

Northumbria Research Link

Citation: Zheng, Bang, Udeh-Momoh, Chinedu, Watermeyer, Tamlyn, de Jager Loots, Celeste A., Ford, Jamie K., Robb, Catherine E., Giannakopoulou, Parthenia, Ahmadi-Abhari, Sara, Baker, Susan, Novak, Gerald P., Price, Geraint and Middleton, Lefkos T. (2022) Practice Effect of Repeated Cognitive Tests Among Older Adults: Associations With Brain Amyloid Pathology and Other Influencing Factors. *Frontiers in Aging Neuroscience*, 14. p. 909614. ISSN 1663-4365

Published by: Frontiers Media

URL: <https://doi.org/10.3389/fnagi.2022.909614>
<<https://doi.org/10.3389/fnagi.2022.909614>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/49496/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

1 **Practice effect of repeated cognitive tests among older adults:**
2 **associations with brain amyloid pathology and other influencing**
3 **factors**

4 **Bang Zheng^{1,2}, Chinedu Udeh-Momoh¹, Tamlyn Watermeyer³, Celeste A de Jager Loots¹,**
5 **Jamie K Ford¹, Catherine E Robb¹, Parthenia Giannakopoulou¹, Sara Ahmadi-Abhari¹, Susan**
6 **Baker⁴, Gerald P Novak⁴, Geraint Price^{1*}, Lefkos T Middleton^{1,5}.**

7 ¹Ageing Epidemiology (AGE) Research Unit, School of Public Health, Imperial College London,
8 London, UK

9 ²Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical
10 Medicine, London, UK

11 ³Edinburgh Dementia Prevention, Centre for Clinical Brain Sciences, University of Edinburgh,
12 Edinburgh, UK

13 ⁴Janssen Research and Development LLC, Titusville NJ, USA

14 ⁵Public Health Directorate, Imperial College NHS Healthcare Trust, London, UK

15 *** Correspondence:**

16 Dr. Geraint Price

17 g.price@imperial.ac.uk

18 **Keywords: practice effect, cognitive test, older adults, amyloid pathology, memory**

19 **Abstract**

20 **Background:** Practice effects (PE), after repeated cognitive measurements, may mask cognitive
21 decline and represent a challenge in clinical and research settings. However, an attenuated practice
22 effect may indicate the presence of brain pathologies. This study aimed to evaluate practice effects on
23 the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) scale, and their
24 associations with brain amyloid status and other factors in a cohort of cognitively unimpaired older
25 adults enrolled in the CHARIOT-PRO SubStudy.

26 **Methods:** 502 cognitively unimpaired participants aged 60-85 years were assessed with RBANS in
27 both screening and baseline clinic visits using alternate versions (median time gap of 3.5 months).
28 We tested PE based on differences between test and retest scores in total scale and domain-specific
29 indices. Multiple linear regressions were used to examine factors influencing PE, after adjusting for
30 age, sex, education level, *APOE*- ϵ 4 carriage and initial RBANS score. The latter and PE were also
31 evaluated as predictors for amyloid positivity status based on defined thresholds, using logistic
32 regression.

33 **Results:** Participants' total scale, immediate memory and delayed memory indices were significantly
34 higher in the second test than in the initial test (Cohen's $d_z = 0.48, 0.70$ and $0.35, P < 0.001$). On the
35 immediate memory index, the PE was significantly lower in the amyloid positive group than the
36 amyloid negative group ($P = 0.022$). Older participants (≥ 70 years), women, non-*APOE*- ϵ 4 carriers,
37 and those with worse initial RBANS test performance had larger PE. No associations were found

38 between brain MRI parameters and PE. In addition, attenuated practice effects in immediate or
39 delayed memory index were independent predictors for amyloid positivity ($P < 0.05$).

40 **Conclusion:** Significant practice effects on RBANS total scale and memory indices were identified
41 in cognitively unimpaired older adults. The association with amyloid status suggests that practice
42 effects are not simply a source of measurement error but may be informative with regard to
43 underlying neuropathology.

44 1 Introduction

45 Valid instruments and implementations of cognitive tests are essential for the evaluation of cognitive
46 status, decline and subsequent dementia diagnosis, and the screening of at-risk participants for
47 clinical trials and population intervention programs for dementia prevention. However, practice
48 effects (PE) after repeated cognitive measurements, which refer to improvements in test performance
49 due to repeated exposure to test materials or procedures (Hausknecht et al., 2007; Goldberg et al.,
50 2015), often mask a potential cognitive decline and remain a major issue in clinical and research
51 settings (Houx et al., 2002; Sanderson-Cimino et al., 2022). Failing to account for practice effects in
52 cognitive tests could delay diagnosis and clinical care for patients with cognitive deficits. PE
53 resulting from task familiarity occurring with test repetition is distinct from learning effects which
54 refer to the recall of correct answers from previous tests. The latter is often extenuated in
55 neuropsychological practice through administration of alternate versions of the same task (e.g.,
56 different word lists in verbal memory tests).

57 Exploring factors that influence practice effects can be informative of potential heterogeneity of
58 measurement bias and in developing mitigation strategies to minimise such bias (Calamia et al.,
59 2012). On the other hand, the magnitude of practice effect per se may also have indicative value for
60 cognitive impairment or existing brain pathologies (Duff et al., 2007; Jutten et al., 2021). From this
61 perspective, PE may represent not merely a source of measurement error but potentially valuable
62 information from a clinical and scientific perspective (Duff et al., 2007).

