older in the San Francisco Bay Area of California, primarily in Oakland and Richmond. The LifeAfter90 (LA90) study is an ethnically diverse cohort of individuals aged 90 years and older. The LA90 recruitment is ongoing as long-term KPNC members age into eligibility. To be consistent with the other two studies, we used “wave” to describe the time of the interview in LA90. KHANDLE and STAR participants had completed nearly two waves of in-person data collection before the pandemic began and subsequently completed a 3<sup>rd</sup> wave via phone during the pandemic. At the start of the pandemic, LA90 had collected up to three visits of in-person data; subsequently four visits of data via phone were collected during the pandemic. All three studies obtained approval from the KPNC Institutional Review Board and all participants provided informed consent (see supplement for the inclusion and exclusion criteria of each study).

### Measures

**Cognitive Outcomes**—Executive function and verbal episodic memory were assessed by the Spanish and English Neuropsychological Assessment Scales (SENAS) in all studies.<sup>22,23</sup> The SENAS is a cognitive battery previously validated for comparisons of cognitive change across racial/ethnic and linguistically diverse groups. Executive function scores were obtained using component tasks of category fluency, phonemic (letter) fluency, and working memory (digit-span backward, visual-span backward, list sorting). Verbal episodic memory scores were derived from a multi-trial word-list-learning test. We z-standardized each domain using the baseline mean and standard deviation (SD).

**Grip Strength**—We chose grip strength as the exposure because of consistent evidence linking it to cognition and because it can be objectively measured.<sup>24</sup> Grip strength was assessed in kilograms (kg) at enrollment by trained interviewers, using a Jamar Hydraulic Hand Dynamometer calibrated by B & L Engineering.<sup>24</sup> Three measures were taken from

the dominant hand, with a grip strength score calculated by averaging across measures. Given the variation in hand grip strength between men and women, this score was converted to sex-specific SDs.<sup>24,25</sup>

### Covariates

We considered three variables relevant to age/aging: current age (in years) at each cognitive assessment; age (in years) at enrollment; and years elapsed since enrollment. Current age was grand mean centered by subtracting the mean age of the baseline sample from current age. All models adjusted for sex/gender (men vs. women), and race and ethnicity (categorized as White, Black, Latino, Asian, or other in LA90 and KHANDLE). All STAR participants self-identified as Black. Educational attainment was assessed by asking participants highest degree or last grade in school completed (categorized as high school or lower, some college but no college credential, associate's degree, and at least a bachelor's degree).

### Statistical Analysis

Analyses were performed in each cohort separately. We evaluated the predicted trajectories of cognition using linear mixed-effects models with random intercepts. All approaches modeled cognition for person  $i$  at assessment  $j$ ,  $Y_{ij}$ , with time since enrollment as the timescale and adjustment for age at enrollment, sex, race, education, and interview mode (phone vs. in-person).

We considered three PE approaches. The first ignored the potential for PE by assuming no PE, denoted with the following model:

$$Y_{ij} = \beta_{0i} + \beta_1 \text{time since enrollment}_{ij} + \beta_2 \text{age at enrollment}_i + \beta_3 \text{sex}_i + \beta_4 \text{race}_i + \beta_5 \text{education}_i + \beta_6 \text{interview mode}_{ij} + \varepsilon_{ij} \quad \text{Model 1 (no acknowledgement of PE):}$$

where  $\beta_{0i} = \beta_0 + U_i$  with  $U_i \sim N(0, \sigma_u^2)$ .

The second approach acknowledged the potential for PEs with the inclusion of a binary indicator for the first wave as a fixed effect:

$$Y_{ij} = \beta_{0i} + \beta_1 \text{time since enrollment}_{ij} + \beta_2 \text{age at enrollment}_i + \beta_3 \text{sex}_i + \beta_4 \text{race}_i + \beta_5 \text{education}_i + \beta_6 \text{first-wave}_{ij} + \beta_7 \text{interview mode}_{ij} + \varepsilon_{ij} \quad \text{(Model 2 (first-wave indicator):)}$$

The third approach, the "APM-based constrained PE approach," required two steps: First, we estimated PEs based on a preliminary model using in-person data collected before lockdowns in March 2020. By restricting the sample to *in-person* interviews only, we were able to avoid period effects introduced by the pandemic as well as mode effects that could have arisen from switching from in-person to phone interviews. Specifically, we fit a preliminary linear mixed-effects model using *current age* as the timescale, adjusted for sex, race, education, and wave indicator (i.e., the practice effect):

$$\begin{aligned}
 Y_{ij} = & \alpha_0 + \alpha_1 \text{current age}_{ij} \\
 & + \alpha_2 \text{wave}_{ij} + \alpha_3 \text{sex}_i + \alpha_4 \text{race}_i \\
 & + \alpha_5 \text{education}_i + \varepsilon_{ij}
 \end{aligned}
 \quad \text{Model 3 (step 1 of the APM-based constrained PE approach):}$$

where  $\alpha_{0i} = \alpha_0 + V_i$  with  $V_i \sim N(0, \sigma_v^2)$ .

