

Reliability of P3 Event-Related Potential during Working Memory Across the Spectrum of Cognitive Aging

1 Hannes Devos^{1*}, Jeffrey Burns^{2,3}, Ke Liao⁴, Pedram Ahmadnezhad,¹ Jonathan D. Mahnken,^{3,4}
2 William M. Brooks,^{2,3,4} Kathleen Gustafson^{3,4}

3 ¹Laboratory for Advanced Rehabilitation Research in Simulation, Department of Physical Therapy
4 and Rehabilitation Science, University of Kansas Medical Center, Kansas City, KS, USA

5 ²Department of Neurology, University of Kansas Medical Center, Kansas City, KS

6 ³University of Kansas Alzheimer's Disease Center, University of Kansas Medical Center, Kansas
7 City, KS

8 ⁴Hoglund Brain Imaging Center, University of Kansas Medical Center, Kansas City KS

9 ⁵Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas City, KS

10

11 *** Correspondence:**

12 Hannes Devos

13 Department of Physical Therapy and Rehabilitation Science, University of Kansas Medical Center,
14 MS 2002 3901 Rainbow Blvd., Kansas City KS

15 E-mail: hdevos@kumc.edu

16 **Keywords: event-related potentials, electro-encephalography, reliability, working memory,**
17 **older adults, mild cognitive impairment, Alzheimer disease, preclinical**

18 **Abstract**

19 Event-related potentials (ERPs) offer unparalleled temporal resolution in tracing distinct
20 electrophysiological processes related to normal and pathological cognitive aging. The stability of
21 ERPs in older individuals, who inherently show more intraindividual variability in cognitive
22 functions, has not been established. In this test-retest reliability study, 39 older individuals (age 74.10
23 (5.4) years; 23 (59%) women; 15 non β -amyloid elevated, 16 β -amyloid elevated, 8 cognitively
24 impaired) with scores on the Montreal Cognitive Assessment (MOCA) ranging between 3 and 30
25 completed a working memory (n-back) test with three levels of difficulty at baseline and two-week
26 follow-up. Stability of the ERP was evaluated on grand averaged task effects for both visits. P3 peak
27 amplitude and latency were measured in frontal channels. P3 peak amplitude at Fz, our main outcome
28 variable, showed excellent reliability in 0-back (intraclass correlation coefficient (ICC), 95%
29 confidence interval = 0.82 (0.67 – 0.90) and 1-back (ICC = 0.87 (0.76 – 0.93), however, only fair
30 reliability in 2-back (ICC = 0.53 (0.09 – 0.75). Reliability of P3 peak latencies was substantially
31 lower, with ICCs ranging between 0.17 for 2-back and 0.54 for 0-back. Generalized linear mixed
32 models showed no confounding effect of age, group, or task difficulty on stability of P3 amplitude
33 and latency of Fz. By contrast, MOCA scores tended to negatively correlate with P3 amplitude of Fz
34 ($p=0.07$). We conclude that P3 peak amplitude and latency provide a stable measure of
35 electrophysiological processes in older individuals. However, impaired cognition may affect the
36 stability of the ERP response.

37

38 **1 Introduction**

39 The aging process is characterized by gradual decline in physical, neurobiological and cognitive
40 functions that may impact instrumental activities of daily living (iADL) such as driving, doing
41 household chores, managing finances, medication adherence, or grocery shopping.(1, 2)
42 Deterioration in these iADL becomes more apparent with age-related neurodegeneration such as mild
43 cognitive impairment (MCI) and Alzheimer’s disease (AD).(3) Executive functions in particular are
44 paramount in carrying out numerous iADL, but are also vulnerable to the effects of normal and
45 pathological cognitive aging.(4, 5) Working memory is one core executive function that relates to the
46 ability to temporarily store, process, and manipulate the information necessary for higher order
47 cognitive tasks such as decision making, learning, and reasoning.(6) Working memory stems from
48 the interaction between attention, short-term retention and manipulation of information, carried out
49 by the coordinated activation of many brain regions.(7)

50 The prefrontal cortex has particularly been associated with working memory.(8) Consequently, the
51 prefrontal cortex is highly susceptible to the effects of aging and early neurodegeneration.(9, 10) A
52 recent meta-analysis pooling functional magnetic resonance imaging studies suggested a gradual and
53 linear decline in prefrontal cortex engagement in older individuals.(11) Similarly,
54 electrophysiological processes also decline with age. The P3, a positive peak that appears with a
55 latency between 250 to 500ms in the event-related potential (ERP), has been implicated in attention
56 and working memory processes across the lifespan.(12) A previous study showed reduced positivity
57 in P3 central-frontal and parietal ERPs in older adults,(13) whereas others demonstrated frontal
58 hyperactivity in P3 coupled with parietal or posterior hypoactivity.(14, 15). Despite the ambiguity in
59 ERP findings, most studies conclude that the abnormal ERP response in older individuals reflects
60 inefficient or compensatory use of neural resources due to frontal cortex dysfunction.(13, 14)

61 Therefore, electrophysiological responses to working memory tasks are convenient measures to test
62 hypotheses related to frontal cortex function, normal cognitive aging, and early neurodegeneration.