63 Given the long preclinical stage of late-onset dementia (Elias et al., 2000) with progressively
64 accumulating neuropathology, it is early detection in at-risk individuals that may prove essential in
65 reducing the burden of cognitive and functional decline and dementia in the elderly population.
66 Therefore, a deeper understanding and characterisation of PE in validated cognitive assessment tools
67 among asymptomatic population is warranted.

68 This study aimed to evaluate PE in the Repeatable Battery for the Assessment of Neuropsychological
69 Status (RBANS) (Randolph et al., 1998), and its associations with brain amyloid status and other
70 factors in a cohort of cognitively unimpaired older adults in the UK Cognitive Health in Ageing
71 Register: Investigational, Observational, and Trial Studies in Dementia Research: Prospective
72 Readiness cOhort Study (CHARIOT-PRO) SubStudy (Udeh-Momoh et al., 2021).

73 2 Methods

74 2.1 Study population

75 CHARIOT-PRO SubStudy is an on-going prospective cohort study of cognitively unimpaired older
76 adults in the UK, which aims to examine longitudinal cognitive changes in those with and without
77 brain amyloid-beta ($A\beta_{42}$) pathology, and factors and markers of subsequent decline (Udeh-Momoh
78 et al., 2021). Following screening of 2425 individuals, including amyloid status determination and

79 multiple cognitive tests, an equal number of participants above and below a binary threshold of A β 42
 80 positivity were enrolled at baseline and in subsequent longitudinal study. During screening,
 81 participants whose performance on any RBANS index fell >1.5 standard deviation (SD) below the
 82 population mean (population norms from Randolph, 1998) were referred to an adjudication panel of
 83 neurologists, psychiatrists and neuropsychologists to detect any undiagnosed cognitive impairment
 84 which was an exclusion criterion. The detailed inclusion/ exclusion criteria and study procedures
 85 have been described in previous papers of our group (Nalder et al., 2021; Udeh-Momoh et al., 2021).
 86 The study received approval from the National Research Ethics Service (NRES) Committee London
 87 Central (reference 15/LO/0711 (IRAS 140764)), as well as independent ethics review by committees
 88 from the local sites. All participants provided informed consent before participating in the study.

89 A total of 502 participants aged 60-85 years completed RBANS assessments in both screening and
 90 baseline clinic visits and were included in this study (Udeh-Momoh et al., 2021). The median time
 91 gap between the screening visit and the baseline visit was 3.5 months, which allowed us to examine
 92 the practice effects in RBANS scale within a relatively short time period with less concern that the
 93 test-retest score differences are (partially) due to the cognitive decline during this time interval.

94 2.2 Measurements

95 Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) (Randolph et al.,
 96 1998) is a validated and widely-used neuropsychological assessment. It is a 20-minute composite
 97 battery which consists of twelve subtests that measure five cognitive domain indices (immediate
 98 memory, delayed memory, visuospatial construction, language, attention). The sum of the five index
 99 scores is converted to a total scale score based on a distribution with a mean of 100 and SD of 15.
 100 This assessment was administered by trained assistant psychologists during the in-person clinic
 101 visits. Version C and Version A of the RBANS were administered at the screening and baseline
 102 assessments, respectively, to avoid learning effects (i.e., recalling answers from the same test
 103 received before).

104 Amyloid burden was determined during the screening visit either by amyloid positron emission
 105 tomography (PET) scans (in ~90% of participants) or cerebrospinal fluid (CSF) A β 42 measurements
 106 via lumbar punctures (in the remaining 10%). A β positive was defined as above-threshold brain A β
 107 deposition on PET (based on tracer-specific thresholds of the composite cortical standardised uptake
 108 value ratio, SUVR) or below-threshold CSF A β 42 concentration (\leq 600 ng/L). Three F18-
 109 radiolabeled amyloid tracers were used: florbetapir (Amyvid), flutemetamol (Vizamyl) and
 110 florbetaben (Neuraceq). The composite cortical SUVR threshold was 1.14 for Amyvid and 1.23 for
 111 Vizamyl (both with whole cerebellum as reference region), and 1.20 for Neuraceq (with cerebellar
 112 grey matter as reference region) (Udeh-Momoh et al., 2021).

113 Screening also included a brain magnetic resonance imaging (MRI). Bilateral volumetric MRI
 114 parameters were obtained, including whole brain volume (mL³), ventricular volume (mL³),
 115 hippocampal volume (mm³) and AD signature cortical thickness (mm) (Schwarz et al., 2016).
 116 Intracranial volume (ICV) was used as the proxy variable for premorbid brain volume to be adjusted
 117 for in the analyses of MRI parameters. All study procedures and cut-off points have previously been
 118 reported (Udeh-Momoh et al., 2021).

119 We also collected other information including age, sex, ethnicity, education level, *APOE* genotype
 120 and National Adult Reading Test (NART) score (as a proxy for premorbid intelligence quotient (IQ))
 121 (Nelson, 1991).

122 2.3 Statistical analyses

123 Demographic and clinical characteristics of study participants were compared according to amyloid
 124 pathology status (amyloid positive vs. negative) using independent samples *t*-test, chi-squared test,
 125 rank-sum test or general linear regression, where appropriate. We assessed the internal consistency
 126 reliability (Cronbach's α coefficient) and test-retest reliability (Pearson correlation coefficient *r*) of
 127 the RBANS scale in this cohort. PE was estimated based on differences between test and retest scores
 128 (i.e., measurements at the screening and baseline visits) in RBANS total scale and domain-specific
 129 indices. Paired *t* test was used to test the statistical significance of PE; Cohen's d_z for the within-
 130 subjects design (Cohen, 1988) was calculated as the standardised effect size for PE (i.e., scaled
 131 difference scores).