Conceptually, using current age as the timescale mitigates issues with collinearity between time since enrollment and the wave indicator, in turn allowing separation of the effect of age from PEs. We also tested whether PEs (in these models, the effect of “wave”) differed by covariates by including interactions between covariates (i.e., sex, race, and education) and wave indicators. PEs did not differ by race and education in all three cohorts and by sex in KHANDLE and STAR. Because the interactions between sex and wave indicators were significant in LA90, we included sex by PE interactions for LA90. Throughout, we used balanced data (i.e., only participants with cognitive assessments at all pre-pandemic waves) to minimize potential bias since between-person estimates are highly sensitive to unbalanced data if cognition is associated with attrition.

In the second step, we estimated the predicted cognitive trajectory using data from all waves and incorporated the PE correction as estimated from the preliminary model. In practice, this can be done by either incorporating a PE variable in the model with a constrained coefficient or, equivalently, creating an adjusted outcome,  $Z_{ij}$ , by subtracting the estimated PEs from each individual’s cognitive outcomes after wave 1 (wave 1 was unaltered); that is, we computed  $Z_{ij} = Y_{ij} - \hat{\alpha}_2 \text{wave}_{ij}$ . We then estimated the predicted cognitive trajectory using data from all waves:

$$\begin{aligned}
 Z_{ij} = & \beta_0 + \beta_1 \text{time since} \\
 & \text{enrollment}_{ij} + \beta_2 \text{age at enrollment}_i \\
 & + \beta_3 \text{sex}_i + \beta_4 \text{race}_i + \beta_5 \text{education}_i \\
 & + \beta_6 \text{interview mode}_{ij} + \varepsilon_{ij}
 \end{aligned}
 \quad \text{Model 4 (step 2 of the APM-based constrained PE approach):}$$

Intuitively, using outcome  $Z_{ij}$  serves to constrain PEs across all participants or subgroups of participants, with three assumptions implicitly invoked: 1) PEs do not vary by covariates at enrollment; 2) PEs do not vary by cognitive function at enrollment; and 3) PEs do not differ by interview mode (see Directed Acyclic Graph in Supplemental Figure 1). As mentioned above, we used interaction terms to evaluate whether PEs differed by demographics at enrollment (assumption 1) and in the case when we found evidence that PEs differed by sex, we calculated sex-specific PEs. We fit quantile regression models at the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentiles to test whether PEs differed across levels of cognitive function to evaluate assumption 2. We used the Wald test to test for equality of PEs across percentiles. Because only one wave of data in KHANDLE and STAR was conducted via phone, we could not evaluate assumption 3.

To compare the three approaches, we took advantage of the fact that the rate of cognitive decline can be estimated by either comparing older to younger individuals (between-person estimates) or comparing the same individuals to themselves at different ages (within-person estimates). Between-person age estimates conflate effects of aging with subtle differences

between people born in earlier years and people born in recent years (i.e., cohort effects). On the other hand, within-person age estimates conflate the effects of aging with practice and period or mode effects. In a model correctly accounting for cohort, practice, and period/mode effects, the between-person and within-person age coefficients should be the same.<sup>3,26,27</sup> Therefore, we compared the within-person and between-person age estimates to evaluate approach performance. We also had *a priori* expectations that age-related cognitive change would on average be negative but modest, unlikely to be faster than 0.1 to 0.2 SD per year. Age coefficients outside the range of 0 to -0.2 therefore likely reflect model misspecification.

As a further illustration, we compared estimated effects of grip strength on cognition across the PE approaches. We first added interactions between handgrip strength and wave to the balanced, pre-pandemic data to evaluate whether PEs differed by levels of handgrip strength. We found no evidence that PEs differed by handgrip strength in any of the samples. Then in the whole sample, we ran linear mixed-effects models with time since enrollment as the timescale and adjusted for age at enrollment, sex, race, education, and interactions between time and grip strength.

### Sensitivity Analysis

In our first sensitivity analysis, we repeated the analysis using unbalanced data. Second, to understand whether differences in PEs across cohorts were primarily driven by age differences, we restricted the KHANDLE sample to participants aged 85 years and older (more comparable to LA90) and the STAR sample to participants aged 65 years and older (more comparable to KHANDLE). Third, we repeated the analysis using *current age* as the timescale instead of time since enrollment to understand whether results were sensitive to choice of timescale. Last, we conducted sensitivity analyses using verbal episodic memory as the outcome. We did not include verbal episodic memory from LA90 because, unlike KHANDLE and STAR, participants could see and hear the words during the word list learning component.

All analyses were performed using R version 4.2.1. Operationally, we constrained PEs in Model 4 by adding an “offset” into the linear mixed-effects models (see supplemental texts for R, SAS, and Stata code examples). R code is available at <https://github.com/KHANDLE-STAR-LA90/KHANDLE-practice-effects>.

### Results

Table 1 presents the sample characteristics of KHANDLE ( $N=1,681$ ), STAR ( $N=749$ ), and LA90 ( $N=907$ ). The average age at enrollment was 76 years (SD 6.8) in KHANDLE, 69 years (SD 8.8) in STAR, and 92 years (SD 2.3) in LA90. In KHANDLE, the average interval between waves 1 and 2 was 1.46 years (SD 0.37) and the average interval between waves 2 and 3 was 1.29 years (SD 0.61). In STAR, the average interval between waves 1 and 2 was 2.04 years (SD=0.66) and the average interval between wave 2 and 3 was 1.18 years (SD=0.24). The average interval between waves in LA90 was about 0.5 years.