63 The ability to distinguish natural variability and measurement error from biologically relevant
64 cognitive changes due to aging or early neurodegeneration is valuable to provide informed decisions
65 on diagnosis, monitoring, and treatment of cognitive impairments.(16) However, older adults show
66 more intraindividual variability in performance measures of working memory compared to younger
67 adults. The age-related changes in intraindividual variability of performance measures become even
68 more apparent with increasing cognitive demand.(9) This increased intraindividual may also stem
69 from the heterogeneity of cognitive profiles in older individuals, especially when patients with MCI
70 and AD are included.(17) The intraindividual variability observed in performance measures is
71 believed to be linked to frontal cortex dysfunction,(9) which may therefore also affect intraindividual
72 variability of the ERP response in older adults.(18) To date, few studies have investigated test-retest
73 reliability of P3 ERP in healthy older adults(19-21) The test-retest reliability of the P3 ERP in older
74 individuals with a heterogeneous cognitive profile has yet to be established.

75 The aims were to (1) characterize test-retest reliability of P3 amplitude and latency, particularly in
76 the frontal-midline channels; and (2) investigate the impact of age, disease groups, cognitive status,
77 and task difficulty P3 amplitude and latency.

78 **2 Materials and Methods**

79 **2.1 Participants**

80 This test-retest reliability study included 39 right-handed participants recruited from the KU
81 Alzheimer's Disease Center between 05/03/2018 and 03/10/2020. Inclusion criteria were (1) age
82 older than 65; (2) having undergone a previous amyloid PET scan of the brain; (3) ability to

83 understand the instructions in English; and (4) informed consent. Exclusion criteria were: (1)
84 currently taking steroids, benzodiazepines, or neuroleptics; (2) history of any substance abuse; and
85 (3) history of a neurological disorder other than MCI or AD. Sixteen were cognitively normal older
86 adults with no elevated amyloid PET scans ($A\beta^-$), 15 were cognitive normal with elevated amyloid
87 PET scans ($A\beta^+$), and eight had a clinical diagnosis of mild cognitive impairment with positive
88 amyloid PET scans. Participants completed their two-week follow-up session 16 ± 8 days after the
89 first session. Each session lasted about 60 minutes including rest breaks.

90 **2.2 Procedure**

91 **2.2.1 Demographic and Clinical Information**

92 Age, sex, and education were recorded. General cognitive functions were evaluated with the
93 Montreal Cognitive Assessment (MOCA).(22) Scores on the MOCA range between 0 and 30.

94 **2.2.2 N-back Test**

95 In the n-back test, participants are shown a series of letters and are instructed to press a button when
96 the current stimulus is the same as the item presented n-positions back. The cognitive demand of the
97 n-back task increases with each number, while the perceptual and motor demands remain constant.
98 Whereas in the 0-back and 1-back the stimulus on screen is held in the focus of attention, the more
99 difficult versions of the n-back tests require constant switching from the focus of attention to short-
100 term memory.(23) Higher levels of difficulty require continuous mental effort to update information
101 of new stimuli and maintain representations of recently presented stimuli.(24)

102 In this study, the 0-back, 1-back, and 2-back tests were administered. Participants sat in a
103 comfortable chair at 26 inches in front of the computer screen with the center of the screen at eye
104 level. White letters appeared on a black screen. Prior to each test, participants were given a practice

105 trial consisting of 7 non-targets and 3 targets. The practice trials were repeated until the participant
106 felt comfortable with the instructions. Each test comprised 180 trials, including 60 trials that needed a
107 response (target, 33.3%) and 120 trials for which a response was not required (nontarget, 66.7%).
108 Each letter was presented for 500 ms on the computer screen followed by a blank interstimulus
109 interval for 1700 ms, with a random jitter of +/-50 ms. The maximum time to accept the response was
110 2150 ms. The total task time was ~7 minutes. In the 0-back test, participants were instructed to press
111 the left mouse button as soon as the letter “X” (target) appeared on the screen while ignoring the
112 other letters (nontarget). In the 1-back test, participants were instructed to press the button if the
113 current letter on the screen was the same as the letter previously shown (target). In the 2-back test,
114 participants were instructed to press the button when the current letter was the same as the one
115 presented two places before (target). The number of hits (accuracy) and response times to the hits
116 were the main behavioral performance outcome measures.

117 **2.2.3 P3 ERP**

118 Continuous electro-encephalogram (EEG) was acquired using a Philips EGI high-density system
119 from 256 scalp electrodes, digitized at 1,000 Hz. Data were filtered from 0.50 to 30 Hz using EGI
120 software. Data were online referenced to Cz and offline rereferenced to the left and right mastoid. All
121 other EEG processing was done in EEGLab(25) and in ERPLab.(26) Various artifacts unrelated to
122 cognitive functions, including ocular and muscular movement, cardiovascular signal were identified
123 and removed using independent component analysis (ICA). Signal from bad electrodes were
124 interpolated using surrounding electrode data. Stimulus-locked ERPs were extracted from the n-back
125 tests and segmented into epochs of 100 ms before to 1000ms after stimulus onset, and baseline
126 corrected using the prestimulus interval. Scalp locations and measurement windows for the P3
127 component were based on their spatial extent and latency after inspection of grand average
128 waveforms (collapsed across the two sessions). P3 peak amplitude and peak latency of the task effect