132 Multiple linear regression model was used to examine whether the magnitude of PE varies by
 133 amyloid status, with the test-retest difference score in RBANS total scale or domain-specific index as
 134 the dependent variable, amyloid status as the independent variable of interest, while adjusting for age,
 135 sex, education level, *APOE-ε4* carriage and initial RBANS level. Following the same procedure, we
 136 also explored other potential influencing factors of PE in separate linear regression models, including
 137 age group (60-69 years vs. 70-85 years), sex, education level (below/above upper secondary
 138 education), *APOE-ε4* (carrier vs. non-carrier), test-retest time interval (1-3 months vs. 4-6 months),
 139 MRI parameters (below/above mean), National Adult Reading Test score (below/above median), and
 140 initial RBANS scores (below/above mean).

141 To assess the robustness of our main findings, we conducted the following sensitivity analyses: (1)
 142 modelling MRI parameters, age, test-retest time interval, initial RBANS score and NART score as
 143 continuous variables instead of dichotomised variables when exploring their associations with PE; (2)
 144 excluding 52 participants who waited for over 6 months after the screening visit to attend the baseline
 145 visit to avoid the loss of PE or occurrence of possible cognitive decline during the prolonged time
 146 gap; (3) additionally adjusting for test-retest time interval and modality of amyloid (PET or CSF)
 147 when assessing the amyloid-PE association.

148 Finally, to explore the predictive value of PE, PE was also assessed as a predictor together with
 149 initial RBANS score for amyloid positive status using binary logistic regression, adjusting for age,
 150 sex, education level, and *APOE-ε4* carriage. The odds ratio (OR) and 95% confidence interval (CI) of
 151 standardised PE scores (i.e., centred and scaled) was reported, which reflects the relative risk of the
 152 presence of amyloid pathology per 1 SD increase in PE.

153 Statistical analyses were conducted using Stata (version 15; College Station, TX: StataCorp LLC).
 154 All statistical analyses are two-sided. A *P* value of < 0.05 indicates a statistically significant result.

155 3 Results

156 3.1 Population characteristics

157 Of the 502 participants assessed with RBANS scale in both screening and baseline clinic visits with
 158 median time gap of 3.5 months (interquartile range: 2.9 - 4.4), the mean (SD) age was 71.4 (5.5)
 159 years, and 254 (50.6%) were females. 192 participants (38.2%) were *APOE-ε4* carriers and 247
 160 (49.2%) were A β positive based on CSF A β 42 level or PET scans. Nearly all participants (95.8%)
 161 were White. Most participants (85.7%) had completed upper secondary education or above.

Participant characteristics are presented by amyloid pathology status in Table 1. A β + participants were slightly older and more likely to be *APOE*- ϵ 4 carriers compared with A β - participants ($P < 0.05$). Differences in MRI parameters were also observed between amyloid groups, with A β + group having lower hippocampal volume, whole brain volume, and AD signature cortical thickness ($P < 0.05$). The RBANS test-retest time interval was similar between A β + group and A β - group ($P = 0.728$).

3.2 Practice effects in RBANS assessment

The internal consistency reliability of RBANS scale in our study sample measured by Cronbach's α was 0.64, and the test-retest reliability measured by Pearson correlation coefficient r was 0.79.

Participants had significantly higher scores in RBANS total scale and immediate and delayed memory indices in the second test than in the initial test (increased score = 3.9, 7.6 and 3.3, respectively; $P < 0.001$; Table 2). After taking into account the differences in variances of these indices, the calculation of within-subject Cohen's d_z revealed a strong effect size for PE in immediate memory index (0.70), and a low-to-moderate effect size for PE in RBANS total scale (0.48) and delayed memory index (0.35). In contrast, no significant PEs were identified for the rest of the three domain indices (Cohen's d_z ranged from 0.05 to 0.06; $P > 0.05$; Table 2).

3.3 Practice effects in RBANS by amyloid pathology status

We examined the practice effects in RBANS total scale and memory indices by amyloid pathology status (Figure 1). After adjusting for potential confounding factors, the amyloid positive group had significantly lower PE in immediate memory index than the amyloid negative group (Cohen's $d_z = 0.60$ vs. 0.81; $P = 0.022$). Similarly, a borderline statistical significance was observed for lower PE in delayed memory index, in the amyloid positive group (Cohen's $d_z = 0.26$ vs. 0.44; $P = 0.059$). However, the difference in PE in RBANS total scale by amyloid status did not reach statistical significance (Cohen's $d_z = 0.46$ vs. 0.50; $P = 0.387$; Figure 1). We also generated spaghetti plots by amyloid status to visualise the heterogeneity in practice effects across individuals (Supplementary Figures 1-3).

3.4 Other influencing factors on practice effect in RBANS

In the exploratory analyses for brain MRI parameters and PE, we observed no significant associations of hippocampal volume, whole brain volume, ventricular volume or AD signature cortical thickness with the magnitude of PE in RBANS total scale or memory indices (Supplementary Table 1).