Figure 2 shows estimated PEs using pre-pandemic, balanced data. In KHANDLE, the estimated PE for wave 2 versus 1 was 0.06 SD (95% CI 0.03, 0.09). In STAR, there was no apparent PE between waves 2 and 1 ( $\beta = -0.002$ , 95% CI  $-0.05, 0.05$ ). In LA90, the estimated wave 2 versus 1 PE was 0.27 SD (95% CI 0.15, 0.39) and the wave 3 versus 1 PE was 0.45 SD (95% CI 0.32, 0.58). PEs were smaller in women than in men in LA90 ( $\beta_{\text{women*wave 2}} = -0.20$  (95% CI  $-0.35, -0.04$ ),  $\beta_{\text{women*wave 3}} = -0.28$  (95% CI  $-0.43, -0.12$ )). There were no significant differences in PEs across the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentiles of executive function scores in KHANDLE ( $p = 0.76$ ), STAR ( $p = 0.76$ ), or LA90 (wave 2  $p = 0.27$ ; wave 3  $p = 0.53$ ) (Supplemental Figure 2). PE estimates were similar when unbalanced data were used (Supplemental Figure 3).

Table 2 presents within- and between-person age estimates using the three approaches using both pre-pandemic and intra-pandemic data. In KHANDLE and LA90, compared with the “no acknowledgment of PE” and the first-wave indicator approach, the APM-based constrained PE approach resulted in the smallest difference between the age-at-enrollment (between-person age effect) and the time-since-enrollment (within-person age effect) estimates (KHANDLE age-at-enrollment:  $-0.05$  (95% CI  $0.06, -0.05$ ) vs. time-since-enrollment:  $-0.02$  (95% CI  $-0.04, -0.004$ ); LA90 women age-at-enrollment:  $-0.08$  (95% CI  $-0.11, -0.05$ ) vs. time-since-enrollment:  $-0.06$  (95% CI  $-0.09, -0.03$ ); LA90 men age-at-enrollment:  $-0.08$  (95% CI  $-0.12, -0.05$ ) vs. time-since-enrollment:  $-0.06$  (95% CI  $-0.10, -0.02$ )). In STAR, the cohort in which PEs were not observed, the “no acknowledgment of PE” approach and the APM-based constrained PE approach resulted in identical estimates for time since enrollment.

The estimated slope of time since enrollment in KHANDLE was positive both in models that did not specify PE and in models including a first-wave indicator (Figure 3), suggesting *improvements* in executive function over time. In models constraining PEs, the time-since-enrollment slope was negative. In STAR, the “no acknowledgement of PE” approach and the APM-based constrained PE approach resulted in identical slopes; the slopes for the first-wave indicator approach were similar. In LA90 women, all three approaches resulted in negative slopes, but the slopes for the first-wave indicator approach and the APM-based constrained PE approach were steeper than the “no acknowledgment of PE” approach. In LA90 men, the slope was positive when a PE was not specified. In contrast, both the first-wave indicator approach and the APM-based constrained PE approach resulted in negative slopes, but the slope for the APM-based constrained PE approach was steeper.

Across the three studies, we found no evidence of an association between grip strength and rate of decline in executive function (Figure 4), with similar findings across the PE approaches and timescales used (see similar results using age as the timescale in Supplemental Figure 4). The findings were generally consistent for verbal memory (Supplemental Figure 5-8).

## Discussion

The COVID-19 pandemic has introduced potential period and mode effects into cognitive aging research. As time, period effects, mode effects, and PEs are highly correlated,

conventional methods to account for PEs may yield imprecise or biased results. Using data from three prospective, racially and ethnically diverse cohorts, we compared estimated cognitive trajectories and associations between grip strength and cognitive decline using three PE approaches. We checked the estimated cognitive decline with each PE approach against three criteria that would imply model misspecification: positive slopes; extremely large negative slopes; and large discrepancies between estimated rates of cognitive decline based on between-person and within-person age effect comparisons. The APM-based constrained PE met all three criteria in all three cohorts, while the “no acknowledgment of PE” approach and the “first-wave indicator” approach resulted in positive slopes or large discrepancies in between- and within-person age estimates in one or more cohorts.

We found substantial variation in the magnitude of PEs across cohorts, suggesting no simple one-size-fits-all solution to modeling PEs is appropriate. Variations in test-retest intervals may account for differences in PE magnitudes across the three studies. LA90 participants were, on average, much older than KHANDLE and STAR participants, and we found the largest PEs in LA90, but almost no PEs in STAR. By restricting KHANDLE and STAR to common age groups, we did not find evidence that differences in PEs by cohort were due to age differences. As LA90 had the shortest test-retest interval among the three cohorts (6 months in LA90 vs. about 1.5 years in KHANDLE and STAR), these findings suggest that test-retest intervals may be more important than age in determining PEs. Furthermore, although PEs did not differ by covariates in KHANDLE and STAR, they were higher in men than in women in LA90. Together, these findings suggest that PEs behave differently across groups; researchers should carefully evaluate within-sample PEs before deciding on the best PE model specification. However, in many cases, it is not possible to evaluate all possible PE specifications, and simplifying assumptions are necessary.

