

Towards Early Detection of Cognitive Decline: Disentangling Retest Effects with a Bayesian Ornstein-Uhlenbeck Process

Grace J. Chen, Riccardo Guidotti, & Zita Oravecz
The Pennsylvania State University
Human Development and Family Studies

September 30, 2026

Outline

1 Background

2 MindTrack Study

3 Methods

4 Results

5 Discussion

Alzheimer's Disease and Related Dementias

- ADRD have a **long preclinical phase** that begins decades before diagnosis (Sperling et al., 2011)
 - We know this largely from **biomarkers**, which begin to change years before symptoms (Jack et al., 2013)
 - Much less evidence comes from **cognitive performance** data
- Once symptoms emerge, neuronal loss is largely irreversible — **prevention and intervention work best early**

Biomarkers

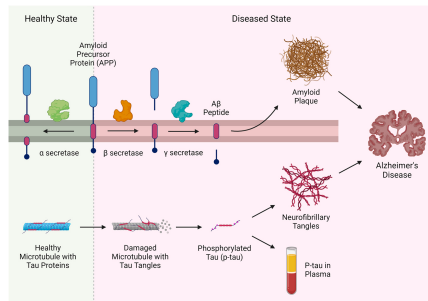
Biomarkers of early-stage AD:

- $A\beta_{42/40}$, pTau217: amyloid and tau
- GFAP: inflammation
- NfL: neuronal damage

How they are collected:

- PET imaging, spinal fluid (CSF), or a blood draw (plasma)

Could be measured **repeatedly**, but **costly** and **not scalable** right now



Lai et al. (2024), *Biomedicines* 12:1836. CC BY 4.0.

Cognitive Performance

- Cognitive tests are more cost-efficient, but usually given as **single-shot assessments**
 - **Limited ecological validity:** a participant may be at their very best that day, or sick, tired, or disengaged — neither reflects daily functioning
 - Cannot tell whether a score is unusual *for that person*
- Alternative: **repeated** testing in people's **natural environments**, now feasible at scale on smartphones (Sliwinski et al., 2018)

So the Question Becomes ...

**Can we extract cognitive markers that reflect early
ADRD risk from repeated assessments?**

Cognitive markers: features of a person's cognitive performance

Outline

1 Background

2 MindTrack Study

3 Methods

4 Results

5 Discussion

The MindTrack Study

Motivation

MindTrack aims to identify **cognitive markers** of ADRD risk in midlife that can help us:

Detect Risk

Find **who is at risk**, as targets for early intervention

Monitor Change

Track **change over time**, e.g., to evaluate whether an intervention is working

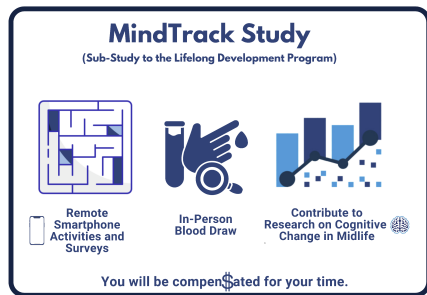
To be useful, these cognitive markers will need to be validated, for example against known AD biomarkers.

The MindTrack Study

Design

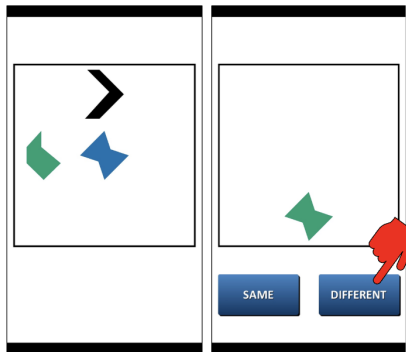
Ongoing longitudinal study of
English-speaking,
community-residing
middle-aged adults

- Repeated **smartphone cognitive assessments** in daily life, in three planned bursts
- **Plasma biomarkers** from an in-person blood draw: $A\beta_{42/40}$, pTau217, GFAP, NfL



Cognitive Assessment

The Color Shapes Task



- Targets **feature binding**, a specific skill where **subtle decline** appears early in preclinical ADRD (Parra et al., 2010)
- **Single-probe** version, optimized for ambulatory assessment (Kim et al., 2025)



demo: original version, not the single-probe variant

Current Analysis

First burst only

Sample ($N = 200$)