129 were considered the main electrophysiological outcome measures. The difference waveform was
130 calculated by subtracting the average ERP elicited from the targets from the average ERP elicited by
131 non-targets for each participant. The P3 component time window was established between 200 ms
132 and 400 ms for all three tests. Because of the prefrontal cortex involvement in working memory, we
133 identified a priori the Fz channel as the main channel, but also calculated reliability of other pre-
134 identified electrode locations, i.e., Cz, Pz, F3, and F4. Cz was interpolated using the surrounding five
135 channels. No participants were removed from the analyses because of artifacts. However, one
136 participant disengaged during the 2-back test and was therefore excluded from the analyses.

137 **2.3 Data Analysis**

138 Descriptive analysis including mean (standard deviation) and frequency count of participants'
139 general, performance measures, and ERP data were performed as appropriate. Intra-class correlation
140 coefficients (ICC) were used to calculate test-retest reliability of performance measures and P3
141 amplitude and latency. ICCs reflect the consistency of a measure taking into account variance related
142 to the time of testing.(27) ICC values less than 0.40 were considered poor; values between 0.40 and
143 0.59 fair, values between 0.60 and 0.74 good, and values between 0.75 and 1.00 excellent.(28).
144 Bland-Altman plots were used to visualize the measurement precision of amplitude and latency
145 across the test moments.(29) Intersubject stability according to subject rankings was calculated using
146 the Pearson r correlation coefficient. Generalized linear mixed models were employed to evaluate the
147 effect of age, diagnosis (A β -; A β +; MCI/AD), MOCA scores, and task difficulty on stability of the
148 P3 amplitude and latency. Stability of P3 amplitude (latency) was calculated as the squared
149 difference of P3 amplitude (latency) at follow-up and baseline. The Kolmogorov–Smirnov test was
150 employed to test the normality of our data distribution in addition to visualization of Q-Q plots. All
151 analyses were done using SAS 9.4 software. The threshold of significance was set at $p = 0.05$.

152 **3 Results**

153 **3.1 Participant Characteristics**

154 Participants (n = 39) were on average 74.05 (5.37) years old and scored 26.44 (4.76) on the MOCA
155 scale. However, MOCA scores ranged from 3 to 30, indicating a wide range of severe to normal
156 cognitive functions. About 60% (n = 23) were women.

157 **3.2 Test-Retest Reliability of Performance Measures**

158 All ICC values of hits (accuracy) and response times of each n-back test demonstrated excellent
159 reliability (Supplementary Table 1). ICCs of hits ranged between 0.92 (1- and 2-back) and 0.99 (0-
160 back) and were slightly higher than the ICCs of response times, ranging between 0.76 (2-back) and
161 0.89 (1-back). Pearson r correlations ranged from 0.65 (0-back response time) to 0.99 (0-back hits).

162 **3.3 Test-Retest Reliability of ERP Measures**

163 Grand average waveforms of the task effect from all channels at baseline and follow-up are displayed
164 in Figure 1. The 3D scalp map is embedded in the figure to demonstrate the task effect at P3.
165 Considerable overlap in ERP response within the P3 time window (200 – 400 ms post-stimulus) was
166 observed at baseline and two-week follow-up.

167 The ICC values of P3 peak amplitude and peak latency of the key electrode locations are displayed in
168 Table 1. Overall, P3 amplitude showed greater reliability compared to P3 latency across channels and
169 task difficulty levels. Also, ICCs of the 0-back and 1-back were consistently higher than those
170 calculated for the 2-back.

171 For the main channel location Fz, excellent reliability was found in P3 amplitude for 0-back (ICC =
172 0.82) and 1-back (ICC = 0.87). P3 amplitude of Fz for 2-back only showed fair reliability. Reliability
173 scores of P3 latency at Fz were fair for 0-back (ICC = 0.54) and 1-back (ICC = 0.47), but poor for 2-

174 back (ICC = 0.17). Figure 2 shows the Bland-Altman plots for P3 peak amplitude and peak latency at
175 the Fz channel. All plots demonstrated equal distribution of the data around zero, indicating no bias
176 in the results and no heteroscedasticity within the data.

177 Finally, generalized linear mixed models were employed to evaluate the effect of age, disease
178 diagnosis (A β -; A β +; MCI/AD), cognitive status, and task difficulty on stability of squared P3 peak
179 amplitude and latency at the Fz channel. Age ($p=0.74$), disease diagnosis ($p = 0.67$), and task
180 difficulty ($p = 0.70$) did not affect the stability of the P3 amplitude response, although individuals
181 with lower MOCA cognitive scores tended to show more variability in P3 amplitude ($p = 0.07$).

182 Age ($p = 0.60$), disease diagnosis ($p = 0.55$), MOCA ($p = 0.52$), or task difficulty ($p = 0.95$) did not
183 affect the stability of the P3 latency response.