Older adults (≥ 70 years), women, and *APOE*- ϵ 4 non-carriers had larger PE in one or more RBANS indices ($P < 0.05$; Table 3). Those with worse performance in the initial RBANS test had larger PE in both total scale and the individual memory indices ($P < 0.05$; Table 3). Test-retest time interval, education level and NART score had no significant association with the magnitude of PE (Supplementary Table 1). Sensitivity analyses revealed consistent results with the main findings (Supplementary Tables 2-5).

3.5 Attenuated practice effect is indicative of above threshold amyloid pathology

We further explored the indicative value of PE for brain amyloid pathology. Results of multiple logistic regressions showed that, besides age (OR = 1.09, 95% CI: 1.05-1.13 per year) and *APOE*- ϵ 4 carriage (OR = 5.50, 95% CI: 3.60-8.40), worse initial performance and lower PE in delayed memory

203 index were independent predictors for amyloid positivity, with similar magnitudes of association (OR
 204 per 1 SD increase = 0.78, 95% CI: 0.63-0.97). As for immediate memory, lower PE (OR = 0.75, 95%
 205 CI: 0.61-0.94) but not performance in the initial test (OR = 0.82, 95% CI: 0.66-1.02) was a
 206 significant predictor for amyloid positivity. We did not find an association between PE in RBANS
 207 total scale and existing amyloid pathology (OR = 0.92, 95% CI: 0.75-1.12).

208 4 Discussion

209 In this prospective cohort study of cognitively unimpaired older adults, enriched with fluid and
 210 neuroimaging biomarker data, we comprehensively assessed the practice effect in RBANS
 211 assessment and its potential influencing factors, with a focus on brain amyloid pathology. We
 212 observed significant practice effects for RBANS total scale and two memory indices, where
 213 participants performed better after repeated measurement using alternate versions of these tasks. The
 214 magnitude of practice effects differed by amyloid pathology status, age, sex, *APOE*- ϵ 4 carriage and
 215 initial RBANS scores, but had no association with brain MRI parameters, education level or NART
 216 score.

217 Our findings suggest that PE in cognitive tests may be domain-specific. Of the five cognitive
 218 domains assessed by RBANS scale, only the two memory indices presented significant practice
 219 effects, whilst participants' performance in visuospatial construction, language and attention domains
 220 remained similar between the first and second tests over a median of 3.5 months. Our results were in
 221 line with a previous study of a much smaller sample of 36 healthy adults (Bartels et al., 2010), where
 222 clinically relevant PE was observed during high-frequency testing within three months in learning
 223 and memory tests but not in language and visuospatial tests. Similarly, a study of 947 cognitively
 224 normal older adults from the Mayo Clinic Study of Aging showed large PE in learning and memory
 225 tests but low PE in language tests, using the Mayo Clinic neurocognitive battery (Machulda et al.,
 226 2013).

227 Regarding the memory domain indices, we observed a much larger effect size of PE for immediate
 228 memory index than that for delayed memory index or the RBANS total scale. This implied that PE
 229 may be more pronounced in immediate memory tasks where people tend to get better at doing these
 230 tasks following familiarisation with the test materials or procedures, even when assessed with
 231 different word lists (Houx et al., 2002). Thus, the immediate memory test seems to be a more
 232 sensitive measure of PE, compared with other domains or the global composite score. The contrast
 233 between immediate and delayed memory PEs might alternatively reflect differences in the content of
 234 the measures. Specifically, the RBANS immediate memory index is derived solely from tests of
 235 verbal recall, whereas the delayed memory index also incorporates verbal recognition and visual-
 236 constructional recall. Future systematic evaluation of practice effects in individual test scores rather
 237 than the overall indices, with larger sample size and careful control of multiple testing, may help
 238 identify even more sensitive metrics.

239 Our data are in line with previous reports, suggesting the predictive value of PE for the presence of
 240 amyloid pathology and subsequent cognitive decline, in addition to merely evaluating cognitive
 241 measurement. To be noted, on average, the RBANS scores in our study participants were within
 242 "cognitively healthy" boundaries, even in the amyloid positive group and would not prompt further
 243 testing in a clinical scenario. This observation underscores the potential value of diminished practice
 244 effects as an adjunct metric to traditional assessments for the sensitive detection of preclinical AD.
 245 Several previous studies have consistently shown that diminished PE over repeated cognitive testing
 246 (mainly episodic memory measures) was associated with subsequent cognitive decline and increased

247 risk of mild cognitive impairment (MCI) or dementia (Duff et al., 2007; Sanchez-Benavides et al.,
248 2016; Jutten et al., 2020; Jutten et al., 2021). In contrast, previous evidence on the association
249 between PE and AD biomarkers and neuropathology remained inconsistent (Duff et al., 2018; Ihara
250 et al., 2018; Jutten et al., 2020). A previous systematic review on PE in cognitive assessment
251 identified four papers reporting an association between higher amyloid uptake on amyloid PET scans
252 and lower PE, whereas two papers did not detect this association (Jutten et al., 2020). In our study,
253 the attenuated PE in memory indices was associated with the presence of high amyloid burden but
254 not with brain MRI features, including hippocampal volume, implying that PE in memory tests could
255 be more indicative of β -amyloidosis (which is specific for Alzheimer's disease (AD)) instead of
256 biomarkers of neurodegeneration or neuronal injury (Jack et al., 2016). Consistent with our results, a
257 recent report from the Harvard Aging Brain Study, of 114 cognitively unimpaired older adults,
258 showed that lower PE in a self-administered computerised cognitive composite battery over the first
259 3 months was associated with more global amyloid burden (based on PiB-PET imaging) and tau
260 deposition in the entorhinal cortex and inferior-temporal lobe (based on Flortaucipir PET imaging)
261 (Jutten et al., 2021). These findings imply the usefulness of PE as an early detection tool for signs of
262 disease burden prior to the emergence of cognitive impairment, which might inform participant
263 stratification and biomarker testing strategies for clinical trials.