The APM-based constrained PE approach resulted in smaller discrepancies between the within-person (i.e., time since enrollment) and between-person (i.e., age at enrollment) age effects, compared with the “no acknowledgment of PE” approach and the first-wave indicator approach. Because the discrepancy between the between-person and the within-person age effects is partially attributable to PEs, when PEs are properly accounted for, the within-person age trends should closely resemble between-person age trends.<sup>26</sup> The APM-based constrained PE approach uses between-person age coefficients to estimate practice effects at the beginning of the study, when PEs are likely to be larger than age-related decline. Using age at assessment in balanced data to estimate PEs, we can then correct longitudinal data for PEs and model predictions of cognitive changes using within-person estimates. Additionally, by constraining PEs using the estimates derived from balanced, pre-pandemic data, we were able to isolate potential mode and period effects. Although other methods can be used to account for PEs, the APM-based constrained PE approach is relatively easy to implement and does not require changes in study design. This approach is flexible and can incorporate additional data that may be available to estimate PEs, such as randomized substudies.

Our estimates of the association between grip strength and cognitive decline using alternative PE specifications were consistent with prior research.<sup>9</sup> The extent to which PE specifications matter depends on the question being examined.<sup>9</sup> Whenever estimating



cognitive trajectories is the primary goal of the research, PEs must be correctly specified. On the other hand, we did not find substantial differences in PEs across quantiles of cognition, suggesting that when the main goal of research is to examine the effects of a given exposure on levels of or changes in cognition, findings may be less sensitive to alternative PE specifications. We did not find evidence that PEs differed by grip strength, so it is not surprising that the findings for grip strength were not sensitive to PE specification. For exposures that modify PEs, the APM-based approach can conveniently incorporate such effect modification.

This study has several limitations. First, by assigning the constraints estimated from in-person-interview data to phone-interview data, we assumed PEs did not differ by interview mode. Unfortunately, not enough intra-pandemic data are currently available to test this assumption or to evaluate the mode and period effects on cognitive trajectories in our samples. Second, participants were KPNC members with higher socioeconomic status than the general California population.<sup>29</sup> We found no difference in PEs by covariates except sex in LA90; however, these findings may not hold for different populations. Although we did not find evidence that PEs differed by levels of cognition, all samples excluded those with diagnosed dementia at enrollment. Thus, it is unclear whether this holds for a wider range of cognitive performances. Major strengths of this study include evaluation of three distinct cohorts with diverse study participants from a wide age variation, heterogeneous designs, but fully harmonized measures.

## Conclusion

This study addresses an important gap in the literature by presenting the APM-based constrained PE approach, which accounts for PEs in the presence of period and mode effects. This approach can be applied widely in longitudinal cognitive aging research and is especially helpful with data collected during the COVID-19 pandemic, where period/mode effects and PEs must be accounted for in the same model.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the staff and participants of the KHANDLE, STAR, and LA90 studies for their important contributions.

## Funding Sources

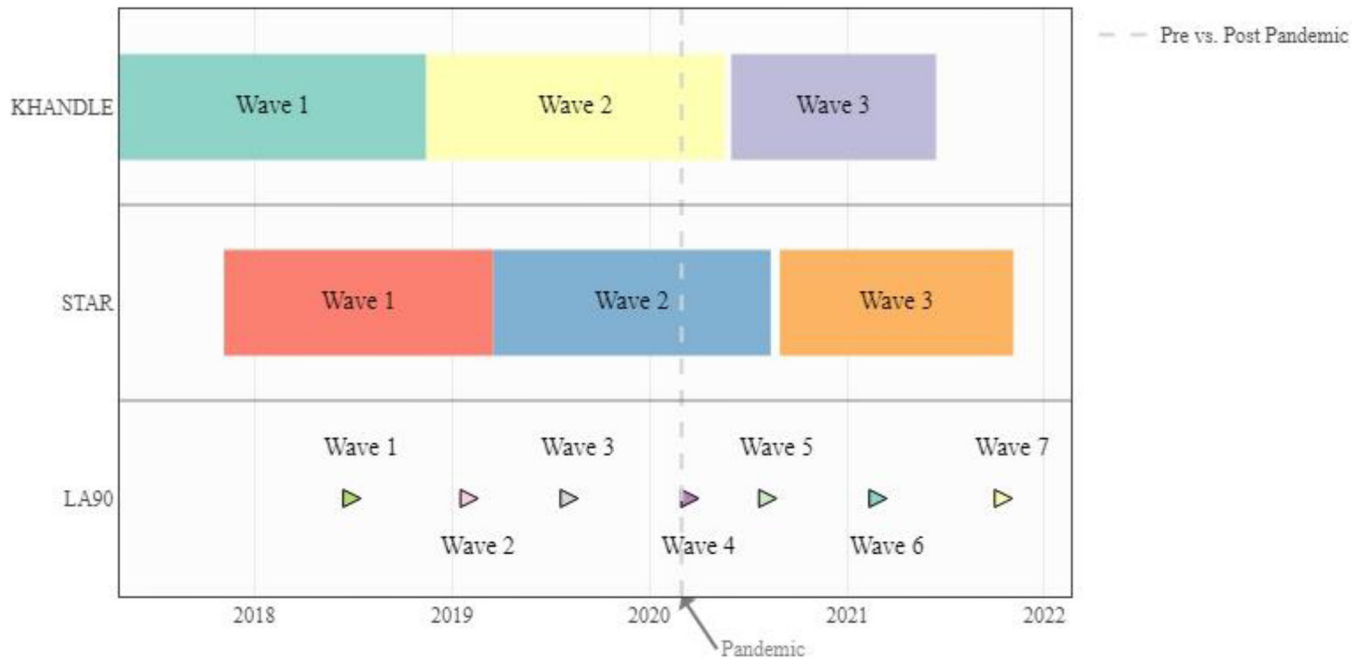
This study is supported by National Institute on Aging (NIA) R01AG052132, RF1AG050782, R01AG05651. RC is supported in part by NIA (K00AG068431). RLP is supported by NIA (R00AG073457-02). EHL is supported by NIA (K99AG075317). SA is supported by NIA (K99AG073454). PG and YS are supported by the Alzheimer's Association/The Judy Fund 2019-AARGD-644788 and NIA (R01AG066132).

