Research Article

Naturalistic assessment of reaction time variability in older adults at risk for Alzheimer's disease

Matthew S. Welhaf[1] ⓘ, Hannah Wilks[1], Andrew J. Aschenbrenner[2] ⓘ, David A. Balota[1], Suzanne E. Schindler[2], Tammie L.S. Benzinger[3], Brian A. Gordon[1,3], Carlos Cruchaga[4], Chengjie Xiong[5], John C. Morris[2] and Jason Hassenstab[1,2] ⓘ

[1]Department of Psychological & Brain Sciences, Washington University in St. Louis. St. Louis, MO, USA, [2]Department of Neurology. Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, MO, USA, [3]Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA, [4]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA and [5]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA

Abstract

Objective: Maintaining attention underlies many aspects of cognition and becomes compromised early in neurodegenerative diseases like Alzheimer's disease (AD). The consistency of maintaining attention can be measured with reaction time (RT) variability. Previous work has focused on measuring such fluctuations during in-clinic testing, but recent developments in remote, smartphone-based cognitive assessments can allow one to test if these fluctuations in attention are evident in naturalistic settings and if they are sensitive to traditional clinical and cognitive markers of AD. **Method:** Three hundred and seventy older adults (aged 75.8 +/− 5.8 years) completed a week of remote daily testing on the Ambulatory Research in Cognition (ARC) smartphone platform and also completed clinical, genetic, and conventional in-clinic cognitive assessments. RT variability was assessed in a brief (20-40 seconds) processing speed task using two different measures of variability, the Coefficient of Variation (CoV) and the Root Mean Squared Successive Difference (RMSSD) of RTs on correct trials. **Results:** Symptomatic participants showed greater variability compared to cognitively normal participants. When restricted to cognitively normal participants, APOE ε4 carriers exhibited greater variability than noncarriers. Both CoV and RMSSD showed significant, and similar, correlations with several in-clinic cognitive composites. Finally, both RT variability measures significantly mediated the relationship between APOE ε4 status and several in-clinic cognition composites. **Conclusions:** Attentional fluctuations over 20–40 seconds assessed in daily life, are sensitive to clinical status and genetic risk for AD. RT variability appears to be an important predictor of cognitive deficits during the preclinical disease stage.

Keywords: attention consistency; mobile testing; digital biomarkers; apolipoprotein e4; preclinical alzheimer disease

(Received 1 August 2023; final revision 28 November 2023; accepted 30 November 2023; First Published online 29 January 2024)

Introduction

Cognitive decline is a hallmark of Alzheimer's disease (AD), and capturing cognitive changes as early as possible is critical for potential therapeutic intervention. Conventional assessment approaches rely on classic cognitive measures that were designed to detect more overt impairment and may not be sensitive to subtle changes that occur prior to onset of clinical symptoms. One promising approach uses data from high-frequency cognitive assessments to examine variability in performance at different timescales.

The study of cognitive variability over larger timescales (i.e., weeks, months, and years) in healthy aging and AD has revealed that older adults and those at risk for developing AD or already

showing AD symptoms are more inconsistent across testing sessions (e.g., Bielak et al., 2019; Cerino et al., 2021; Hultsch et al., 2000; MacDonald & Stawski, 2020; Martin & Hofer, 2004; Stawski et al., 2015). However, moment-to-moment variability (over the course of several seconds, rather than hours) might serve an important role in basic psychological and everyday functional processes (Unsworth & Miller, 2021; Unsworth, 2015; Welhaf & Kane, 2023). For example, the ability to maintain consistent attention and focus over the course of several seconds is important for many everyday functions like remembering a shopping list, driving, following basic instructions, and even focusing during a conversation. Failing to maintain consistent attention can result in a host of errors ranging from simple forgetting to automobile accidents. In the case of cognitive testing, consistency of reaction

times (RT) might provide unique information about an individual's attentional ability. Throughout a task, participants' attention may naturally wax and wane, leading to periods of focused, accurate performance, and periods of worse performance (Esterman et al., 2013). This information is overlooked by only examining mean performance on a task.

Using measures of intra-individual RT variability within a single task has previously been useful in discriminating healthy aging from mild cognitive impairment and from individuals diagnosed with symptomatic AD (Christ et al., 2018; Dixon et al., 2007; Gorus et al., 2008; Hultsch et al., 2000) and even predicting increased risk of developing mild cognitive impairment (for a review see Haynes et al., 2017). For example, Duchek et al. (2009) found that individuals diagnosed with very mild AD showed increased variability on difficult trials (measured as RT coefficient of variation, [CoV]) compared to healthy older adults and young adults, in three attention control tasks. Also, Gyurkovics et al. (2018) found increased CoV in a go/no-go task in early-stage AD participants compared to healthy older adults and healthy young adults. Increased RT variability has also been proposed to, at least partially, reflect efficiency of attention control (West et al., 2002), differences in white matter integrity (Jackson et al., 2012), or breakdowns in functional connectivity (Duchek et al., 2013). If RT variability reflects efficiency of attention control (e.g., West et al., 2002) then it should also be related to conventional cognitive measures which rely on consistent attention for successful performance. Previous work suggests that attention control appears to be impaired in the early stages of AD and might be an important precursor to memory-related impairments associated with AD (see Balota & Faust, 2001; Perry & Hodges, 1999).

This previous work has examined the sensitivity of RT variability between participants who are healthy and beginning to show early signs of AD. However, a major effort of AD research is to identify markers that are sensitive during the preclinical stage of the disease, especially in individual who might be at an increased risk for developing AD. For example, carrying at least one *APOE* ε4 allele is a major risk factor for developing AD (Corder et al., 1993) and is often associated with increased amyloid burden (Mishra et al., 2018; Morris et al., 2010). Thus, it is possible that RT variability might be a sensitive measure to help identify cognitive decline earlier among *APOE* ε4 carriers. Several studies have tested for such a relationship, but the findings are mixed. Some have found that carrying at least one copy of *APOE* ε4 is associated with increases in RT variability in otherwise healthy older adults (Duchek et al., 2009). Other studies have found no difference in RT variability measures between ε4 carriers and noncarriers (e.g., Anstey et al., 2007; Kay et al., 2017; Lu et al., 2020). One possibility for these differences across studies may be due to the tasks used. Specifically, Duchek et al. (2009) assess RT variability in the context of attention control tasks (e.g., Stroop), whereas studies reporting null effects assessed RT variability in either simple or choice RT tasks (Anstey et al., 2007; Lu et al., 2020) or a motor timing task (Kay et al., 2017). Thus, more work is needed to understand how *APOE* ε4 status is related to RT variability and how results may vary by the cognitive demands of the task.

Prior studies of RT variability have, for pragmatic reasons, all taken place in clinic or lab settings. These typically require participants to travel to the assessment facility and undergo testing for extended periods of time. These "one-shot" assessments of cognition may be biased due to many factors including stress from traveling to the site, or other performance decrements due to day-to-day hassles and contextual confounds like poor mood, or increased fatigue that naturally occur in daily life. Further, when participants come into the clinic, they may enter a "testing mode," which might unintentionally generate different cognitive performance than they would otherwise display in real life. Additionally, traditional in-clinic testing occurred in a structured testing environment that is highly controlled and minimizes distractions. Measuring aspects of cognition, like attention, in everyday life where people are more vulnerable to distractions can be more reflective of real-world functioning compared to traditional in-clinic testing.