264 In our exploratory analyses, practice effects in RBANS total scale or memory indices were more
265 pronounced in older adults, women, *APOE*- ϵ 4 non-carriers and those with worse performance in the
266 initial RBANS assessment (probably due to larger space for improvement). Of note, these factors
267 were associated with different indices, indicating a complex domain-specific PE population
268 heterogeneity. Our finding of a positive association between age and PE was inconsistent with a
269 previous meta-analysis report (Calamia et al., 2012) of a negative association, in a much younger
270 population (mean age of around 40 to 50 years). In the afore-mentioned Mayo Clinic report
271 (Machulda et al., 2013), no significant PE differences were found on memory test scores between
272 those aged below and above 80 years. A previous systematic review identified three papers reporting
273 an association between presence of ≥ 1 *APOE*- ϵ 4 allele and lower PE, whereas three papers did not
274 detect this association (Jutten et al., 2020). Further studies are warranted to elucidate the nature and
275 extent of these population heterogeneities in PE, which could be crucial for clinical trials in obtaining
276 unbiased effect estimate for tested treatment or intervention. If the factors affecting PE are not well
277 balanced between placebo and treatment groups, the two groups may have different levels of PE, in
278 which case researchers need to control for these factors so that the estimate of difference in cognitive
279 outcomes between groups can be attributed to treatment.

280 The availability of extensive phenotypic (including fluid and neuroimaging biomarker) data is a key
281 strength of our study. Moreover, the relatively short test-retest interval (median of 3.5 months) was
282 essential in minimising the risk of a potential cognitive decline during the test-retest interval affecting
283 the presence and extent of PE. If given a long test-retest period, PE may be masked by progressive
284 cognitive decline over time and it would be difficult to distinguish one from the other.

285 Several limitations need to be taken into consideration when interpreting our results. Since we
286 explored multiple influencing factors on PE in our study, the risk of inflated Type 1 error in multiple
287 testing cannot be ruled out. Therefore, our exploratory analyses need further validation. Moreover,
288 RBANS does not provide an isolated scale of executive function, a domain which has been
289 independently associated with early amyloidosis rather than memory performance decrements in
290 cognitively normal adults (Tideman et al., 2022). Assessing diminished practice effects in this
291 domain may yet provide even more sensitive markers of subtle cognitive signs. Due to the different
292 modalities and tracers used for amyloid testing in this study, we did not evaluate the amyloid

293 pathology on a quantitative scale which is worth to be considered in future studies. In addition, we
294 only used data from two time points; future studies on longitudinal PE across multiple measurements
295 (with short between-test intervals) are needed. For instance, it is worth exploring whether the PE
296 beyond the second test is not as large as that between the first two tests, which may have important
297 implications for research and clinical purposes (e.g., recommending the second assessment to be
298 considered as baseline measure to minimise PE in outcome assessment). Furthermore, since our test-
299 retest time gap mainly fell between 3-4 months, future large-scale studies with time gaps of wider
300 distribution could provide insights for what might be too short vs. too long for detecting PE, though it
301 is possible that the optimal time gap could be different for different cognitive domains or tasks.
302 Finally, our study population are cognitively unimpaired older adults; it would also be interesting to
303 investigate PE in MCI or AD patients, which may show different profiles (Machulda et al., 2013).
304 Similarly, the study sample lacks ethnic and racial diversity (95.8% White people) thereby limiting
305 the generalisability of our findings.

306 In conclusion, we identified significant PE in RBANS total scale and memory indices among a
307 cohort of cognitively unimpaired older adults. PE is not simply a source of measurement bias in
308 cognitive assessment, but may be informative with regard to a significant brain amyloid pathology
309 burden.

310

311 **5 Author Contributions**

312 GP, LTM, CUM, BZ and TW contributed to study design and conception. BZ and CUM carried out
313 data analysis and interpretation. BZ, LTM and CUM drafted the first version of the manuscript. All
314 authors critically reviewed and revised the manuscript.

315 **6 Data Availability Statement**

316 The datasets for this study cannot be made publicly available yet for ethical reasons.

317 **7 Ethics Statement**

318 The study received approval from the National Research Ethics Service (NRES) Committee London
319 Central (reference 15/LO/0711 (IRAS 140764)), as well as independent ethics review by committees
320 from the local sites. All participants provided informed consent before participating in the study.