## References

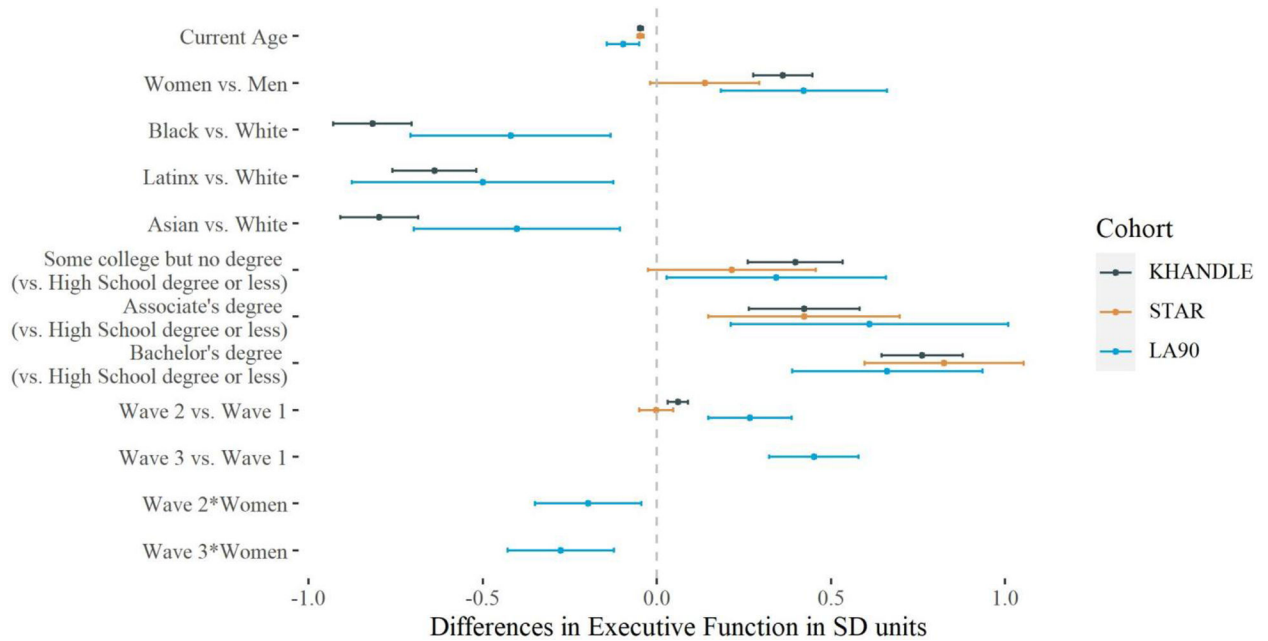
1. Weuve J, Prout-Lima C, Power MC, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimer's & dementia*. 2015;11(9):1098–1109.