Our assessment approach (see Nicosia et al., 2023; Wilks et al., 2021) uses a custom-designed smartphone app to repeatedly assess cognition with a measurement "burst" design. In this type of design, participants are remotely assessed several times per day for several consecutive days with extremely brief cognitive tests that are designed for repeated administration. Participants might complete these assessment bursts annually or in our case, semiannually, to investigate cognitive change (see also Cerino et al., 2021; Sliwinski et al., 2018; Stawski et al., 2019). Using principles of ecological momentary assessment (EMA; Shiffman et al., 2008), we can capture behavior and cognition in daily life across a variety of contexts and environments to better characterize fluctuations in RT.

Overview of the present study

The present paper focuses on how *moment-to-moment variability* in performance (assessed via RT variability) on a brief processing speed task given multiple times over the course of a week might be a sensitive measure to detect subtle cognitive changes in older adults at risk for AD. Below we detail the major aims of the study.

Aim 1: Can RT variability measure be reliably assessed using EMA methods?

First, we examined if two measures of RT variability can be reliably assessed over the course of a week using EMA. Figure 1 depicts our general analytic approach. We examined moment-to-moment variability within each session and then calculated the mean of that variability across sessions to obtain a "person-level" estimate of variability. Thus, each visit can be equated to an individual "task" that might be given at a typical in-clinic assessment. Using this approach, we aimed to minimize influences of between-session variability, which appears to be sensitive to AD risk factors like *APOE* status (see Aschenbrenner et al., 2023).

Aim 2: How is RT variability related to AD risk factors?

If RT variability can be reliably assessed using EMA, the next logical question is to see if it is sensitive to important AD risk factors. In line with previous research (Duchek et al., 2009; Gyurkovics et al., 2018), we hypothesized that RT variability would be greater among those already showing clinical symptoms (e.g., CDR 0.5 vs. CDR 0 participants). Further, we hypothesized that RT variability would also be a sensitive measure in a preclinical sample of participants who were at increased risk for developing AD dementia (e.g., *APOE* ε4 carriers).
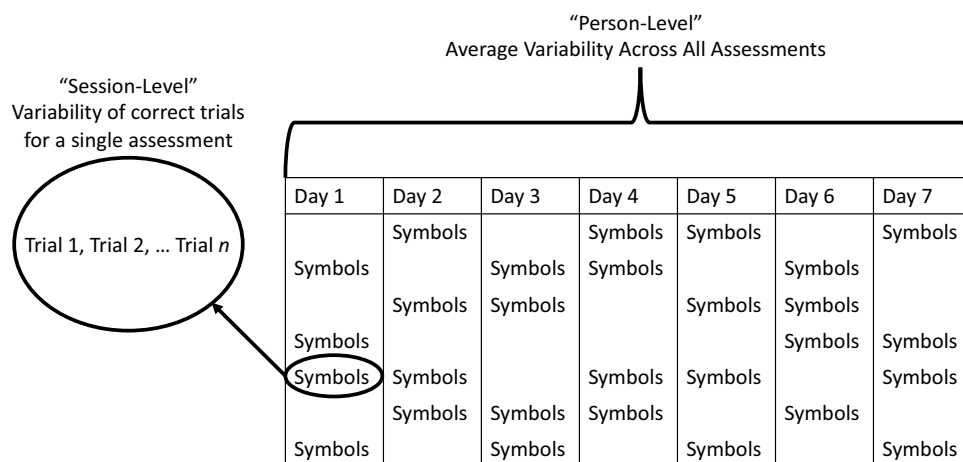
Figure 1. Depiction of different levels of RT variability analyses across a typical ARC assessment visit period.

Aim 3: How is RT variability related to conventional cognitive measures and can it explain the relationship between AD risk and cognitive performance?

We hypothesized that measures of RT variability would be modestly correlated with in-clinic measures of cognition and perhaps more strongly correlated with in-clinic measures of attention control compared to memory or speed measures. Further, we hypothesized that RT variability would be a significant mediator in the relationship between AD risk (indexed by *APOE* ε4 status) and in-clinic cognition. That is, part of the reason *APOE* ε4 carriers might perform worse on traditional cognitive measures is because of their poorer ability to maintain consistent focus (i.e., greater RT variability).

Methods

Participants

Participants were recruited from ongoing studies of aging and dementia at the Knight Alzheimer Disease Research Center (Knight ADRC) at Washington University School of Medicine in St Louis. Data collection began in March of 2020 for our ongoing high-frequency cognitive assessment study of older adults at risk for developing AD, the Ambulatory Research in Cognition (ARC) study (Nicosia et al., 2023) at Washington University in St Louis. Nearly all participants were invited to enroll after completing their annual cognitive and clinic assessment and began ARC testing shortly after enrolling. As data collection is still ongoing, we used a data freeze from June 2023 for the current analyses. All participants provided informed consent, and all procedures were approved by the Human Research Protections Office at Washington University in St Louis and the research was conducted in accordance with the Helsinki Declaration.

The final sample consisted of 377 participants. Table 1 provides demographic and ARC information on the full sample, grouped by CDR status. Table 2 provides demographic information and cognitive assessment descriptive statistics participants in the CDR 0 subsample, grouped by *APOE* ε4 status. Participants were highly educated, most self-identified their race as White, and were majority female. Adherence was high with participants completing, on average, 23.07 ± 4.69 out of a possible 28 sessions (82.39% adherence).

Clinical assessments

Clinical status was assessed using the Clinical Dementia Rating® (Morris, 1993). The CDR® rates cognitive and functional performance on a 5-point scale across six domains (memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care; see Morris, 1993). Participants and a collateral source (a close family member or friend) both underwent semi-structured interviews to determine CDR scores with CDR = 0 indicating cognitive normality, 0.5 indicating very mild dementia, 1 indicating mild dementia, 2 indicating moderate dementia, and 3 indicating severe dementia. For the current study, participants were categorized as either cognitively normal (CDR = 0), or diagnosed with very mild dementia (CDR = 0.5). By design, most participants were cognitively normal (CDR 0), with the final sample consisting of 345 cognitively normal participants and 32 CDR 0.5 participants. CDR assessments were completed 8.02 ± 7.53 weeks from the ARC assessment.

APOE genotyping was conducted using blood samples drawn from participants. Participants were classified as *APOE* ε4 carriers (4/4, 3/4, 2/4) or *APOE* ε4 noncarriers (3/3, 2/3, 2/2) (see Cruchaga et al., 2013). During the start of the ARC study, biomarker assessments at the Knight ADRC were suspended temporarily due to the COVID-19 pandemic. Because of this, we focus our analyses on *APOE* ε4 status as our primary AD risk factor (as in Aschenbrenner et al., 2023).