321 **8 Conflict of Interest**

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

324

325 **9 References**

326 Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., and Ehrenreich, H. (2010). Practice effects in
327 healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci.*
328 11, 118. doi: 10.1186/1471-2202-11-118

- 329 Calamia, M., Markon, K., and Tranel, D. (2012). Scoring higher the second time around: meta-
 330 analyses of practice effects in neuropsychological assessment. *Clin. Neuropsychol.* 26(4),
 331 543-570. doi: 10.1080/13854046.2012.680913
- 332 Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. NJ: Lawrence
 333 Erlbaum Associates, Publishers.
- 334 Duff, K., Anderson, J.S., Mallik, A.K., Suhrie, K.R., Atkinson, T.J., Dalley, B.C.A., et al. (2018).
 335 Short-term repeat cognitive testing and its relationship to hippocampal volumes in older
 336 adults. *J. Clin. Neurosci.* 57, 121-125. doi: 10.1016/j.jocn.2018.08.015
- 337 Duff, K., Beglinger, L.J., Schultz, S.K., Moser, D.J., McCaffrey, R.J., Haase, R.F., et al. (2007).
 338 Practice effects in the prediction of long-term cognitive outcome in three patient samples: a
 339 novel prognostic index. *Arch. Clin. Neuropsychol.* 22(1), 15-24. doi:
 340 10.1016/j.acn.2006.08.013
- 341 Elias, M.F., Beiser, A., Wolf, P.A., Au, R., White, R.F., and D'Agostino, R.B. (2000). The preclinical
 342 phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch.*
 343 *Neurol.* 57(6), 808-813. doi: 10.1001/archneur.57.6.808
- 344 Goldberg, T.E., Harvey, P.D., Wesnes, K.A., Snyder, P.J., and Schneider, L.S. (2015). Practice
 345 effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease
 346 randomized controlled trials. *Alzheimers Dement. (Amst)* 1(1), 103-111. doi:
 347 10.1016/j.dadm.2014.11.003
- 348 Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., and Moriarty Gerrard, M.O. (2007). Retesting in
 349 selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *J.*
 350 *Appl. Psychol.* 92(2), 373-385. doi: 10.1037/0021-9010.92.2.373
- 351 Houx, P.J., Shepherd, J., Blauw, G.J., Murphy, M.B., Ford, I., Bollen, E.L., et al. (2002). Testing
 352 cognitive function in elderly populations: the PROSPER study. PROSpective Study of
 353 Pravastatin in the Elderly at Risk. *J. Neurol. Neurosurg. Psychiatry* 73(4), 385-389. doi:
 354 10.1136/jnnp.73.4.385
- 355 Ihara, R., Iwata, A., Suzuki, K., Ikeuchi, T., Kuwano, R., Iwatsubo, T., et al. (2018). Clinical and
 356 cognitive characteristics of preclinical Alzheimer's disease in the Japanese Alzheimer's
 357 Disease Neuroimaging Initiative cohort. *Alzheimers Dement. (N Y)* 4, 645-651. doi:
 358 10.1016/j.trci.2018.10.004
- 359 Jack, C.R., Jr., Bennett, D.A., Blennow, K., Carrillo, M.C., Feldman, H.H., Frisoni, G.B., et al.
 360 (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease
 361 biomarkers. *Neurology* 87(5), 539-547. doi: 10.1212/WNL.0000000000002923
- 362 Jutten, R.J., Grandoit, E., Foldi, N.S., Sikkes, S.A.M., Jones, R.N., Choi, S.E., et al. (2020). Lower
 363 practice effects as a marker of cognitive performance and dementia risk: A literature review.
 364 *Alzheimers Dement. (Amst)* 12(1), e12055. doi: 10.1002/dad2.12055
- 365 Jutten, R.J., Rentz, D.M., Fu, J.F., Mayblyum, D.V., Amariglio, R.E., Buckley, R.F., et al. (2021).
 366 Monthly At-Home Computerized Cognitive Testing to Detect Diminished Practice Effects in
 367 Preclinical Alzheimer's Disease. *Front. Aging Neurosci.* 13, 800126. doi:
 368 10.3389/fnagi.2021.800126
- 369 Machulda, M.M., Pankratz, V.S., Christianson, T.J., Ivnik, R.J., Mielke, M.M., Roberts, R.O., et al.
 370 (2013). Practice effects and longitudinal cognitive change in normal aging vs. incident mild
 371 cognitive impairment and dementia in the Mayo Clinic Study of Aging. *Clin. Neuropsychol.*
 372 27(8), 1247-1264. doi: 10.1080/13854046.2013.836567
- 373 Nalder, L., Zheng, B., Chiandret, G., Middleton, L.T., and de Jager, C.A. (2021). Vitamin B12 and
 374 Folate Status in Cognitively Healthy Older Adults and Associations with Cognitive
 375 Performance. *J. Nutr. Health Aging* 25(3), 287-294. doi: 10.1007/s12603-020-1489-y
- 376 Nelson, H.E., and Willison, J. (1991). *National adult reading test (NART)*. Windsor: Nfer-Nelson.