184 We recalculated ICC for 0-back, 1-back, and 2-back in participants who scored 26 or higher on
185 MOCA ($n = 32$) and those scoring lower than 26 ($n = 7$). ICC values showed more variance in 0-back
186 and in 2-back in the group with lower MOCA scores, but ICC values were not worse across the n-
187 back tests in this group (Supplementary Tables 2 and 3). Whereas ICCs were similar in the A β - and
188 A β + groups, lower ICCs were found for the MCI/AD group (Supplementary Tables 2 and 3).

189 **4 Discussion**

190 This test-retest reliability study provides critical information on the stability of electrophysiological
191 measures related to working memory in healthy older adults, older adults with increased risk of
192 dementia, and those with MCI or AD. Our results showed that most P3 ERPs in the frontal channels
193 provide fair to excellent reliability to measure electrophysiological processes of cognitive aging in
194 older adults with and without cognitive impairments. Similar to previous studies, the reliability is
195 superior in measures of amplitude compared to latency.(20, 21, 30, 31) The robustness of P3 stability
196 is not affected by age, disease diagnosis, or task difficulty, however, there is a trend that lower
197 MOCA scores may affect the stability of the P3 amplitude response.

198 The body of evidence related to reliability of P3 ERPs is sparse, and typically restricted to healthy
199 young,(30-34) middle-aged,(31) and older individuals.(19-21). Few studies have reported reliability
200 measures in neurological conditions.(35) The reliability analyses in our study produced fair to
201 excellent ICC values across the n-back tests. Whereas ICC values provide a single measure of the
202 magnitude of agreement, Bland-Altman plots depict a graphical display of bias across the two test
203 moments.(36) Visual inspection of the Bland-Altman plots showed average difference in ERP
204 responses between first and second testing close to 0, with equal spread of data points around the
205 average difference line. These findings suggest that 2 weeks follow-up is sufficient to wash out any
206 potential adaptation, test, or practice effect of the n-back on ERPs in older individuals.

207 Comparison of our results with other test-retest studies of ERPs in older adults is complicated by lack
208 of consistency in terms of the ERP components that are investigated, the tests of working memory,
209 the choice of channel locations, the extracted P3 metric, the P3 window measurement, and the test-
210 retest reliability intervals.(19-21). Our research design most closely aligns with a study that
211 compared ERPs to novel stimuli collected at baseline and 7-week follow-up in healthy older
212 individuals.(21) Similar to our study, this study also found excellent reliability for P3 mean
213 amplitude (ICC =0.86, 95% CI, 0.78 – 0.92), and poorer reliability for P3 mean latency (ICC = 0.56,
214 0.30 – 0.73). Our study demonstrated larger confidence intervals in some of the amplitude and
215 latency measures, which might have been due to the greater cognitive heterogeneity of our sample.

216 Another reliability study also reported considerably lower reliability in P3 amplitude (ICC = -0.02)
217 and latency (ICC = -0.17) in seven individuals experiencing cognitive difficulties following traumatic
218 brain injury compared to healthy peers (ICC = 0.84 for amplitude and 0.64 for latency).(35)

219 Combined, these findings point towards a potential confounding effect of cognitive impairment on
220 stability of ERPs in neurological conditions.

221 No effect of age, task difficulty, or disease diagnosis was found on stability of the P3 ERP in the n-
222 back task. Most participants in our study were cognitively normal, either without (n = 15) or with (n

223 = 16) elevated beta amyloid. Accumulation of beta amyloid deposits in the brain is known to increase
224 the risk of developing AD.(37) The fair to excellent reliability of P3 amplitude and latency in our
225 group of older adults provides opportunities for studying the effect of amyloid deposits on neural
226 transmission in preclinical AD using ERP. P3 amplitudes are smaller in AD compared to
227 controls.(38) ERPs also show useful in predicting conversion to AD, with accuracy rates ranging
228 between 70% and 94%.(39) Patients with AD exhibit prolonged latency in P3 ERP compared to age-
229 matched controls(40). These prolonged latencies observed in patients with AD become particularly
230 apparent in the cognitive domains of executive function, memory, and language.(41) The ability of
231 P3 ERP to discriminate between mild cognitive impairment (MCI) and AD(42) opens avenues for
232 investigation of ERP in detecting preclinical AD.(43)

233 Although this study included participants with a wide range of cognitive impairment, most were
234 cognitively normal. Future studies should include a larger sample of participants with MCI and AD
235 to confirm the confounding effect of impaired cognition on the stability of the P3 response. Although
236 the n-back is arguably the most ubiquitous working memory test used in ERP studies across the age
237 spectrum,(23) we did not establish reliability of ERP in other cognitive domains known to deteriorate
238 in older age, such as memory and language, and this remains an opportunity for further investigation.
239 Our results are therefore only generalizable to the specific subdomain of working memory. Future
240 research should also include multiple testing sessions over extended periods of time to evaluate the
241 sensitivity of ERP to detect subtle neurobiological changes due to normal and pathological aging.

242 **5 Conclusion**

243 We set out to assess the test-retest reliability of ERP response in older adults with a heterogeneous
244 cognitive profile. Consistent with other studies, P3 amplitude and latency show fair to excellent
245 reliability across different levels of task difficulty. However, impaired cognition may potentially
246 affect the stability of the P3 ERP response.