ARC smartphone assessment

Participants completed three cognitive tasks up to four times per day over a week in their natural environments (see Nicosia et al., 2023 for full details on the ARC platform). For the purposes of the current study, we describe the Symbols test, as this measure features RT as its primary outcome (the other two tasks include a measure of visual working memory and associative memory). A screenshot of the Symbols test is shown in Figure 2. We only included participants if they completed 10 sessions of the Symbols test over the course of the week to ensure they engaged sufficiently with the testing.

During the Symbols task, participants were randomly shown three pairs of abstract shapes and were asked to select which of two possible responses matched one of the three targets. At each assessment, participants completed 12 trials (there was no response deadline for participants in the task). In our previous work

Table 1. Demographics of sample by CDR status

	CDR 0 ($N = 345$)	CDR = 0.5 ($N = 32$)	Statistic	p-value
Age (years)	75.80 (5.77)	77.10 (6.44)	$t(35.87) = -1.10$	.278
Gender (% Female)	199 (58.9%)	12 (37.5%)	$\chi2 = 5.45$	.020
Education (years)	16.50 (2.42)	16.12 (2.35)	$t(37.50) = 0.86$	.395
Race (N)				
White	285	31		
Black/African American	58	1		
Other	2	0		
APOE ε4 status (%)			$\chi2 = 4.81$	.028
ε4 carrier	117 (33.7%)	17 (53.1%)		
ε4 noncarrier	228 (66.3%)	15 (46.9%)		
Sessions Complete	23.07 (4.69)	21.28 (5.28)	$t(35.78) = 1.84$	.074
Symbols CoV	0.33 (0.06)	0.37 (0.09)	$t(34.18) = -2.45$	.020
Symbols RMSSD	1499.68 (818.10)	2040.66 (1100.72)	$t(34.25) = -2.71$	.010

APOE ε4 = Apolipoprotein ε4, CDR = Clinical Dementia Rating.
Means (Standard Deviations) reported for continuous variables. N(%) reported for categorical variables. Gender and Race were self-reported.

Table 2. Demographics and cognitive performance of CDR 0 subsample by APOE ε4 status

	ε4 Noncarrier ($N = 228$)	ε4 Carrier ($N = 117$)	Statistic	p-value
Age (years)	76.12 (6.44)	74.98 (5.37)	$t(255.96) = 1.79$	.075
Gender (% Female)	137 (60.1%)	67 (57.3%)	$\chi2 = 0.043$	.836
Education (years)	16.57 (2.38)	16.41 (2.47)	$t(223.93) = 0.56$	.576
Race (N)				
White	195	91		
Black/African American	34	24		
Other	1	1		
Sessions Complete	26.43 (12.34)	24.86 (8.63)	$t(311.71) = 1.39$	.167
Symbols CoV	0.32 (0.06)	0.34 (0.07)	$t(216.12) = -1.57$	.118
Symbols RMSSD	1445.23 (618.18)	1600.83 (1108.99)	$t(151.99) = -1.41$	.162
FCSRT Free Recall	32.71 (5.61)	31.29 (6.43)	$t(169.57) = 1.85$	.065
WMS associates Recall	15.39 (3.35)	14.43 (3.30)	$t(229.55) = 2.52$	.013
Craft Story Recall Immediate	17.68 (3.54)	16.57 (3.39)	$t(238.30) = 2.83$	.005
Craft Story Recall Delay	17.27 (3.54)	15.84 (3.65)	$t(225.91) = 3.49$	.001
Animal Fluency	20.94 (5.69)	20.80 (5.60)	$t(234.19) = 0.214$	.831
Vegetable Fluency	14.33 (4.03)	13.72 (4.02)	$t(231.94) = 1.31$	.191
Multilingual Naming Test	30.29 (1.93)	30.14 (1.74)	$t(212.81) = 0.65$	.519
Number Symbol	39.36 (7.55)	39.40 (9.02)	$t(164.51) = -0.04$	.966
Digit Span Forward	8.35 (2.41)	8.62 (2.19)	$t(253.08) = -1.05$	.293
CVOE Switch Accuracy	0.95 (0.09)	0.95 (0.10)	$t(199.77) = 0.15$	.879
Stroop Incongruent Accuracy	0.94 (0.14)	0.95 (0.07)	$t(324.00) = -1.19$	.235
Episodic Memory Composite	0.10 (0.67)	-0.20 (0.74)	$t(232.57) = 3.50$	.001
Semantic Memory Composite	0.03 (0.86)	-0.05 (0.83)	$t(244.91) = 1.00$	.317
Processing Speed Composite	-0.02 (0.82)	0.03 (0.65)	$t(216.45) = -0.72$	.574
Attention Control Composite	-0.02 (0.70)	0.04 (0.76)	$t(261.80) = -0.56$	.577

APOE ε4 = Apolipoprotein ε4; In-clinic cognitive measures were only analyzed for the CDR 0 group.
Means (Standard Deviations) reported for continuous variables. N(%) reported for categorical variables. Gender and Race were self-reported.

(Nicosia et al., 2022; Wilks et al., 2021) the primary dependent measure was median RT on correct trials. For the current study, because we were interested in the variability in RTs, we calculated CoV (standard deviation of RTs / mean RT) to correct RTs.

We chose CoV for multiple reasons. First, mean RT and RT variability are highly correlated, especially in older adults (Faust et al., 1999), so we used CoV to account for age- and disease-related slowing of RTs. Second, we used CoV to maintain consistency with our past work (e.g., Duchek et al., 2009). Higher scores indicated greater variability and thus poorer moment-to-moment attention above and beyond overall differences in processing speed.

To test the robustness of our methods, we also calculated an alternative measure of RT variability with the Root Mean Squared Successive Difference (RMSSD; Von Neuman et al., 1941). The equation for RMSSD is shown in Equation (1):

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (X_i - X_{i+1})^2}{N - 1}}$$

where X_i is the interval between adjacent trials, X_{i+1} the next trial interval, and N is the number of trials.