- 377 Randolph, C. (1998). *Repeatable battery for the assessment of neuropsychological status (RBANS)*.
 378 San Antonio, TX: Psychological Corporation.
- 379 Randolph, C., Tierney, M.C., Mohr, E., and Chase, T.N. (1998). The Repeatable Battery for the
 380 Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *J. Clin.*
 381 *Exp. Neuropsychol.* 20(3), 310-319. doi: 10.1076/jcen.20.3.310.823
- 382 Sanchez-Benavides, G., Gispert, J.D., Fauria, K., Molinuevo, J.L., and Gramunt, N. (2016).
 383 Modeling practice effects in healthy middle-aged participants of the Alzheimer and Families
 384 parent cohort. *Alzheimers Dement. (Amst)* 4, 149-158. doi: 10.1016/j.dadm.2016.07.001
- 385 Sanderson-Cimino, M., Elman, J.A., and Tu, X. (2022). Practice Effects in Mild Cognitive
 386 Impairment Increase Reversion Rates and Delay Detection of New Impairments. *Front. Aging*
 387 *Neurosci.* In press.
- 388 Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., et al. (2016).
 389 A large-scale comparison of cortical thickness and volume methods for measuring
 390 Alzheimer's disease severity. *Neuroimage Clin.* 11, 802-812. doi: 10.1016/j.nicl.2016.05.017
- 391 Tideman, P., Stomrud, E., Leuzy, A., Mattsson-Carlgen, N., Palmqvist, S., Hansson, O.; Alzheimer's
 392 Disease Neuroimaging Initiative (2022). Association of β -Amyloid Accumulation With
 393 Executive Function in Adults With Unimpaired Cognition. *Neurology* 98(15), e1525-e1533.
 394 doi: 10.1212/WNL.0000000000013299
- 395 Udeh-Momoh, C.T., Watermeyer, T., Price, G., de Jager Loots, C.A., Reglinska-Matveyev, N.,
 396 Ropacki, M., et al. (2021). Protocol of the Cognitive Health in Ageing Register:
 397 Investigational, Observational and Trial Studies in Dementia Research (CHARIOT):
 398 Prospective Readiness cOhort (PRO) SubStudy. *BMJ Open* 11(6), e043114. doi:
 399 10.1136/bmjopen-2020-043114

401 10 Tables and Figures

402 Table 1. Population characteristics by amyloid status (N = 502)

Characteristics	Total	Amyloid positive	Amyloid negative	<i>P</i> value
N	502	247	255	
Age (years), $\bar{x} \pm SD$	71.4 \pm 5.5	72.3 \pm 5.6	70.4 \pm 5.4	< 0.001
Female, %	50.6	48.6	52.6	0.374
Ethnicity (White), %	95.8	96.8	94.9	0.298
Below upper secondary education, %	14.3	17.0	11.8	0.094
<i>APOE</i> - ϵ 4 carrier, %	38.2	54.7	22.4	< 0.001
NART score, $\bar{x} \pm SD$	9.9 \pm 6.7	9.5 \pm 5.9	10.3 \pm 7.3	0.202
Days between test and retest, median (IQR)	107 (87-133)	106 (86-133)	108 (87-136)	0.728
RBANS score (first test), $\bar{x} \pm SD$				
Total scale	102.7 \pm 11.8	102.6 \pm 11.7	102.9 \pm 11.9	0.734
Immediate memory index	101.6 \pm 12.7	101.0 \pm 12.3	102.2 \pm 13.0	0.268
Delayed memory index	100.7 \pm 10.1	99.8 \pm 10.9	101.6 \pm 9.2	0.045
Visuospatial construction index	95.7 \pm 14.1	96.7 \pm 13.9	94.9 \pm 14.3	0.148
Language index	104.1 \pm 11.5	104.5 \pm 11.0	103.7 \pm 12.0	0.422
Attention index	108.8 \pm 14.5	108.5 \pm 13.8	109.2 \pm 15.2	0.607
MRI parameters, $\bar{x} \pm SD$				
Hippocampal volume (mm ³)	7754 \pm 852	7621 \pm 899	7883 \pm 794	<0.001
Whole brain volume (mL ³)	1094629 \pm 107552	1087603 \pm 109462	1101408 \pm 105861	0.005
Ventricular volume (mL ³)	35701 \pm 16987	36381 \pm 16991	35045 \pm 16886	0.304
AD signature cortical thickness (mm)	2.80 \pm 0.12	2.79 \pm 0.13	2.81 \pm 0.12	0.028

403 Note: SD = standard deviation; NART = National Adult Reading Test; IQR = interquartile range;
404 RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; MRI = magnetic
405 resonance imaging; AD = Alzheimer's disease. *P* values were calculated by chi-squared tests, *t* tests,
406 rank-sum test, or general linear regressions to adjust for intracranial volume for volumetric MRI
407 parameters.

408 **Table 2. Differences between test and retest performance in RBANS (N = 502)**