247 **References**

- 248 1. Carmona-Torres JM, Rodríguez-Borrego MA, Laredo-Aguilera JA, López-Soto PJ,
249 Santacruz-Salas E, Cobo-Cuenca AI. Disability for basic and instrumental activities of daily living in
250 older individuals. *PLOS ONE* (2019) 14(7):e0220157. doi: 10.1371/journal.pone.0220157.
- 251 2. Moon S, Ranchet M, Abiodun, Tant M, David, Mukaila, et al. The Impact of Advanced Age
252 on Driving Safety in Adults with Medical Conditions. *Gerontology* (2018) 64(3):291-9. doi:
253 10.1159/000486511.
- 254 3. Jekel K, Damian M, Wattmo C, Hausner L, Bullock R, Connelly PJ, et al. Mild cognitive
255 impairment and deficits in instrumental activities of daily living: a systematic review. *Alzheimer's*
256 *Research & Therapy* (2015) 7(1). doi: 10.1186/s13195-015-0099-0.
- 257 4. Overdorp EJ, Kessels RPC, Claassen JA, Oosterman JM. The Combined Effect of
258 Neuropsychological and Neuropathological Deficits on Instrumental Activities of Daily Living in
259 Older Adults: a Systematic Review. *Neuropsychology Review* (2016) 26(1):92-106. doi:
260 10.1007/s11065-015-9312-y.
- 261 5. Tabira T, Hotta M, Murata M, Yoshiura K, Han G, Ishikawa T, et al. Age-Related Changes in
262 Instrumental and Basic Activities of Daily Living Impairment in Older Adults with Very Mild
263 Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders Extra* (2020):27-37. doi:
264 10.1159/000506281.
- 265 6. Baddeley A. Working memory. *Science* (1992) 255(5044):556-9. doi:
266 10.1126/science.1736359.
- 267 7. Eriksson J, Edward, Lansner A, Bergström F, Nyberg L. Neurocognitive Architecture of
268 Working Memory. *Neuron* (2015) 88(1):33-46. doi: 10.1016/j.neuron.2015.09.020.
- 269 8. Bahmani Z, Clark K, Merrikhi Y, Mueller A, Pettine W, Isabel Vanegas M, et al. Prefrontal
270 Contributions to Attention and Working Memory. *Curr Top Behav Neurosci* (2019) 41:129-53. Epub
271 2019/02/11. doi: 10.1007/7854_2018_74. PubMed PMID: 30739308; PubMed Central PMCID:
272 PMC6689265.
- 273 9. West R, Murphy KJ, Armilio ML, Craik FIM, Stuss DT. Lapses of Intention and Performance
274 Variability Reveal Age-Related Increases in Fluctuations of Executive Control. (2002) 49(3):402-19.
275 doi: 10.1006/brcg.2001.1507.
- 276 10. Ranchet M, Morgan JC, Akinwuntan AE, Devos H. Cognitive workload across the spectrum
277 of cognitive impairments: A systematic review of physiological measures. *Neuroscience &*
278 *Biobehavioral Reviews* (2017) 80:516-37. doi: 10.1016/j.neubiorev.2017.07.001.
- 279 11. Yapple ZA, Stevens WD, Arsalidou M. Meta-analyses of the n-back working memory task:
280 fMRI evidence of age-related changes in prefrontal cortex involvement across the adult lifespan.
281 *NeuroImage* (2019) 196:16-31. doi: 10.1016/j.neuroimage.2019.03.074.
- 282 12. Van Dinteren R, Arns M, Jongsma MLA, Kessels RPC. P300 Development across the
283 Lifespan: A Systematic Review and Meta-Analysis. *PLoS ONE* (2014) 9(2):e87347. doi:
284 10.1371/journal.pone.0087347.
- 285 13. Lubitz AF, Niedeggen M, Feser M. Aging and working memory performance:
286 Electrophysiological correlates of high and low performing elderly. *Neuropsychologia* (2017)
287 106:42-51. Epub 2017/09/12. doi: 10.1016/j.neuropsychologia.2017.09.002. PubMed PMID:
288 28889995.