We calculated participant-level estimates of RMSSD the same way we did for CoV. That is, we calculated the RMSSD at each individual session and then averaged this over all sessions for a given participant. RMSSD is a well-established measure of variability in the heart rate literature and accounts for gradual shifts in mean performance over time and overall learning effects. However, because CoV can behave differently than other variability measures (e.g., Hultsch et al., 2011; Stawski et al., 2019), it is important to determine if measures of variability might lead to different results or interpretations.
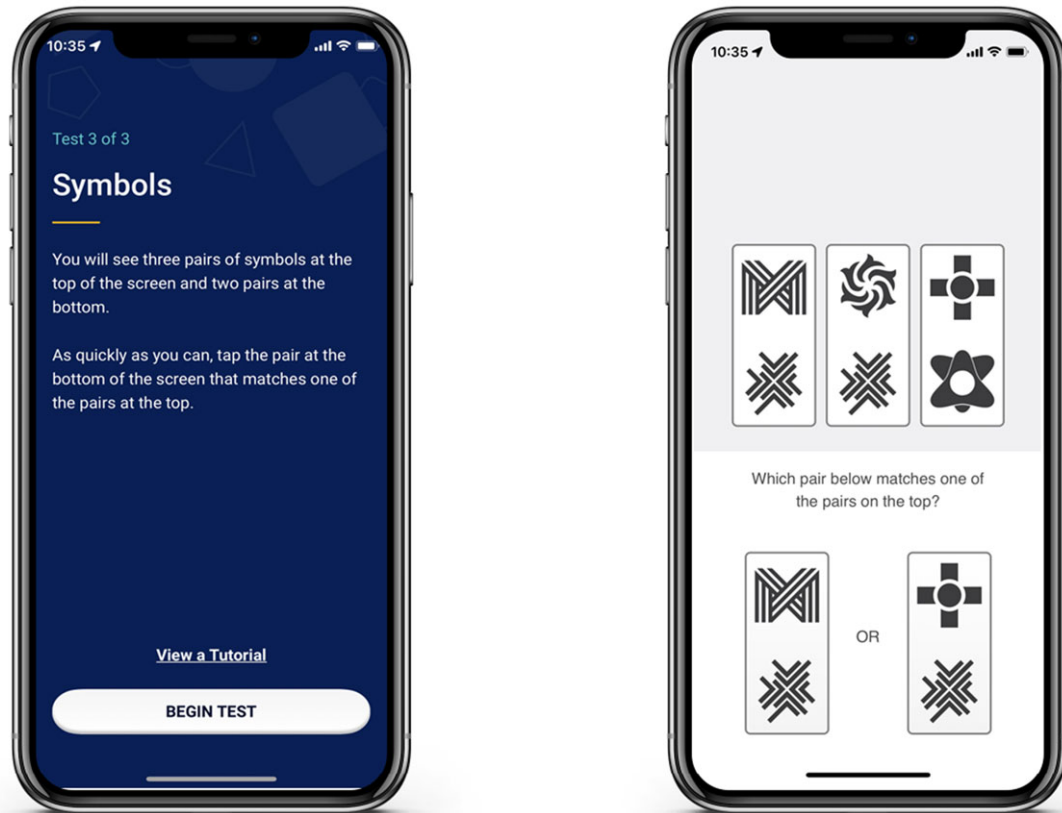
Figure 2. Screenshot of ARC symbols task. Participants completed up to 4 sessions a day for 7 days. Each symbols task included 12 trials of matching items at the bottom to one of the three options in the top row.

RT outliers were handled using the following procedures. From the correct trials, we removed RTs < 200ms as these RTs are below the threshold for participants to accurately encode and respond to stimuli. We next imputed any RTs for each subject beyond their Median + 3*Interquartile Range to that value. Thus, we retained as much data for each subject without undue influence of extremely long RTs (see Welhaf & Kane, 2023). This procedure resulted in < 1% of RTs being imputed for each participant.

Conventional cognitive assessments

Participants completed several in-person cognitive measures as part of their annual assessment at the Knight ADRC (see Hassenstab et al., 2016). For the current study, we examined scores on several measures that spanned three main cognitive domains. Specifically, we included composites for semantic memory/language (Category Fluency for Animals and Vegetables, Multilingual Naming Test), episodic memory (Weschler Memory Sale Paired Associated Recall, Free and Cued Selective Reminding Test, Craft Story 21 immediate and delayed recall), processing speed (Digit Span Forward and Number Symbol Test, a measure similar to the digit symbol coding test), and Attention Control (Stroop Incongruent Accuracy and CVOE Switch Accuracy). We chose to use these composites given their increased reliability over individual tests (see Hassenstab et al., 2016 and Weintraub et al., 2009; 2018 for additional information and references for individual tests). Cognitive assessments were completed 1.29 ± 4.41 weeks from the ARC assessment.

Analytic plan

To address our first aim, we report between-person reliability estimates of the two variability measures used in the present paper. For our second aim, if RT variability measures are sensitive to clinical status, we used linear regression to predict RT CoV and RMSSD (our outcome measures) from two main predictors of interest, *APOE* ε4 status and CDR status, while controlling for age, level of education, self-reported gender, and number of completed sessions. As part of our second aim, we also examined if RT variability could serve as a sensitive preclinical measure. To test this, we again conducted linear regression analysis in only the cognitively healthy (CDR 0) participants to determine if there was a reliable signal of *APOE* ε4 in a preclinical AD sample following from Aschenbrenner et al. (2023). Again, our outcome measure for this analysis was the RT variability measure with the main predictor being *APOE* ε4 status while controlling for age, level of education, self-reported gender, and number of completed sessions.

For our third and final aim, we tested the convergent and discriminant validity of our RT variability measures against conventional in-clinic cognitive measures among cognitively healthy older adults. We approached this question in two ways. First, we examined the simple bivariate correlations between the ARC RT Variability measures (CoV and RMSSD) and several in-clinic measures of episodic memory, semantic memory, processing speed, language, and attention control. Second, we asked if the reason individuals at risk for AD perform worse on in-clinic measures of cognition is due, in part, to poorer ability to maintain

attention. To test this, we conducted several mediation models using the PROCESS macro to test a simple mediation model 4 (Hayes, 2022). A different cognitive composite served as the outcome measure in each model, *APOE ε4* status served as the predictor variable, RT CoV or RMSSD as the proposed mediator, with age, education, and gender entered as covariates. We calculated bootstrapped 95% confidence intervals for the direct and indirect effects in each model. When the confidence interval did not contain zero, that effect can be interpreted as significant.

Results

Between-person reliability of RT variability measures

As previously described, aggregating scores across EMA sessions can increase the reliability of measures compared to traditional "one-shot" measures (e.g., Shiffman et al., 2008). We have previously shown that ARC measures show dramatically higher reliability compared to their in-clinic counterparts in as few as two sessions for the Symbols task and across multiple days for the other two ARC measures of visual working memory and associate memory (e.g., Nicosia et al., 2023). We also think it is important to test if measures of variability (which are inherently noisier) are also reliable when gathered using EMA methods. To test this, we fit a series of unconditional multilevel models for the RT CoV and RMSSD measures to compute between-person reliability (e.g., Raykov & Marcoulides, 2006; Sliwinski et al., 2018). The resulting reliabilities are displayed in Table 3. Between-person reliability for both measures was quite high. Notably, the between-person reliability for RMSSD was impressively high with only a single day worth of ARC tests (i.e., four sessions). RT CoV required roughly two days' worth of tests to reach a similar level of between-person reliability.

RT CoV differences by AD risk factors

Our first set of analyses determined if cognitively normal (asymptomatic) participants (CDR = 0) differed in their RT CoV compared to participants with very mild dementia symptoms (CDR = 0.5). As seen in Table 4, RT CoV was significantly predicted by age and CDR status. As expected, participants with very mild dementia had higher RT CoV over their week of testing indicating poorer moment-to-moment attention (see Fig. 3). The effect of *APOE ε4* status was not significant, but it was in the hypothesized direction with *APOE ε4* carriers showing higher CoV than noncarriers.

We next looked to see if this *APOE ε4* effect was apparent in a cognitively healthy sample of participants to examine a possible sensitive preclinical signal. As seen in Table 4, when restricting the sample to only CDR 0 participants, *APOE ε4* carriers showed greater RT CoV compared to noncarriers (see Fig. 4). Age also predicted RT CoV such that older participants had greater RT CoV. Thus, there appears to be a preclinical effect of *APOE ε4* status on moment-to-moment consistency of attention.