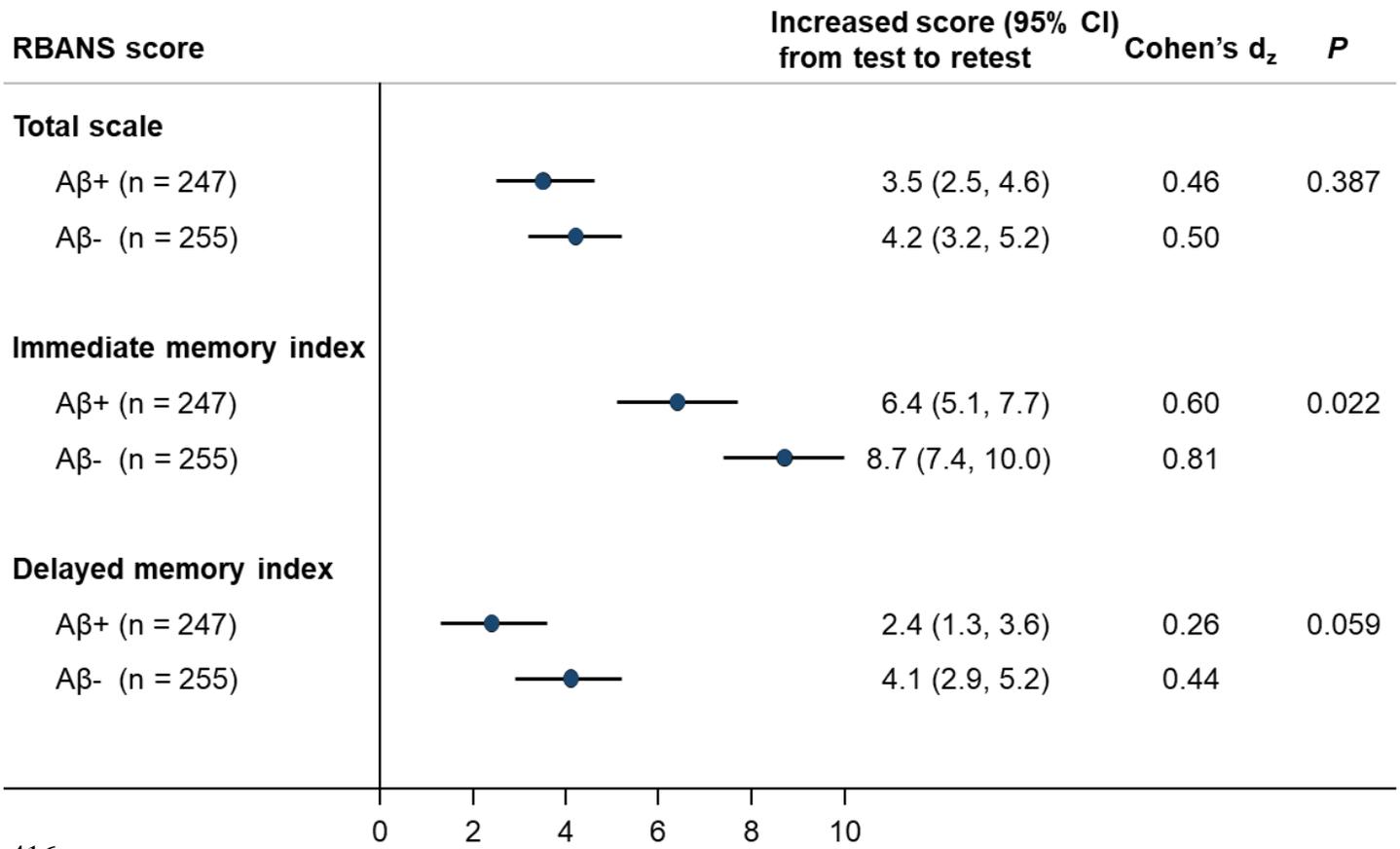
RBANS score, $\bar{x} \pm SD$	Test	Retest	Difference score (mean)	Difference score (range)	Cohen's d_z	<i>P</i> value
Total scale	102.7 \pm 11.8	106.6 \pm 12.9	3.9	-20, 38	0.48	<0.001
Immediate memory index	101.6 \pm 12.7	109.2 \pm 13.4	7.6	-28, 35	0.70	<0.001
Delayed memory index	100.7 \pm 10.1	104.0 \pm 10.6	3.3	-35, 36	0.35	<0.001
Visuospatial construction index	95.8 \pm 14.1	96.6 \pm 14.3	0.8	-37, 41	0.06	0.176
Language index	104.1 \pm 11.5	104.8 \pm 13.0	0.7	-42, 41	0.06	0.209
Attention index	108.8 \pm 14.5	109.3 \pm 14.7	0.5	-31, 32	0.05	0.293

409 Note: RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; SD = standard
 410 deviation. *P* values were calculated by paired *t* tests.

411 **Table 3. Associations between other characteristics and magnitude of RBANS practice effects**
 412 **(N = 502)**

Characteristics	No. of participants	Increase of total scale (95% CI)	<i>P</i> value	Increase of immediate memory index (95% CI)	<i>P</i> value	Increase of delayed memory index (95% CI)	<i>P</i> value
Age (years)			0.565		0.146		0.009
60-69	208	4.1 (3.0, 5.2)		6.8 (5.3, 8.2)		2.0 (0.8, 3.2)	
70-85	294	3.7 (2.8, 4.6)		8.1 (6.9, 9.3)		4.2 (3.1, 5.2)	
Sex			0.018		0.002		0.712
Male	248	3.0 (2.0, 4.0)		6.1 (4.8, 7.4)		3.1 (2.0, 4.2)	
Female	254	4.7 (3.7, 5.7)		9.0 (7.7, 10.2)		3.4 (2.3, 4.6)	
<i>APOE</i> - ϵ 4			0.164		0.004		0.337
Carrier	192	3.2 (2.1, 4.4)		5.9 (4.4, 7.3)		3.8 (2.5, 5.0)	
Non-carrier	310	4.3 (3.4, 5.2)		8.6 (7.5, 9.8)		3.0 (2.0, 4.0)	
Initial RBANS score			0.002		<0.001		<0.001
Higher than mean level	238	2.7 (1.6, 3.7)		4.7 (3.5, 6.0)		0.9 (-0.2, 1.9)	
Lower than mean level	264	5.0 (4.0, 5.9)		10.7 (9.3, 12.0)		6.2 (5.0, 7.4)	

413 Note: RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; CI =
 414 confidence interval. Estimates were adjusted for age, sex, education level, *APOE*- ϵ 4 carriage and
 415 initial RBANS level, where applicable.



416

417 **Figure 1. Associations between amyloid status with magnitude of RBANS practice effects (N =**
 418 **502)**

419 Note: RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; CI =
 420 confidence interval. Estimates were adjusted for age, sex, education level, *APOE-ε4* carriage and
 421 initial RBANS level, where applicable.