2. McCaffrey RJ, Duff K, Westervelt HJ. *Practitioner's Guide to Evaluating Change with Neuropsychological Assessment Instruments*. Springer Science & Business Media; 2000.
3. Salthouse TA. Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*. 2010;24(5):563. [PubMed: 20804244]
4. Gross AL, Benitez A, Shih R, et al. Predictors of retest effects in a longitudinal study of cognitive aging in a diverse community-based sample. *Journal of the International Neuropsychological Society*. 2015;21(7):506–518. [PubMed: 26527240]
5. Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*. 2012;26(4):543–570. [PubMed: 22540222]
6. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience*. 2010;11(1):1–12.
7. Gross AL, Chu N, Anderson L, Glymour MM, Jones RN, Diseases CAM. Do people with Alzheimer's disease improve with repeated testing? Unpacking the role of content and context in retest effects. *Age and ageing*. 2018;47(6):866–871. [PubMed: 30124777]
8. Sanderson-Cimino M, Elman JA, Tu XM, et al. Cognitive practice effects delay diagnosis of MCI: Implications for clinical trials. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2022;8(1):e12228.
9. Vivot A, Power MC, Glymour MM, et al. Jump, hop, or skip: modeling practice effects in studies of determinants of cognitive change in older adults. *American journal of epidemiology*. 2016;183(4):302–314. [PubMed: 26825924]
10. Hale JM, Schneider DC, Gampe J, Mehta NK, Myrskylä M. Trends in the risk of cognitive impairment in the United States, 1996–2014. *Epidemiology (Cambridge, Mass)*. 2020;31(5):745. [PubMed: 32740472]
11. Kremen WS, Sanderson-Cimino ME, Elman JA, et al. Accounting for cognitive practice effects results in earlier detection and more accurate diagnosis of MCI: Biomarker confirmation: Neuropsychology: Longitudinal cognitive assessment in early stages of AD. *Alzheimer's & Dementia*. 2020;16:e044883.
12. Goldberg TE, Goldman RS, Burdick KE, et al. Cognitive improvement after treatment with second-generation antipsychotic medications in first-episode schizophrenia: is it a practice effect? *Archives of general psychiatry*. 2007;64(10):1115–1122. [PubMed: 17909123]
13. Schmitt FA, Bigley JW, McKinnis R, et al. Neuropsychological outcome of zidovudine (AZT) treatment of patients with AIDS and AIDS-related complex. *New England Journal of Medicine*. 1988;319(24):1573–1578. [PubMed: 3059187]
14. Beglinger LJ, Gaydos B, Tangphao-Daniels O, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*. 2005;20(4):517–529. [PubMed: 15896564]
15. Hyun J, Katz MJ, Lipton RB, Sliwinski MJ. Mentally challenging occupations are associated with more rapid cognitive decline at later stages of cognitive aging. *The Journals of Gerontology: Series B*. 2021;76(4):671–680.
16. Fisher GG, Stachowski A, Infurna FJ, Faul JD, Grosch J, Tetrack LE. Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of occupational health psychology*. 2014;19(2):231. [PubMed: 24635733]
17. Schaie KW. *Developmental Influences on Adult Intelligence: The Seattle Longitudinal Study*. Oxford University Press; 2005.
18. Smith JR, Gibbons LE, Crane PK, et al. Shifting of cognitive assessments between face-to-face and telephone administration: Measurement considerations. *The Journals of Gerontology: Series B*. Published online 2022.
19. Daroische R, Hemminghyth MS, Eilertsen TH, Breivte MH, Chwiszczuk LJ. Cognitive impairment after COVID-19—a review on objective test data. *Frontiers in Neurology*. 2021;12:699582. [PubMed: 34393978]
20. Martínez-de-Quel Ó, Suárez-Iglesias D, López-Flores M, Pérez CA. Physical activity, dietary habits and sleep quality before and during COVID-19 lockdown: A longitudinal study. *Appetite*. 2021;158:105019. [PubMed: 33161046]

21. Sepúlveda-Loyola W, Rodríguez-Sánchez I, Pérez-Rodríguez P, et al. Impact of social isolation due to COVID-19 on health in older people: mental and physical effects and recommendations. *The journal of nutrition, health & aging*. 2020;24:938–947.
22. Mungas D, Reed BR, Crane PK, Haan MN, González H. Spanish and English Neuropsychological Assessment Scales (SENAS): further development and psychometric characteristics. *Psychological assessment*. 2004;16(4):347. [PubMed: 15584794]
23. 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. [PubMed: 10791861]
24. George KM, Gilsanz P, Peterson RL, et al. Physical Performance and Cognition in a Diverse Cohort: Kaiser Healthy Aging and Diverse Life Experiences (KHANDLE) Study. *Alzheimer Disease & Associated Disorders*. 2021;35(1):23–29. [PubMed: 33629977]
25. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2001;56(3):M146–M157. [PubMed: 11253156]
26. Salthouse TA, Schroeder DH, 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. [PubMed: 15355168]
27. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. John Wiley & Sons; 2012.
28. Kazlauskaitė R, Janssen I, Wilson RS, et al. Is midlife metabolic syndrome associated with cognitive function change? The study of women’s health across the nation. *The Journal of Clinical Endocrinology & Metabolism*. 2020;105(4):e1093–e1105. [PubMed: 32083676]
29. Hayes-Larson E, Mobley TM, Mungas D, et al. Accounting for lack of representation in dementia research: Generalizing KHANDLE study findings on the prevalence of cognitive impairment to the California older population. *Alzheimer’s & Dementia*. Published online 2022.