RMSSD as an alternative RT variability measure

As described earlier, we also attempted to provide converging evidence from a different measure of variability. To address this concern, we also calculated the RMSSD at each individual session and then averaged across these sessions for participant-level scores. We repeated the above regression analyses replacing RT CoV with RMSSD as the outcome to see if a different measure of variability would produce similar patterns.

Table 3. Reliabilities for symbols CoV and RMSSD measure

Sessions	Symbols CoV	Symbols RMSSD
4	0.723	0.898
7	0.830	0.944
16	0.907	0.980
27	0.941	0.989

ARC Participants received 4 sessions/day for 7 days.

Table 4. Linear regressions predicting symbols RT CoV

	Model 1 – Full Sample	Model 2 – CDR 0 Only
	b (SE)	b (SE)
Age (years)	.003 (.001)***	.003 (.001)***
Gender	−.000 (.007)	.004 (.007)
Education (years)	.000 (.001)	−.001 (.001)
Sessions Complete	−.001 (.000)	−.000 (.000)
APOE ε4 status	.013 (.007)^	.014 (.007)*
CDR Status	.031 (.012)**	
R^2	.105	.104
Adjusted R^2	.090	.091
Observations	377	345

APOE ε4 = Apolipoprotein ε4; CDR = Clinical Dementia Rating.
^ p = .070; * $p < .05$; ** $p < .01$; *** $p < .001$.

The results of the RMSSD models are displayed in Table 5. In the full sample, RMSSD was related to age and the number of completed sessions. Critically, CDR status was again a significant predictor with CDR 0.5 participants showing increased variability (i.e., higher RMSSD) compared to cognitively healthy participants. Restricting this to only the CDR 0 participants, we again found the critical *APOE ε4* effect. Specifically, *APOE ε4* carriers showed greater variability (i.e., higher RMSSD values) compared to noncarriers. Thus, our main results were largely consistent across two different measures of variability.

Evidence for convergent and discriminant validity of naturalistically assessed RT variability

We next examined the correlations between the RT variability measures and several conventional cognitive measures to test if there was an association between the ability to successfully maintain moment-to-moment attention in naturalistic settings and various cognitive abilities (tested in a supervised clinic setting). The resulting correlation matrix is presented in Table 6. First, both measures of variability were strongly correlated with each other ($r = .78$). Thus, these two measures appear to be largely redundant with one another.

In terms of correlations with other cognitive variables, as expected, both measures of variability showed moderate correlations (based on Cohen's 1988 framework) with the in-clinic cognitive composite measures. As expected, cognitively normal (CDR 0) participants who had greater variability in the ARC Symbols task had lower scores on the cognitive composites for episodic memory, semantic memory, attention control, and processing speed.

We next tested if part, or all, of the relationship between *APOE ε4* status and in-clinic cognition could be explained by RT variability. Table 7 shows the resulting direct and indirect effects of several mediation models. In general, the results across both measures of RT variability were rather similar. Aside from episodic memory, all relationships between *APOE ε4* status and the
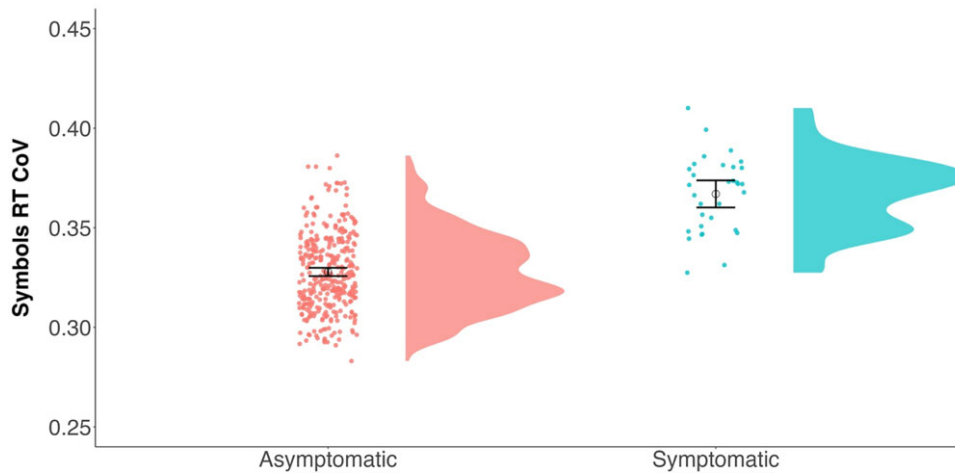
Figure 3. Raincloud plots (Allen et al., 2019) depicting differences in RT CoV by CDR group. Individual dots represent participant values with the corresponding distribution. Open circles reflect the group means and the corresponding 95% confidence interval. Asymptomatic (CDR = 0) $N = 345$; symptomatic (CDR = 0.5) $N = 32$.
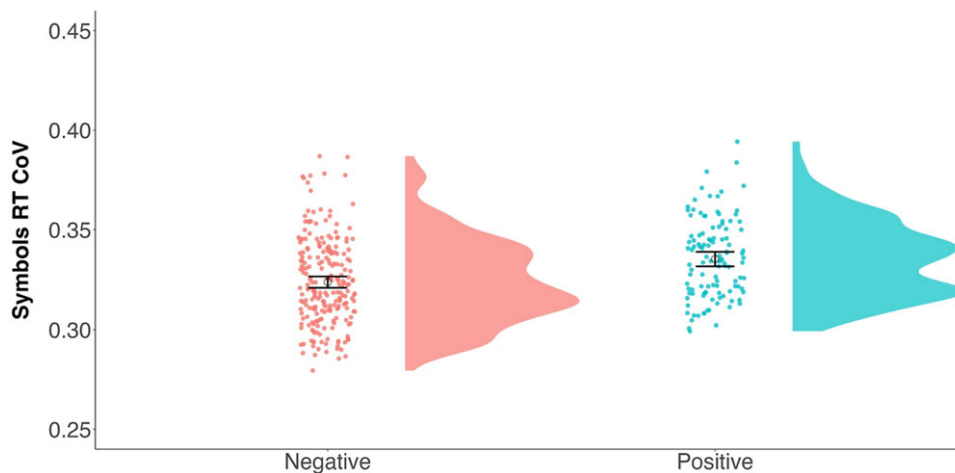

Figure 4. Raincloud plots (Allen et al., 2019) depicting differences in RT CoV by APOE ε4 status. Individual dots represent participant values with the corresponding distribution. Open circles reflect the group means and the corresponding 95% confidence interval. APOE ε4 negative $N = 228$; APOE ε4 positive $N = 117$.

cognitive composites could be fully explained by RT variability. Specifically, *APOE* ε4 carriers exhibited greater variability which in turn predicted poorer cognitive composite scores. The only relationship that could not be fully explained was that of episodic memory. That is, there was still a direct effect of *APOE* ε4 status on episodic memory performance even after accounting for RT variability.