- Age: $M = 54.9$, $SD = 8.89$ (range 40–70); 50.5% female; 91.5% White

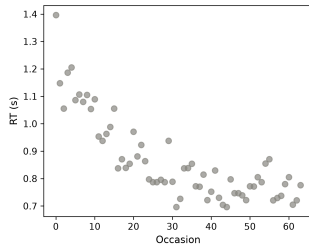
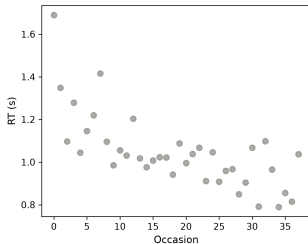
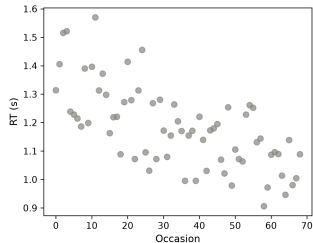
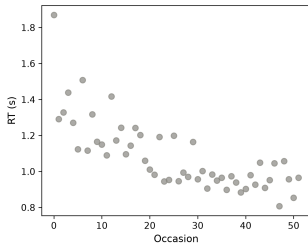
Protocol

- 5 sessions/day for 14 days; 20 trials of the Color Shapes task per session
- Outcome: median response time across the 20 trials in each session

Person-level covariates

- Age, sex, years of education, pTau217

What Do Our Data Look Like?



What Do We See?

- **Repeated measures**, irregularly spaced in time (not shown), with a different number of sessions per person
 - Correlated state: with repeated measurement within each person, we need to consider autocorrelation
- A shared **pattern**: faster with practice, then leveling off
- But people **differ**: where they level off, how much they vary, how much they improve
 - These individual differences give us information about underlying cognitive functioning — cognitive markers!

The Retest Effect

- That pattern is the **retest effect**: performance improves with repeated testing (Calamia et al., 2012)
- Practice gains can **mask** real cognitive decline
- But individual differences related to retest learning are **informative**
 - e.g., smaller practice effects predicted **later progression to AD** in cognitively normal older adults (Hassenstab et al., 2015)

So: model the retest effect rather than treat it as noise

Outline

1 Background

2 MindTrack Study

3 Methods

4 Results

5 Discussion

Modeling Time Series

The Ornstein-Uhlenbeck (OU) Process

Repeated assessments give each person an **irregularly spaced time series**. The OU process, the standard continuous-time model, follows:

$$dY_i(t) = \beta_i(\mu_i - Y_i(t)) dt + \sigma_i dW(t)$$

$\beta_i(\mu_i - Y_i(t)) dt$: pull toward μ_i (deterministic)

$\sigma_i dW(t)$: Brownian motion (stochastic)

Solving the SDE between consecutive sessions gives a Gaussian conditional distribution:

$$Y_{i,s} \mid Y_{i,s-1} \sim \mathcal{N}\left(\mu_i + e^{-\beta_i \Delta_{i,s}} (Y_{i,s-1} - \mu_i), \gamma_i (1 - e^{-2\beta_i \Delta_{i,s}})\right)$$

Note: $\Delta_{i,s} = t_{i,s} - t_{i,s-1}$, elapsed time between sessions, so autocorrelation depends on **actual elapsed time**; $\gamma_i = \sigma_i^2 / 2\beta_i$, the stationary variance. □ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ≡ ▶ ≡ ▶

Modeling the Retest Effect

- Learning theory suggests performance follows an **exponential** curve with practice: improvement slows as it approaches **peak performance** (Heathcote et al., 2000)
 - For response times, peak performance is the **lowest** level (fastest), so the curve **decays**
 - Previously applied to retest effects using **daily averages** (Oravecz et al., 2022)
- Here: the level (μ_i) the OU process pulls toward **follows the exponential curve**
 - Models median response time from **every session**, not daily averages
 - Learning depends on **how many sessions** someone has completed; OU dynamics depend on **how much time** passed between them

Extending the OU Model

A Time-Varying Mean

Replacing μ_i with a time-varying $\mu_{i,s}$ that settles at μ_i :

$$\mu_{i,s} = \mu_i + G_i e^{-R_i s}$$

$$Y_{i,s} \mid Y_{i,s-1} \sim \mathcal{N}\left(\mu_{i,s} + e^{-\beta_i \Delta_{i,s}} (Y_{i,s-1} - \mu_{i,s}), \gamma_i (1 - e^{-2\beta_i \Delta_{i,s}})\right)$$

μ_i	Peak performance	where performance settles after practice
$G_i > 0$	Retest gain	total improvement from practice
$R_i > 0$	Learning rate	how quickly that improvement happens
$\beta_i > 0$	Attractor force	decay rate of autocorrelation $e^{-\beta_i \Delta_{i,s}}$
$\gamma_i > 0$	Volatility	intra-individual variability around $\mu_{i,s}$

s indexes **sessions** completed; $\mu_{i,s}$ is assumed constant between sessions. Initial condition $Y_{i,1} \sim \mathcal{N}(\mu_{i,1}, \gamma_i)$.