- 289 14. Saliassi E, Geerligs L, Lorist MM, Maurits NM. The Relationship between P3 Amplitude and
290 Working Memory Performance Differs in Young and Older Adults. (2013) 8(5):e63701. doi:
291 10.1371/journal.pone.0063701.
- 292 15. Fjell AM, Walhovd KB. P300 and neuropsychological tests as measures of aging: Scalp
293 topography and cognitive changes. *Brain Topography* (2001) 14(1):25-40. doi:
294 10.1023/a:1012563605837.
- 295 16. Feinkohl I, Borchers F, Burkhardt S, Krampe H, Kraft A, Speidel S, et al. Stability of
296 neuropsychological test performance in older adults serving as normative controls for a study on
297 postoperative cognitive dysfunction. *BMC Research Notes* (2020) 13(1). doi: 10.1186/s13104-020-
298 4919-3.
- 299 17. Troyer AK, Vandermorris S, Murphy KJ. Intraindividual variability in performance on
300 associative memory tasks is elevated in amnesic mild cognitive impairment. *Neuropsychologia*
301 (2016) 90:110-6. Epub 2016/06/15. doi: 10.1016/j.neuropsychologia.2016.06.011. PubMed PMID:
302 27297728.
- 303 18. Robertson S, Myerson J, Hale S. Are There Age Differences in Intraindividual Variability in
304 Working Memory Performance? *The Journals of Gerontology Series B: Psychological Sciences and*
305 *Social Sciences* (2006) 61(1):P18-P24. doi: 10.1093/geronb/61.1.p18.
- 306 19. Sandman CA, Patterson JV. The auditory event-related potential is a stable and reliable
307 measure in elderly subjects over a 3 year period. *Clinical Neurophysiology* (2000) 111(8):1427-37.
308 doi: 10.1016/s1388-2457(00)00320-5.
- 309 20. Walhovd KB, Fjell AM. One-year test-retest reliability of auditory ERPs in young and old
310 adults. *Int J Psychophysiol* (2002) 46(1):29-40. Epub 2002/10/11. doi: 10.1016/s0167-
311 8760(02)00039-9. PubMed PMID: 12374644.
- 312 21. Behforuzi H, Feng NC, Billig AR, Ryan E, Tusch ES, Holcomb PJ, et al. Markers of Novelty
313 Processing in Older Adults Are Stable and Reliable. *Frontiers in Aging Neuroscience* (2019) 11. doi:
314 10.3389/fnagi.2019.00165.
- 315 22. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The
316 Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am*
317 *Geriatr Soc* (2005) 53(4):695-9. doi: 10.1111/j.1532-5415.2005.53221.x. PubMed PMID:
318 WOS:000227899200021.
- 319 23. Bopp KL, Verhaeghen P. Aging and n-Back Performance: A Meta-Analysis. *The Journals of*
320 *Gerontology: Series B* (2018). doi: 10.1093/geronb/gby024.
- 321 24. Gevins A, Smith ME, McEvoy LK, Ilan AB, Chan CS, Jiang A, et al. A cognitive and
322 neurophysiological test of change from an individual's baseline. *Clinical Neurophysiology* (2011)
323 122(1):114-20. doi: 10.1016/j.clinph.2010.06.010.
- 324 25. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG
325 dynamics including independent component analysis. *J Neurosci Methods* (2004) 134(1):9-21. Epub
326 2004/04/23. doi: 10.1016/j.jneumeth.2003.10.009. PubMed PMID: 15102499.
- 327 26. Lopez-Calderon J, Luck SJ. ERPLAB: an open-source toolbox for the analysis of event-
328 related potentials. *Frontiers in Human Neuroscience* (2014) 8(213). doi: 10.3389/fnhum.2014.00213.
- 329 27. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*
330 (1979) 86(2):420-8. Epub 1979/03/01. doi: 10.1037//0033-2909.86.2.420. PubMed PMID: 18839484.

- 331 28. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized
332 assessment instruments in psychology. *Psychological assessment* (1994) 6(4):284.
- 333 29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of
334 clinical measurement. *Lancet (London, England)* (1986) 1(8476):307-10. Epub 1986/02/08. PubMed
335 PMID: 2868172.
- 336 30. Cassidy SM, Robertson IH, O'Connell RG. Retest reliability of event-related potentials:
337 Evidence from a variety of paradigms. *Psychophysiology* (2012) 49(5):659-64. doi: 10.1111/j.1469-
338 8986.2011.01349.x.
- 339 31. Kinoshita S, Inoue M, Maeda H, Nakamura J, Morita K. Long-term patterns of change in
340 ERPs across repeated measurements. *Physiol Behav* (1996) 60(4):1087-92. Epub 1996/10/01. doi:
341 10.1016/0031-9384(96)00130-8. PubMed PMID: 8884937.
- 342 32. Segalowitz SJ, Barnes KL. The reliability of ERP components in the auditory oddball
343 paradigm. *Psychophysiology* (1993) 30(5):451-9. Epub 1993/09/01. doi: 10.1111/j.1469-
344 8986.1993.tb02068.x. PubMed PMID: 8416071.
- 345 33. Brunner JF, Hansen TI, Olsen A, Skandsen T, Haberg A, Kropotov J. Long-term test-retest
346 reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo
347 wave in a visual Go/NoGo task. *Int J Psychophysiol* (2013) 89(1):106-14. Epub 2013/06/19. doi:
348 10.1016/j.ijpsycho.2013.06.005. PubMed PMID: 23770084.
- 349 34. Huffmeijer R, Bakermans-Kranenburg MJ, Alink LR, van Ijzendoorn MH. Reliability of
350 event-related potentials: the influence of number of trials and electrodes. *Physiol Behav* (2014)
351 130:13-22. Epub 2014/03/20. doi: 10.1016/j.physbeh.2014.03.008. PubMed PMID: 24642000.
- 352 35. Lew HL, Gray M, Poole JH. Temporal Stability of Auditory Event-Related Potentials in
353 Healthy Individuals and Patients With Traumatic Brain Injury. *Journal of Clinical Neurophysiology*
354 (2007) 24(5):392-7. doi: 10.1097/wnp.0b013e31814a56e3.
- 355 36. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Measures
356 of agreement. *Perspect Clin Res* (2017) 8(4):187-91. doi: 10.4103/picr.PICR_123_17. PubMed
357 PMID: 29109937.
- 358 37. Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, et al. Imaging brain
359 amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol* (2004) 55(3):306-19.
360 Epub 2004/03/03. doi: 10.1002/ana.20009. PubMed PMID: 14991808.
- 361 38. Hedges D, Janis R, Mickelson S, Keith C, Bennett D, Brown BL. P300 Amplitude in
362 Alzheimer's Disease. *Clinical EEG and Neuroscience* (2016) 47(1):48-55. doi:
363 10.1177/1550059414550567.
- 364 39. Chapman RM, McCrary JW, Gardner MN, Sandoval TC, Guillily MD, Reilly LA, et al. Brain
365 ERP components predict which individuals progress to Alzheimer's disease and which do not.
366 (2011) 32(10):1742-55. doi: 10.1016/j.neurobiolaging.2009.11.010.
- 367 40. Pedroso RV, Fraga FJ, Corazza DI, Andreatto CAA, Coelho FGDM, Costa JLR, et al.
368 Latência e amplitude do P300 auditivo na doença de Alzheimer: uma revisão sistemática. *Brazilian*
369 *Journal of Otorhinolaryngology* (2012) 78(4):126-32. doi: 10.1590/s1808-86942012000400023.
- 370 41. Lee M-S, Lee S-H, Moon E-O, Moon Y-J, Kim S, Kim S-H, et al. Neuropsychological
371 correlates of the P300 in patients with Alzheimer's disease. *Progress in Neuro-Psychopharmacology*
372 *and Biological Psychiatry* (2013) 40:62-9. doi: 10.1016/j.pnpbp.2012.08.009.