Discussion

The current study examined variability in RTs using a novel smartphone application that measures cognition repeatedly in a variety of contexts in the participants natural environment over the course of 7 days. There were several key findings. First, RT variability can be reliably assessed using EMA approaches (and this was true across two different metrics of variability, CoV and RMSSD). Second, measures of RT variability appear sensitive to both clinical status (i.e., CDR status) and genetic disposition of AD (i.e., *APOE* ε4 status). Specifically, older adults who were already showing cognitive symptoms of AD had increased RT variability. Among cognitively normal participants, those at risk for AD also showed increased RT variability. Finally, measures of RT variability were correlated with conventional cognitive measures in theoretically meaningful ways and could largely explain the relationship between APOE ε4 status and conventional measures.

Can RT variability be reliably measured using EMA?

Our results suggest that variability in attention over the course of several seconds can be assessed using remote assessments across a variety of contexts and can leverage the strengths of EMA designs to obtain reliable estimates of RT variability. Between-person reliability estimates for both CoV and RMSSD surpassed traditional standards for "good" reliability of > .75 (Bruton et al., 2000). Specifically, both RT CoV and RMSSD only required four tests (which collectively took no more than 4 minutes) to get reliability that exceeds an in-lab test of the same construct and would take more than four minutes to administer. Further, these two measures of variability were strongly correlated with one another suggesting some redundancy. Previous work has also suggested that CoV measures are not recommended as they confound changes in mean RT with changes in the standard deviation (e.g., Hultsch et al., 2011; Stawski et al., 2019). However, the results of the current study suggest there are largely similar patterns regardless of the RT variability metric used. Thus, the heightened reliability and strong correlation between the two measures could minimize concerns that CoV is a confounded measure of variability.

These measures of variability also do not appear to be susceptible to floor or ceiling effects. As displayed in Figures 2 and 3, there was substantial inter-individual variation in RT CoV (and this was also true for RMSSD). One discrepancy between the

Table 5. Linear regressions predicting symbols RT RMSSD

	Model 1 – Full Sample	Model 2 – CDR 0 Only
	b (SE)	b (SE)
Age (years)	39.71 (7.22)***	43.37 (7.27)***
Gender	71.31 (86.29)	102.49 (85.75)
Education (years)	−11.39 (17.75)	−14.59 (17.74)
Sessions Complete	−9.42 (3.87)*	−9.33 (3.84)*
APOE ε4 status	156.56 (88.16)^	181.55 (88.21)*
CDR Status	424.41 (151.91)**	
R^2	.128	.128
Adjusted R^2	.114	.116
Observations	377	345

APOE ε4 = Apolipoprotein ε4; CDR = Clinical Dementia Rating.
^ $p = .077$; * $p < .05$; ** $p < .01$; *** $p < .001$.

RT variability measures appeared in terms of improvement in these scores over time (i.e., practice effects). Specifically, RT CoV was not associated with the number of completed sessions (either in the full sample or in the CDR 0 subsample), while RMSSD was associated with completed sessions. Thus, at least in some cases, how RT variability is measured can produce slightly different results. When assessing trial-to-trial variability using EMA, though, RMSSD might not be a gold standard approach for assessing RT variability. RT CoV, on the other hand, might be a suitable measure for longitudinal assessments that avoids common pitfalls of conventional cognitive measures.

Is RT variability measured with EMA sensitive to AD risk factors?

Our measures of RT variability also showed consistent patterns regarding group differences. Specifically, CDR status predicted unique variance in RT CoV and RMSSD suggesting that individuals showing clinical symptoms of dementia had more difficulty maintaining moment-to-moment attention. These results replicate several studies using in-clinic measures to assess RT variability (e.g., Cherbuin et al., 2010; Duchek et al., 2009; Jackson et al., 2012; Tse et al., 2010).

This effect of increased RT variability was also evident in cognitively normal individuals who carry a genetic risk for developing AD. Our results are consistent with previous work showing that genetic risk for AD (as defined as APOE ε4 status) is associated with increased performance variability within a task (e.g., Duchek et al., 2009). As noted earlier, some studies have reported null effects of APOE ε4 status on RT variability measures when using more basic RT tasks. These studies tested participants using a more traditional "one-shot" assessment of cognition during a supervised clinic session. Such tasks may be too short to allow attention to fluctuate on a trial-to-trial basis. For example, Lu et al. (2020) noted that the choice RT task they used lasted less than three mins. In the current study, the total time to complete Symbols over the week was closer to 8–10 minutes. Thus, capturing variability using a more natural testing platform in ARC highlights the sensitivity of ARC testing.

How is EMA-assessed RT variability related to conventional cognitive measures?

Previous work has proposed that attentional processes might underlie some of the declines in memory, and cognition more broadly, found in AD (e.g., Balota & Faust, 2001). This suggests that individuals who are at risk, or even showing early signs of AD,

Table 6. Correlation matrix of variables of interest for cognitively healthy (CDR = 0) participants

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) Age																			
2) Education	−.16**																		
3) Completed Sessions	−.03	−.03																	
4) Symbols CoV	.29***	−.07	−.11*																
5) Symbols RMSSD	.31***	−.08	−.14**	.78***															
6) FCSRT Free Recall	−.06	−.03	.14**	−.24***	−.22***														
7) WMS Associate Recall	.07	−.01	.12*	−.19***	−.17**	.44***													
8) Craft Story Recall Immediate	−.01	.20***	.07	−.27***	−.25***	.30***	.44***												
9) Craft Story Recall Delayed	−.05	.17**	.13*	−.31***	−.24***	.39***	.51***	.87***											
10) Animal Fluency	−.27***	.21***	−.01	−.28***	−.30***	.31***	.25***	.33***	.33***										
11) Vegetable Fluency	−.14***	.10	.09	−.27***	−.31***	.35***	.34***	.31***	.33***	.47***									
12) Multilingual Naming Test	−.01	.14*	.04	−.18***	−.10	.17**	.33***	.32***	.36***	.33***	.17**								
13) Digit Span Forward	−.01	.15**	.04	−.21***	−.13*	.14*	.18**	.18**	.22***	.16**	.10	.24***							
14) Number Symbol	−.34***	.02	.09	−.46***	−.52***	.27***	.27***	.18**	.24***	.34***	.23***	.21***	.13*						
15) CVOE Switch Accuracy	−.08	.04	.09	−.30***	−.36***	.13*	.19***	.19***	.22***	.19***	.13*	.23***	.04	.19**					
16) Stroop Incongruent Accuracy	−.23***	.05	.07	−.20***	−.19***	.10	.09	.15**	.17**	.09	.11	.07	.05	.16**	.19**				
17) Episodic Memory Composite	−.01	.11*	.14*	−.32***	−.28***	.68***	.76***	.83***	.88***	.37***	.42***	.37***	.23***	.31***	.19***	.16**			
18) Semantic Memory Composite	−.19***	.20***	.05	−.33***	−.33***	.38***	.40***	.42***	.44***	.80***	.75***	.68***	.21***	.35***	.23***	.12*	.52***		
19) Attention Control Composite	−.20***	.06	.10	−.32***	−.33***	.15*	.18**	.22***	.25***	.18**	.15*	.19**	.06	.23***	.77***	.78***	.26***	.23***	
20) Processing Speed Composite	−.21***	.11*	.08	−.41***	−.32***	.27***	.28***	.23***	.28***	.30***	.20***	.29***	.77***	.75***	.13*	.13*	.34***	.35***	.17**