The OU Process with Negative Exponential Mean

An Illustration

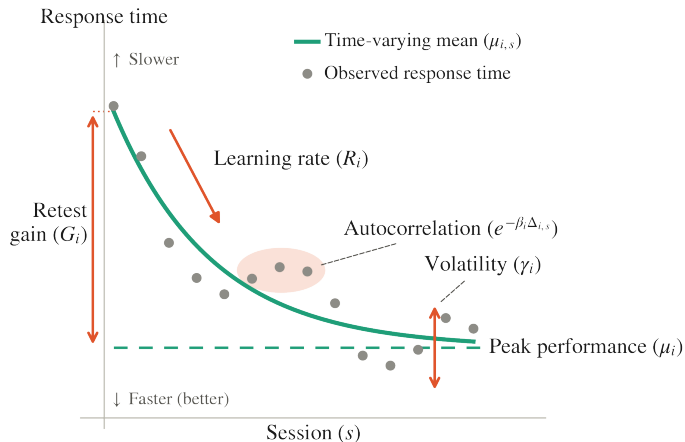


Illustration courtesy of Claude. Hypothetical participant; not fitted data.

The Full Model

Making It Multilevel

Each person gets their own μ_i , β_i , γ_i , G_i , R_i , drawn from population distributions whose means depend on covariates $\mathbf{x}_i$:

$$\mu_i \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\alpha}_\mu, \sigma_\mu^2)$$

$$\log \beta_i \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\alpha}_{\log \beta}, \sigma_{\log \beta}^2)$$

$$\log \sqrt{\gamma_i} \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\alpha}_{\log \sqrt{\gamma}}, \sigma_{\log \sqrt{\gamma}}^2)$$

$$\log G_i \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\alpha}_{\log G}, \sigma_{\log G}^2)$$

$$\log R_i \sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\alpha}_{\log R}, \sigma_{\log R}^2)$$

Note: $\beta_i, \gamma_i, G_i, R_i > 0$ by definition; modeled on the log scale. Autocorrelation at 1-hour interval is bounded below at .01, so $\log(\beta_i) \leq 1.527$, $\beta_i \leq 4.61$. Covariates: age, sex, years of education, pTau217. All but sex standardized.

Model Applied

- **Framework:** MCMC via PyMC; all parameters estimated simultaneously
- **Priors:**
 - Regression coefficients: $\alpha. \sim \mathcal{N}(0, 10^2)$
 - Between-person SDs: $\sigma_{\log \sqrt{\gamma}}, \sigma_{\log G}, \sigma_{\log R} \sim \text{Half-}\mathcal{N}(0, 1)$
 - Between-person variances: $\sigma_{\mu}^2, \sigma_{\log \beta}^2 \sim \text{Gamma}(0.01, 0.01)$
- **Sampling:** NUTS via nutpie; 4 chains $\times$ 25,000 draws, 2,000 tuning steps each

Outline

1 Background

2 MindTrack Study

3 Methods

4 Results

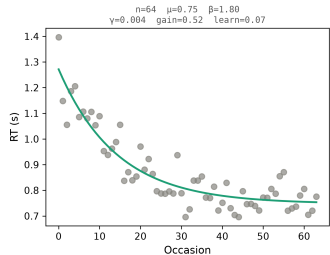
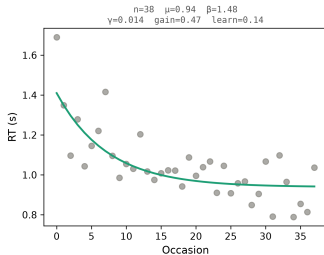
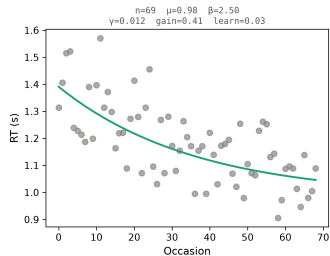
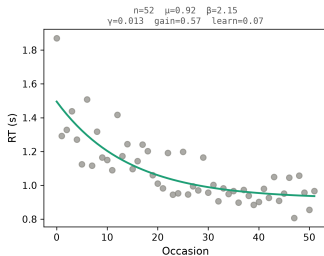
5 Discussion