373 42. Bennys K, Portet F, Touchon J, Rondouin G. Diagnostic value of event-related evoked
 374 potentials N200 and P300 subcomponents in early diagnosis of Alzheimer's disease and mild
 375 cognitive impairment. *J Clin Neurophysiol* (2007) 24(5):405-12. Epub 2007/10/04. doi:
 376 10.1097/WNP.0b013e31815068d5. PubMed PMID: 17912065.

377 43. Rossini PM, Di Iorio R, Vecchio F, Anfossi M, Babiloni C, Bozzali M, et al. Early diagnosis
 378 of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. Report from
 379 the IFCN-sponsored panel of experts. *Clinical Neurophysiology* (2020) 131(6):1287-310. doi:
 380 10.1016/j.clinph.2020.03.003.

381 **Tables**

Table 1. Comparison of task effect (target – nontarget) P3 response at baseline and two-week follow-up.

| Variable | Baseline | Follow-up | Pearson r | ICC, (95% CI) |
|---------------------------------|----------------|----------------|-------------------|----------------------------------|
| 0-back, Fz amplitude (μ V) | 4.87 (3.23) | 5.02 (3.13) | 0.63 ^a | 0.82 (0.67 – 0.90) ^a |
| 0-back, Fz latency (ms) | 291.49 (49.98) | 293.31 (40.26) | 0.38 ^b | 0.54 (0.13 – 0.76) ^b |
| 0-back, Cz amplitude (μ V) | 2.71 (2.46) | 2.59 (2.20) | 0.56 ^a | 0.73 (0.48 – 0.85) ^a |
| 0-back, Cz latency (ms) | 289.49 (55.06) | 290.03 (46.57) | 0.30 ^b | 0.51 (0.07 – 0.74) ^a |
| 0-back, Pz amplitude (μ V) | 1.94 (1.47) | 2.01 (2.00) | 0.38 ^b | 0.54 (0.11 – 0.76) ^b |
| 0-back, Pz latency (ms) | 309.74 (62.36) | 315.92 (63.96) | 0.34 ^b | 0.52 (0.08 – 0.75) ^b |
| 0-back, F3 amplitude (μ V) | 3.99 (2.80) | 4.29 (2.41) | 0.58 ^a | 0.74 (0.49 – 0.86) ^a |
| 0-back, F3 latency (ms) | 297.33 (45.51) | 295.51 (38.37) | 0.31 ^b | 0.47 (-0.02 – 0.72) ^b |
| 0-back, F4 amplitude (μ V) | 4.29 (2.93) | 4.44 (2.72) | 0.62 ^a | 0.77 (0.56 – 0.88) ^a |
| 0-back, F4 latency (ms) | 300.54 (52.41) | 297.95 (35.36) | 0.31 | 0.45 (-0.05 – 0.71) ^b |
| 1-back, Fz amplitude (μ V) | 3.97 (3.27) | 4.08 (3.76) | 0.78 ^a | 0.87 (0.76 – 0.93) ^a |
| 1-back, Fz latency (ms) | 300.90 (38.45) | 300.26 (45.77) | 0.31 | 0.47 (-0.02 – 0.72) ^a |
| 1-back, Cz amplitude (μ V) | 3.48 (3.05) | 3.56 (3.02) | 0.79 ^a | 0.86 (0.74 – 0.92) ^a |
| 1-back, Cz latency (ms) | 302.11 (37.55) | 300.03 (42.39) | 0.33 | 0.44 (-0.04 – 0.70) ^a |
| 1-back, Pz amplitude (μ V) | 1.64 (1.51) | 1.54 (1.70) | 0.58 ^a | 0.73 (0.48 – 0.86) ^a |
| 1-back, Pz latency (ms) | 318.44 (66.66) | 309.97 (64.17) | 0.26 | 0.42 (-0.11 – 0.69) ^b |
| 1-back, F3 amplitude (μ V) | 3.30 (3.43) | 3.42 (3.47) | 0.76 ^a | 0.87 (0.74 – 0.93) ^a |
| 1-back, F3 latency (ms) | 301.92 (44.80) | 300.67 (49.86) | 0.38 ^b | 0.55 (0.14 – 0.77) ^b |
| 1-back, F4 amplitude (μ V) | 3.51 (2.95) | 3.46 (3.37) | 0.68 ^a | 0.81 (0.64 – 0.90) ^a |
| 1-back, F4 latency (ms) | 304.44 (43.16) | 311.03 (45.42) | 0.59 ^a | 0.75 (0.51 – 0.87) ^a |
| 2-back, Fz amplitude (μ V) | 3.39 (2.29) | 3.21 (2.15) | 0.36 ^b | 0.53 (0.09 – 0.75) ^b |
| 2-back, Fz latency (ms) | 300.97 (39.67) | 303.68 (45.65) | 0.09 | 0.17 (-0.60 – 0.49) |
| 2-back, Cz amplitude (μ V) | 2.28 (1.26) | 2.08 (1.65) | 0.39 ^b | 0.51 (0.12 – 0.78) ^b |
| 2-back, Cz latency (ms) | 300.63 (49.62) | 302.21 (45.84) | 0.31 | 0.46 (-0.04 – 0.70) |
| 2-back, Pz amplitude (μ V) | 1.24 (1.20) | 1.38 (1.83) | -0.03 | -0.06 (-1.04 – 0.44) |