FCSRT = Free and Cued Selective Reminder Test.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7. ß weights and bootstrapped 95% confidence intervals for *APOE* ε4–Cognitive relationship as mediated by RT variability measures

	Cov		RMSSD	
	Direct	Indirect	Direct	Indirect
Cognitive Composite				
Episodic Memory	**−.231 [−.391, −.071]**	**−.058 [−.115, −.004]**	−.238 [−.400, .075]	**−.051 [−.093, −.005]**
Semantic Memory/Language	−.042 [−.192, .107]	**−.048 [−.099, −.003]**	−.041 [−.190, .109]	**−.050 [−.090, −.005]**
Processing Speed	.103 [−.045, .252]	**−.069 [−.136, −.007]**	.086 [−.069, .241]	**−.051 [−.106, −.007]**
Attention Control	.075 [−.094, .244]	**−.056 [−.124, −.005]**	.072 [−.097, .240]	−.052 [−.117, .001]

The direct effect represents the direct relationship between *APOE* ε4 status and the cognitive variable. The indirect effect represents the relationship between *APOE* ε4 status and cognitive variable, mediated by either RT variability measure. All models included covariates of age at ARC assessment, years of education, and gender. Significant effects are bolded.

might have cognitive issues partially because they have difficulty maintaining attention. The correlational, and, importantly, the mediation results of the current study appear to largely support this claim. RT variability accounted for significant variance in the relationship between *APOE* ε4 status and in-clinic cognition. The only relationship that was not fully accounted for by RT variability was episodic memory. Given how central episodic memory deficits are to AD, it is perhaps unsurprising that individuals at risk for AD showed poorer episodic memory even after accounting for attentional abilities.

One possible explanation for these results is that older adults who are at risk for developing AD (e.g., *APOE* ε4 carriers) might also be susceptible to, or in the early stages of, accumulation of abnormal levels of β-amyloid and phosphorylated tau, which might be contributing to larger memory impairments beyond the attentional contributions. If this is the case, then one might expect to see episodic memory impairments even after accounting for attentional ability. Although we could not adequately examine AD biomarkers in the current study (due to disruptions in biomarker collection brought on by the COVID-19 pandemic), future research should examine the link between AD biomarkers and RT variability measures assessed in EMA studies of cognition.

The results of the current study also align with recent work from our group examining the utility of RT variability in other repeated testing contexts. Aschenbrenner et al., (2023) recently showed that over the course of 21 days of repeated testing, RT CoV in a go/no-go task was significantly correlated with n-back performance. Thus, brief, and repeated assessments, of attentional fluctuations represent a meaningful and reliable construct, is are related to cognitive performance in other domains.

Limitations

It is worth noting some limitations of the current study. First, an obvious limitation is that we used *APOE* ε4 status as a proxy for AD rather than measuring amyloid directly (although we did so in conjunction with CDR as a marker of AD severity which could be seen as a strength). Given that ARC testing started at the same time at the COVID-19 pandemic, collection of routine biomarker data was halted, and many participants may have outdated biomarker data which might lead to incorrect classification of AD status. Future work should also consider how biomarkers are related to naturalistic assessments of RT variability. Second, while we were able to find reliable differences in RT variability in a brief cognitive task, it is possible that 12 trials, which take roughly 20–40 s, might not fully capture differences in variability. Although some have identified attentional fluctuations in as little as 16 s (Esterman et al., 2013) future studies might consider increasing the number of trials to better capture moment-to-moment RT variability. As is

typical of studies of AD, Knight ADRC participants are typically highly educated and primarily White older adults with high motivation and willingness to engage in studies. Our results might not generalize to a more diverse population where different contextual and environmental factors might make maintaining consistent attention difficult. The Knight ADRC is making continued efforts to connect with more participants from underrepresented groups in the St Louis metropolitan and surrounding areas. It is our hope that through these efforts we will be better able to make more generalizable conclusions from our studies in the future.

Conclusions

This study extends previous findings that trial-to-trial RT variability is sensitive to AD risk factors by assessing variability using EMA methods, rather than in the constraints of a lab. RT variability differences we evident between cognitively normal participants and participants with very mild dementia. Within the cognitively normal participants, increased RT variability was associated with *APOE* ε4 status, such that carriers showed increased variability compared to noncarriers. Finally, measures of naturalistic RT variability were moderately correlated with conventional cognitive tests and could, in most cases, fully explain the relationship between *APOE* ε4 and cognition. This study highlights the utility of using brief, remote, cognitive testing to measure subtle variations in attention during everyday life.

Competing interests. None.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome open research*, 4, 63.

Anstey, K. J., Mack, H. A., Christensen, H., Li, S.-C., Reglade-Meslin, C., Maller, J., Kumar, R., Dear, K., Easteal, S., & Sachdev, P. (2007). Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia*, 45(8), 1911–1920.

Aschenbrenner, A. J., Hassenstab, J., Morris, J. C., Cruchaga, C., & Jackson, J. J. (2023). Relationships between hourly cognitive variability and risk of alzheimer's disease revealed with mixed-effects location scale models. *Neuropsychology*, 38(1), 69–80. https://doi.org/10.1037/neu0000905

Aschenbrenner, A. J., Welhaf, M., Hassenstab, J., & Jackson, J. J. (2023). Antecedents and consequences of mind wandering states in healthy aging and mild cognitive impairment. https://doi.org/10.31234/osf.io/n276e

Balota, D. A., & Faust, M. (2001). Attention in dementia of the Alzheimer's type. *Handbook of Neuropsychology*, 6, 51–80.

Bielak, A. A. M., Mogle, J., & Sliwinski, M. J. (2019). What did you do today? Variability in daily activities is related to variability in daily cognitive performance. *The Journals of Gerontology: Series B*, 74(5), 764–771.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, 86(2), 94–99.

Cerino, E. S., Katz, M. J., Wang, C., Qin, J., Gao, Q., Hyun, J., Hakun, J. G., Roque, N. A., Derby, C. A., Lipton, R. B., & Sliwinski, M. J. (2021). Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: Results from the einstein aging study. *Frontiers in Digital Health*, 3, 758031.

Cherbuin, N., Sachdev, P., & Anstey, K. J. (2010). Neuropsycho- logical predictors of transition from healthy cognitive aging to mild cognitive impairment: The PATH through life study. *American Journal of Geriatric Psychiatry*, 18(8), 723–733. https://doi.org/10.1097/Jgp.0b013e3181cdecf1

Christ, B. U., Combrinck, M. I., & Thomas, K. G. (2018). Both reaction time and accuracy measures of intraindividual variability predict cognitive performance in Alzheimer's disease. *Frontiers in Human Neuroscience*, 12, 124.