MCMC Diagnostics

Diagnostic	Value	Criterion
Divergent transitions	0 / ($4 \times 25,000$)	0%
Iterations at max tree depth	0 / ($4 \times 25,000$)	0%
Maximum $\hat{R}$	1.006	< 1.01
Minimum ESS	925	> 400

Results

Remember those examples from earlier?



Results

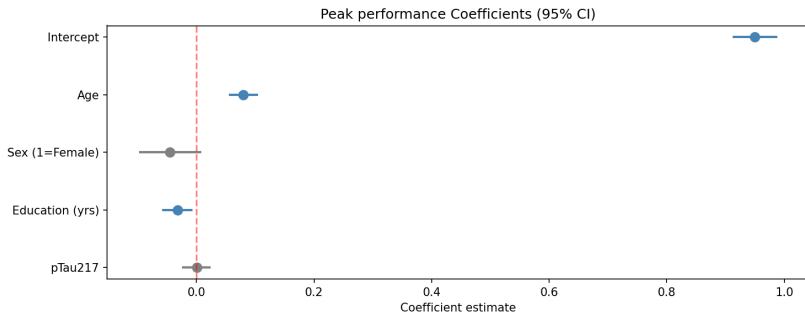
Population-Level Estimates

		Posterior	95% CI	
		Median	Low	High
Peak performance (μ)	Intercept	0.950	0.912	0.988
	SD	0.145	0.122	0.170
Attractor force (β)	Intercept	3.397	1.674	4.547
	SD	1.127	0.943	1.215
Volatility ($\sqrt{\gamma}$)	Intercept	0.121	0.110	0.132
	SD	0.055	0.047	0.067
Retest gain (G)	Intercept	0.314	0.276	0.352
	SD	0.150	0.121	0.193
Learning rate (R)	Intercept	0.059	0.041	0.081
	SD	0.166	0.087	0.340

Note: Original scale (μ , $\sqrt{\gamma}$, G in seconds; β per hour; R per session), for a person at the covariate reference values (mean age, years of education, and pTau217; male). SD = between-person SD after covariates.

Results

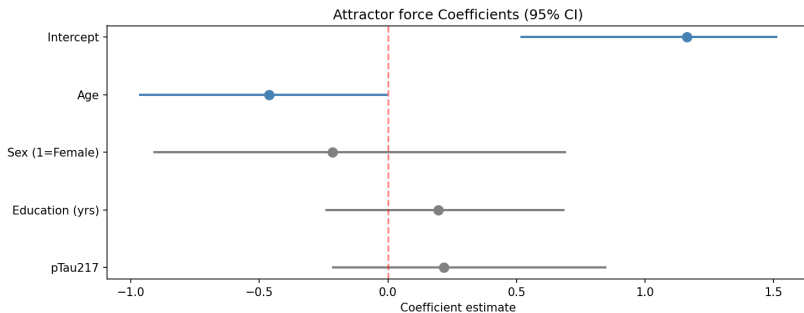
Coefficients by Parameter



Note: Difference in peak performance (seconds) per unit of each predictor (0 = no effect).

Results

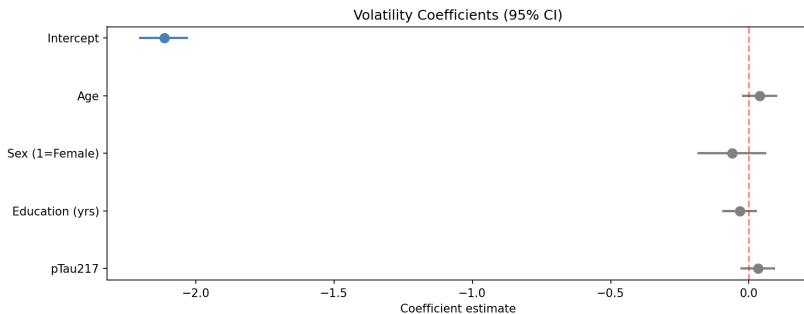
Coefficients by Parameter



Note: Change in the *log* of the parameter per unit of each predictor (0 = no effect).