| | | | | |
|---------------------------------|----------------|----------------|-------------------|----------------------------------|
| 2-back, Pz latency (ms) | 318.74 (66.70) | 321.63 (63.18) | 0.08 | 0.15 (-0.63 – 0.56) |
| 2-back, F3 amplitude (μ V) | 2.89 (2.03) | 2.30 (1.65) | 0.38 ^b | 0.54 (0.11 – 0.76) ^b |
| 2-back, F3 latency (ms) | 297.29 (47.56) | 292.89 (46.84) | 0.30 | 0.47 (-0.03 – 0.72) ^b |
| 2-back, F4 amplitude (μ V) | 3.37 (2.32) | 3.23 (1.89) | 0.46 ^b | 0.63 (0.28 – 0.81) ^a |
| 2-back, F4 latency (ms) | 306.26 (44.99) | 308.92 (43.12) | 0.09 | 0.16 (-0.61 – 0.56) |

Abbreviations: ICC, Intraclass correlation coefficient; CI, confidence interval

^ap<0.0001

^bp<0.05

382 **Ethics Statement**

383 The study was carried out in accordance with the Declaration of Helsinki. The study was approved by
384 the University of Kansas Medical Center Internal Review Board. Informed consent was obtained
385 from each participant. Each participant received \$100 for participating.

386 **Conflict of Interest**

387 The authors declare that the research was conducted in the absence of any commercial or financial
388 relationships that could be construed as a potential conflict of interest.

389 **Author Contributions**

390 HD, JB, JM, WMB, and KG conceptualized the study. HD, KL, and KG worked out the EEG data
391 processing steps. HD, PA, and KL administered the tests. HD and JM analysed the data. HD wrote
392 the initial manuscript. JB, KL, PA, JM, WMB, and KG reviewed the manuscript and provided
393 valuable comments.

394 **Funding**

395 Research reported in this publication was supported by the National Institute on Aging of the
396 National Institutes of Health under Award Number K01 AG058785. This study was supported in part
397 by a pilot grant of the KU Alzheimer Disease Center (P30 AG035982). The Hoglund Biomedical

398 Imaging Center is supported in part by S10 RR29577 and generous gifts from Forrest and Sally
399 Hoglund. The content is solely the responsibility of the authors and does not necessarily represent the
400 official views of the National Institutes of Health.

401 **Acknowledgments**

402 The authors thank the volunteers for their time and willingness to participate in this research. We are
403 also grateful for the staff at the KU Alzheimer Disease Center.

404 **Figure Legends**

405 Figure 1. Grand Average Event-related Potential Waveform of (a) 0-back, (b) 1-back and (c) 2-back.

406 Figure 2. Bland Altman Plots of (a) 0-back Fz Peak Amplitude (b) 1-back Fz Peak Amplitude; (c) 2-
407 back Fz Peak Amplitude; (d) 0-back Fz Peak Latency; (e) 1-back Peak Latency; (f) 2-back Peak
408 Latency.