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of alzheimer's disease in late onset families. *Science*, 261(5123), 921–923. https://doi.org/10.1126/science.8346443

Cruchaga, C., Kauwe, J. S. K., Harari, O., Jin, S. C., Cai, Y., Karch, C. M., Benitez, B. A., Jeng, A. T., Skorupa, T., Carrell, D., Bertelsen, S., Bailey, M., McKean, D., Shulman, J. M., De Jager, P. L., Chibnik, L., Bennett, D. A., Arnold, S. E., Harold, D., Sims, R., Gerrish, A., Williams, J., Van Deerlin, V. M., Lee, V. M.-Y., Shaw, L. M., Trojanowski, J. Q., Haines, J. L., Mayeux, R., Pericak-Vance, M. A., Farrer, L. A., Schellenberg, G. D., Peskind, E. R., Galasko, D., Fagan, A. M., Holtzman, D. M., Morris, M., A. M., Goate, & GERAD Consortium, Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimer Disease Genetic Consortium (ADGC) (2013). GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron*, 78(2), 256–268. https://doi.org/10.1016/j.neuron

Dixon, R. A., Garrett, D. D., Lentz, T. L., MacDonald, S. W., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*, 21(3), 381–399.

Duchek, J. M., Balota, D. A., Thomas, J. B., Snyder, A. Z., Rich, P., Benzinger, T. L., Fagan, A. M., Holtzman, D. M., Morris, J. C., & Ances, B. M. (2013). Relationship between Stroop performance and resting state functional connectivity in cognitively normal older adults. *Neuropsychology*, 27(5), 516–528. https://doi.org/10.1037/a0033402

Duchek, J. M., Balota, D. A., Tse, C. S., Holtzman, D. M., Fagan, A. M., & Goate, A. M. (2009). The utility of intraindividual variability in selective attention tasks as an early marker for Alzheimer's disease. *Neuropsychology*, 23(6), 746–758.

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, 23(11), 2712–2723.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, 125(6), 777–799.

Gorus, E., De Raedt, R., Lambert, M., Lemper, J. C., & Mets, T. (2008). Reaction times and performance variability in normal aging, mild cognitive impairment, and Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology*, 21(3), 204–218.

Gyurkovics, M., Balota, D. A., & Jackson, J. D. (2018). Mind-wandering in healthy aging and early stage Alzheimer's disease. *Neuropsychology*, 32(1), 89–101.

Hassenstab, J., Chasse, R., Grabow, P., Benzinger, T. L. S., Fagan, A. M., Xiong, C., Jasielec, M., Grant, E., & Morris, J. C. (2016). Certified normal: Alzheimer's disease biomarkers and normative estimates of cognitive functioning. *Neurobiology of Aging*, 43, 23–33.

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Haynes, B. I., Bauermeister, S., & Bunce, D. (2017). A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia, and mortality. *Journal of the International Neuropsychological Society*, 23(5), 431–445.

Hultsch, D. F., MacDonald, S. W., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, 14(4), 588–598.

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. (2011). Intraindividual variability, cognition, and aging. In *The Handbook of Aging and Cognition* (pp. 497–562). Psychology Press.

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual variability in healthy aging and early-stage alzheimer disease. *Neuropsychologia*, 50(3), 357–366.

Kay, C. D., Seidenberg, M., Durgerian, S., Nielson, K. A., Smith, J. C., Woodard, J. L., & Rao, S. M. (2017). Motor timing intraindividual variability in amnestic mild cognitive impairment and cognitively intact elders at genetic risk for alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 39(9), 866–875.

Lu, K., Nicholas, J. M., James, S. N., Lane, C. A., Parker, T. D., Keshavan, A., Keuss, S. E., Buchanan, S. M., Murray-Smith, H., Cash, D. M., Sudre, C. H., Malone, L. B., Coath, W., Wong, A., Henley, S. M. D., N. C., Fox, M., Richards, J. M., Schott, & S. J., Crutch (2020). Increased variability in reaction time is associated with amyloid beta pathology at age 70. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 12(1), e12076.

MacDonald, S. W., & Stawski, R. S. (2020). Longitudinal changes in response time mean and inconsistency exhibit predictive dissociations for risk of cognitive impairment. *Neuropsychology*, 34(3), 264–275.

Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontologia*, 50(1), 7–11.

Mishra, S., Blazey, T. M., Holtzman, D. M., Cruchaga, C., Su, Y., Morris, J. C., Benzinger, T. L. S., & Gordon, B. A. (2018). Longitudinal brain imaging in preclinical Alzheimer disease: Impact of APOE ε4 genotype. *Brain*, 141(6), 1828–1839.

Morris, J. C. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43(11), 2412–2414.

Morris, J. C., Roe, C. M., Xiong, C., Fagan, A. M., Goate, A. M., Holtzman, D. M., & Mintun, M. A. (2010). APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Annals of Neurology*, 67(1), 122–131. https://doi.org/10.1002/ana.21843

Nicosia, J., Aschenbrenner, A. J., Balota, D. A., Sliwinski, M. J., Tahan, M., Adams, S., & Hassenstab, J. (2023). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. *Journal of the International Neuropsychological Society*, 29(5), 459–471.

Perry, R. J., & Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease: A critical review. *Brain*, 122(3), 383–404.

Raykov, T., & Marcoulides, G. A. (2006). On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling*, 13(1), 130–141.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4(1), 1–32.

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. *Assessment*, 25(1), 14–30.

Stawski, R. S., MacDonald, S. W., Brewster, P. W., Munoz, E., Cerino, E. S., & Halliday, D. W. (2019). A comprehensive comparison of quantifications of intraindividual variability in response times: A measurement burst approach. *The Journals of Gerontology: Series B*, 74(3), 397–408.

Stawski, R. S., MacDonald, S. W., & Sliwinski, M. J. (2015). *Measurement burst design* (pp. 1–5). The Encyclopedia of Adulthood and Aging.

Tse, C. S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, 24(3), 300–315.

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, 49, 110–128.

Unsworth, N., & Miller, A. L. (2021). Individual differences in the intensity and consistency of attention. *Current Directions in Psychological Science*, *30*(5), 391–400.

Von Neumann, J., Kent, R. H., Bellinson, H. R., & Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, *12*(2), 153–162.

Weintraub, S., Besser, L., Dodge, H. H., Teylan, M., Ferris, S., Goldstein, F. C., Giordani, B., Kramer, J., Loewenstein, D., Marson, D., Mungas, D., Salmon, D., Welsh-Bohmer, K., Zhou, X.-H., Shirk, S. D., Atri, A., Kukull, W. A., Phelps, C., & Morris, J. C. (2018). Version 3 of the Alzheimer disease centers' neuropsychological test battery in the uniform data set (UDS). *Alzheimer Disease and Associated Disorders*, *32*(1), 10–17.

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., Cummings, J., DeCarli, C., Foster, N. L., Galasko, D., Peskind, E., Dietrich, W., Beekly, D. L., Kukull, W. A., & Morris, J. C. (2009). The alzheimer's disease centers' uniform data set (UDS): The neuropsychological test battery. *Alzheimer Disease and Associated Disorders*, *23*(2), 91–101.

Welhaf, M. S., & Kane, M. J. (2023). *A Nomothetic span approach to assessing the construct validity of sustained attention measures*. Psychological Research.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition*, *49*(3), 402–419.

Wilks, H., Aschenbrenner, A. J., Gordon, B. A., Balota, D. A., Fagan, A. M., Musiek, E., & Hassenstab, J. (2021). Sharper in the morning: Cognitive time of day effects revealed with high-frequency smartphone testing. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 825–837.