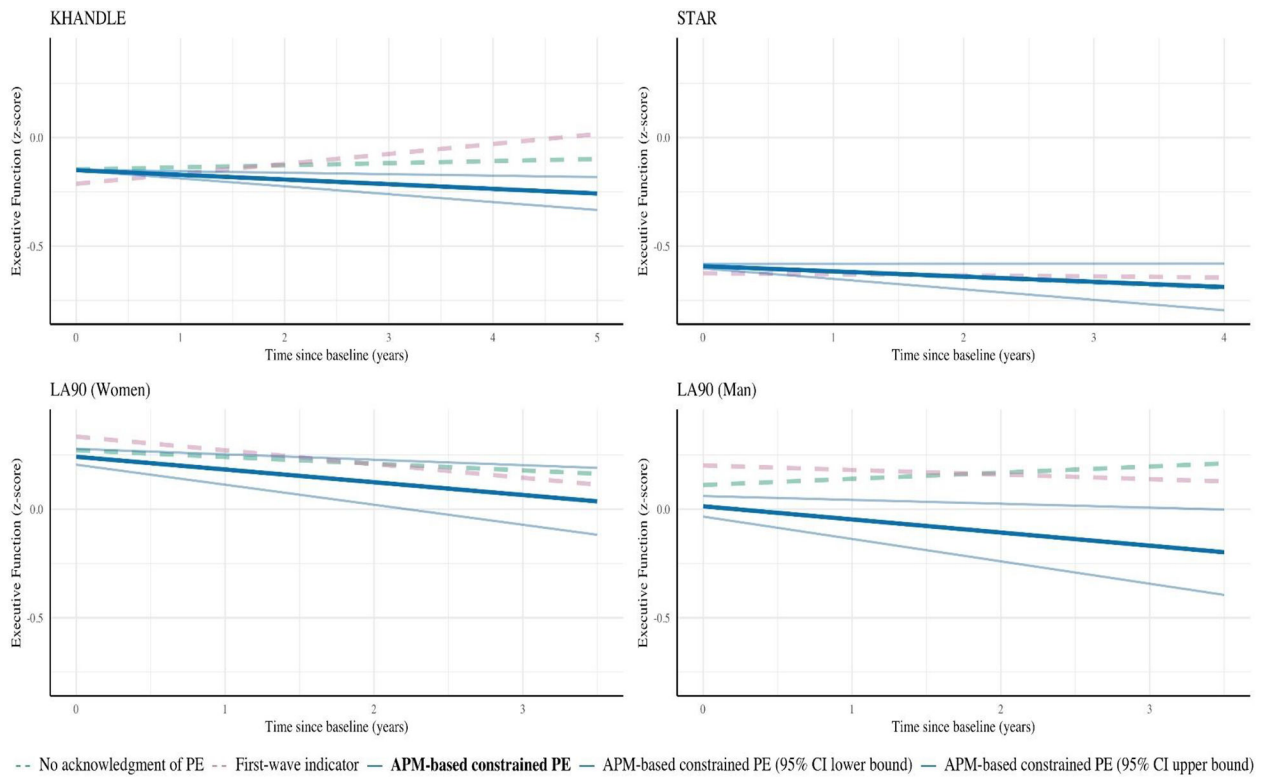


**Figure 1.** KHANDLE, STAR, and LA90 Timelines. LA90 enrollment is ongoing with Kaiser Permanente members who age into eligibility continuously invited so there is no end date for LA90. All interviews were switched from in person to phone after March 2020.



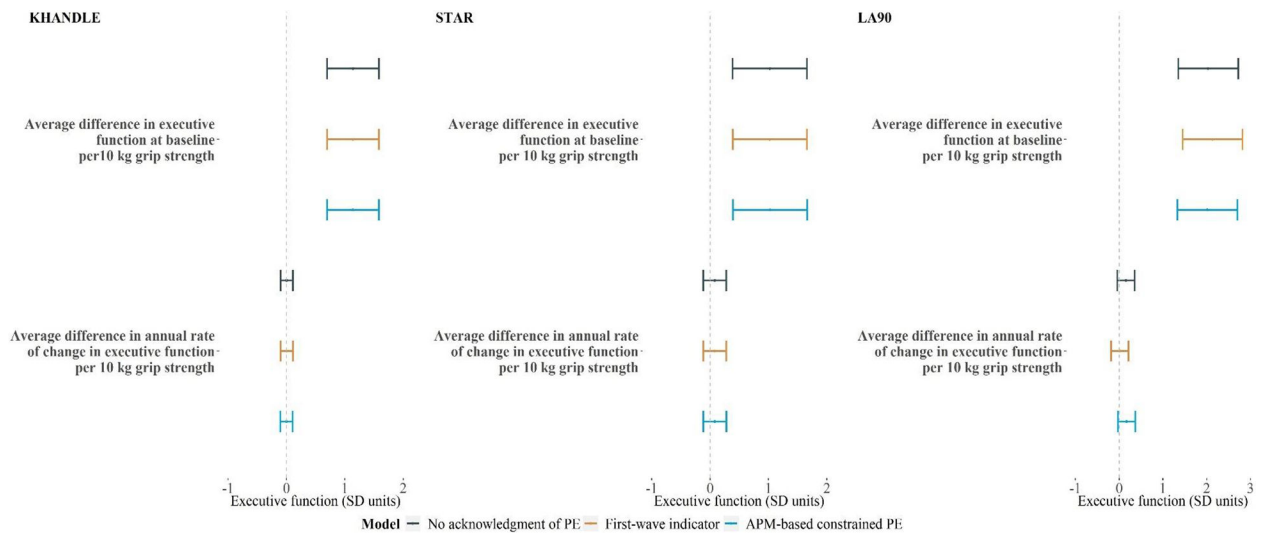
**Figure 2.**

Estimated associations ( $\beta$  and 95% CI) of age, sex, race/ethnicity, education, and indicators for wave with executive function (z-scores) from mixed-effects linear regression models (Model 3 step 1) using KHANDLE ( $n=1333$ ), STAR ( $n=404$ ) and LA90 ( $n=251$ ) data collected before March 2020. All analyses were performed using balanced, pre-pandemic data. All models included random intercepts and used current age as the timescale. Because PE in LA90 was modified by sex, we included interactions between sex and PE in the models. The wave indicators coefficients represent our estimates for the practice effects, so the wave 3 PE could be estimated only for LA90.