Results

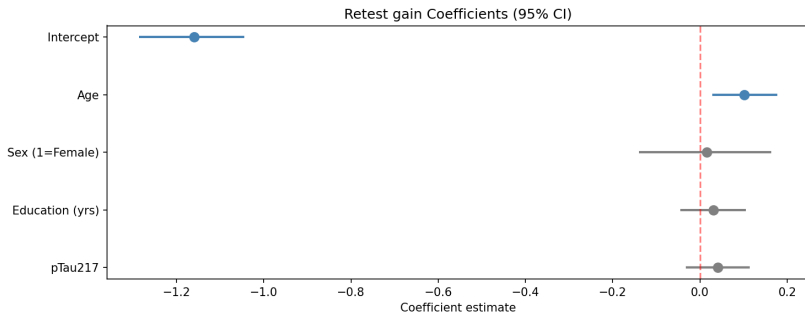
Coefficients by Parameter



Note: Change in the *log* of the parameter per unit of each predictor (0 = no effect).

Results

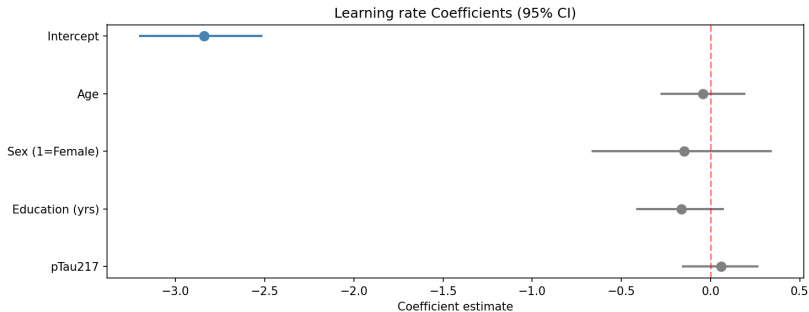
Coefficients by Parameter



Note: Change in the *log* of the parameter per unit of each predictor (0 = no effect).

Results

Coefficients by Parameter



Note: Change in the *log* of the parameter per unit of each predictor (0 = no effect).

Results

Credible Predictors of Cognitive and Retest Features

Model parameter	Person covariate	Posterior mean	<i>pd</i>
Peak performance (μ)	Age	0.080	>99.9%
	Education	-0.032	99.2%
Attractor force (β)	Age	-0.461	97.5%
Retest gain (G)	Age	0.102	99.6%

Note: Coefficients on the model scale: peak performance in seconds; β , $\sqrt{\gamma}$, G , and R on the log scale. Credible predictors: 95% credible interval excludes zero. *pd* = probability of direction. Age, years of education, and pTau217 are standardized. No predictor was credible for volatility or learning rate, and no pTau217 effect was (maximum *pd* = 86.3%).

Outline

1 Background

2 MindTrack Study

3 Methods

4 Results

5 Discussion

Key Takeaways

- Individual-specific features of retest learning are not a nuisance but **important cognitive markers** with **potential clinical utility**.
- Time-varying mean following a negative exponential implemented in an OU framework allows us to **separate the retest effect out from cognitive performance**.
 - All in one step!
- Empirically: older adults have a **larger retest gain** but reach a **slower peak performance**
 - More room to improve, yet they still don't catch up to younger adults

Limitations

- **Attractor force** (β) sits near its upper bound: little carryover detectable at hourly spacing
 - Sessions may be further apart than the timescale of carryover
- **Single burst**: cannot yet estimate across-burst change (e.g., forgetting) or relate *change* in biomarkers to *change* in cognition
- **pTau217** is upstream of cognition, and values skew low in a healthy mid-age sample
 - Its link to cognitive markers may lie in *changes over time* rather than in a cross-sectional snapshot

Future Directions

- **Multi-burst extension:** across-burst parameters to estimate change in latent cognitive ability (Oravecz et al., 2022)
- **Hypothesized markers** of early AD-related change:
 - **Forgetting** between bursts
 - Change in **peak performance** between bursts
 - Possibly, change in **autocorrelation** and **variability** over time
- **Validating** these markers against biomarkers

That's a wrap!

Thank you!

Questions?

gjchen@psu.edu

A very special thank you to Riccardo Guidotti and Zita Oravecz!

Template courtesy of Jonathan Park & Lindy Williams.