**Figure 3.**

Predicted values of executive function model from mixed-effects linear regression models using different approaches to handling PEs in KHANDLE, STAR, and LA90. All models included random intercepts and used time since enrollment as the timescale. All models were adjusted for age at enrollment, sex, race, education, and an indicator for phone assessment. Analyses were performed separately in each cohort. Predictions based on models with no PE specification (green lines) or PE indicators only (pink lines) suggest average increases in cognition in some samples, whereas predictions from models using the APM-based constrained PE approach (blue line) suggest modest annual cognitive declines in all samples.



**Figure 4.**

Estimated associations ( $\beta$  and 95% CI) between grip strength and executive function from mixed-effects linear regression models using different approaches for handling PEs in KHANDLE, STAR, and LA90 ( $n = 595$ , right panel) data. All models included random intercepts only and used time since enrollment as the time scale. All models were adjusted for age at enrollment, sex, race, education, and an indicator for phone assessment. Analyses were performed separately in each cohort. The STAR sample included African Americans only and thus no estimates for Black, Latinx, or Asian participants were shown. Model 1 (green) omits PE, model 2 (orange) adjusts for PE by including an indicator for first assessment for KHANDLE and STAR, and first and second assessments for LA90, and model 3 (blue) forces the PE by constraining the coefficients for the wave indicators based on the preliminary model estimates.

**Table 1.**

Descriptive Statistics for KHANDLE, STAR, and LA90

	<b>KHANDLE</b> N=1,681	<b>STAR</b> N=749	<b>LA90</b> N=907
<b>Age (mean (SD))</b>	75.96 (6.76)	68.63 (8.77)	92.42 (2.32)
<b>Female (%)</b>	998 (59.4)	513 (68.5)	550(60.6)
<b>Race (%)</b>			
Asian	411 (24.4)		233 (25.7)
Black	435 (25.9)	749 (100.0)	224 (24.7)
Latinx	342 (20.3)		190 (20.9)
White	493 (29.3)		260 (28.7)
<b>Education (%)</b>			
High school or lower	353 (21.0)	166 (22.2)	313 (34.5)
Some college but no degree	334 (19.9)	217 (29.0)	182 (20.1)
Associate degree	180 (10.7)	104 (13.9)	88 (9.7)
Bachelor's degree	814 (48.4)	262 (35.0)	324 (35.7)
<b>Longitudinal information (pre-pandemic)</b>			
Duration of follow-up (year), mean (SD)	1.44 (0.35)	1.18 (0.15)	1.1 (0.63)
Average interval (year) between wave 1 and wave 2	1.44 (0.35)	1.18 (0.15)	0.47 (0.2)
Average interval (year) between wave 2 and wave 3	N/A	N/A	0.49 (0.31)
<b>Longitudinal information (Whole study)</b>			
Duration of follow-up (years), mean (SD)	2.58 (0.83)	2.04 (0.66)	1.75 (1.01)
Average interval (years) between wave 1 and wave 2	1.46 (0.37)	1.18 (0.24)	0.6 (0.15)
Average interval (years) between wave 2 and wave 3	1.29 (0.61)	1.2 (0.24)	0.63 (0.23)
Average interval (years) between wave 3 and wave 4	N/A	N/A	0.6 (0.22)
Average interval (years) between wave 4 and wave 5	N/A	N/A	0.55 (0.15)
Average interval (years) between wave 5 and wave 6	N/A	N/A	0.57 (0.18)
Average interval (years) between wave 6 and wave 7	N/A	N/A	0.52 (0.08)

Note. SD = Standard Deviation



Estimates of age at enrollment (within-person age) and time since enrollment (between-person age) in KHANDLE, STAR, and LA90 comparing three approaches to handling Practice Effects

**Table 2.**

	No Acknowledgment of Practice Effect (Model 1)		First-Wave Indicator (Model 2)		APM-Based Constraint on Practice Effect (Model 3)	
	Age at enrollment	Time since enrollment	Age at enrollment	Time since enrollment	Age at enrollment	Time since enrollment
	$\beta$ (95% CI)	$\beta$ (95% CI)	$\beta$ (95% CI)	$\beta$ (95% CI)	$\beta$ (95% CI)	$\beta$ (95% CI)
KHANDLE	-0.05 (-0.06, -0.05)	0.01 (-0.01, 0.03)	-0.05 (0.06, -0.05)	0.05 (0.02, 0.08)	-0.05 (0.06, -0.05)	-0.02 (-0.04, -0.004)
STAR	-0.05 (-0.06, -0.05)	-0.02 (-0.06, 0.01)	-0.05 (0.06, -0.05)	-0.01 (-0.05, 0.05)	-0.05 (0.06, -0.05)	-0.02 (-0.06, 0.01)
LA90 Female	-0.08 (-0.11, -0.05)	-0.03 (-0.06, -0.004)	-0.08 (-0.11, -0.05)	-0.06 (-0.10, -0.03)	-0.08 (-0.11, -0.05)	-0.06 (-0.09, -0.03)
LA90 Male	-0.09 (-0.02, -0.05)	0.03 (-0.01, 0.07)	-0.08 (-0.12, -0.05)	-0.02 (-0.07, 0.03)	-0.08 (-0.12, -0.05)	-0.06 (-0.10, -0.02)

Note. Linear mixed-effects models were adjusted for age at enrollment, sex, race, education, and an indicator for phone assessment. All models included random intercept only and used *time since enrollment* as the timescale.