Neurobiology of Aging 112 (2022) 111–121

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging.org

Regular article

Predicting Alzheimer's disease with practice effects, APOE genotype and brain metabolism

Javier Oltra-Cucarella, Miriam Sánchez-SanSegundo*, Rosario Ferrer-Cascales, for the Alzheimer Disease Neuroimaging Initiative#

Department of Health Psychology, University of Alicante, Alicante, Spain)

ARTICLE INFO

Article history:
Received 12 March 2019
Revised 17 December 2021
Accepted 28 December 2021
Available online 2 January 2022

Keywords:
Biomarkers
Memory
Practice effects
Recognition
Reliable change

ABSTRACT

After the paper Cognition or genetics. Predicting progression to Alzheimer's disease with practice effects, APOE genotype and brain metabolism [Neurobiol Aging, 2018; 71:234–240] was published, we identified a coding error of one of the variables analyzed. To correct, update and expand the previous work, we compared simple and complex regression-based Reliable Change Index (RCI_{RB}) to analyze the risk of progression to AD (AD-risk) after six years using either delayed recall or recognition scores. Auditory Verbal Learning Test scores at six months for 394 individuals with normal cognition from the ADNI were used to build the regression. In 816 individuals with amnestic mild cognitive impairments, the AD-risk was associated with age, brain metabolism, APOE-ε4, recognition hits, the discrimination index, and low practice effects in the complex RCI_{RB} only. The complex RCI_{RB} outperformed the simple RCI_{RB}. Small correlations were found between practice effects and both Aβ (highest r = 0.218) and TAU (highest r = -0.183). RCI_{RB} are computationally simple and provide sensitive AD-risk estimates in combination with APOE-ε4 and FDG-PET.

1. Introduction

The rate of progression from amnestic Mild Cognitive Impairment (aMCI) to Alzheimer's disease (AD) varies depending on several factors such as the source of recruitment and the time frame under investigation. For example, Mitchell and Shiri-Feshki (2009) reported that the rate of progression over a period of 4.5 years on average was 17.1%–28.9% for community settings and 33.1%–33.6% for specialist settings, according to the diagnostic criteria used. Oltra-Cucarella et al., (2018a) reported that the rate of progression was 10.4%–20.6% for community settings over a period of 3.5 years on average, and 8.6%–40.4% for clinical settings over a period of 2.6 years on average. Both studies agreed in two main issues regarding the risk of AD (AD-risk): that the risk-AD was lower than the risk of remaining stable or reverting to normal; and that the rate of progression varied across MCI subtypes, with non-amnestic MCI showing a lower rate of progression compared to amnestic MCI (aMCI). For these reasons, new ways of identifying individuals at the greatest AD-risk have been investigated. One of these is repeated testing in memory tests, a useful statistical approach to identify memory impairments using change over time rather than using performance at a single time point.

Repeated cognitive testing can lead to incorrect conclusions if an individual's performance is compared against the same normative data on two occasions, because an increase in performance is expected for a number of cognitive tests due to the exposure to the same test in a previous occasion. This phenomenon, known as *practice effects* (Duff, 2012), has been documented in several populations including MCI (Calamia et al., 2012). Practice effects on memory tests have been reported in individuals with aMCI within the same session (Duff et al., 2012) and over periods of one week (Duff et al., 2017a), eighteen months (Campos-Magdaleno et al., 2017) and even 5 years (Gavett et al., 2016), have proven useful to identify individuals with aMCI who will show larger cog-

* Corresponding author at: Department of Health Psychology, University of Alicante, Alicante, Spain.
E-mail address: miriam.sanchez@ua.es (M. Sánchez-SanSegundo).

Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

0197-4580/$ – see front matter © 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
https://doi.org/10.1016/j.neurobiolaging.2021.12.011

nitive decline after one year follow-up (Duff et al., 2011), and also for improving the identification of individuals with MCI or those who will progress to normal cognition to MCI (Elman et al., 2018; Kremen et al., 2020). However, some researchers have reported no practice effects in individuals with aMCI over different periods (Darby et al., 2002; Schrijnemaekers et al., 2006), so their utility remains controversial.

Practice effects have been associated with APOE-ε4 genotype (Machulda et al., 2013; Zehnder et al., 2009) and with brain metabolism, which in turn have been associated with the AD-risk. It is known that APOE-ε4 carriers, mostly those carrying two copies of the allele, have an increased AD-risk compared to APOE-ε4 non-carriers and carriers of APOE-ε2 and ε3 alleles (Elias-Sonnenschein et al., 2011; Qian et al., 2017; Yu et al., 2014). In the study by Machulda et al., (2013), APOE-ε4 carriers failed to sustain their initial practice effects over one year, with a level of performance similar to baseline after approximately 6 years of follow-up. Regarding brain metabolism, although data on the accuracy of FDG-PET are highly variable (Smailagic et al., 2015), FDG-PET has been suggested as a more sensitive tool than cognitive scores for predicting AD in aMCI (Herholz et al., 2011). FDG-PET has been associated with practice effects on tests of visual and verbal memory, with more brain hypometabolism being associated with worse cognitive performance and lower practice effects (Duff et al., 2015, 2014). However, the associations between practice effects and the AD-risk, and also the differential predictive value of AD for practice effects, APOE-ε4 genotype and brain metabolism were not analyzed in either of these previous studies.

In our previous report (Oltra-Cucarella et al., 2018b), we aimed to analyze whether practice effects for delayed recall scores were useful to predict AD-risk in individuals with aMCI. After the paper was published, we notified the editor of an unintentional coding error. We noticed that we erroneously coded recognition scores as delayed recall scores. Free recall and recognition are two ways of assessing learning and memory. Whereas free recall relies on recollection through access to previously studied material (Tversky, 1973), recognition relies on both recollection and familiarity (Yonelinas et al., 2010). For this reason, performance on recognition tasks is usually higher than that on free recall tasks, even in individuals with memory impairments, because items not retrieved in free recall tasks can be identified as familiar in recognition tasks. The importance of recognition scores in MCI (Bennett et al., 2006) has been reported in previous works. Recognition scores have been found to improve the prediction of AD (Russo et al., 2017), have shown a lower false positive identification of performance validity test than AVLT delayed recall when combined with reliable digit span (Loring et al., 2016), and might be useful to discriminate between AD and other neurological diseases (Van Liew et al., 2016).

When analyzing performance on recognition tasks it is necessary to take account of both true positive and false positive responses. Thus, indices such as discrimination and response bias have been used to assessing memory (Stanislaw and Todorov, 1999). In aMCI, discrimination scores have proven more useful than both true positives and delayed recall scores for predicting progression to AD (De Simone et al., 2019). In the assessment of memory impairments, delayed recall scores show larger differences between aMCI and cognitively healthy individuals than recognition scores. For example, using the Consortium to Establish a Registry for Alzheimer's disease list learning test (range 0–15), Bennett, Golob, Parker and Starr (2006) found that individuals with aMCI recalled on average 6.1 fewer words than cognitively normal individuals on the 30-min free delayed recall task, whereas the difference on the 30-minute recognition task (range 0–45) was 3.7. De Simone, Perri, Fadda, Caltagirone and Carlesimo (2019) found a larger difference (effect size = 1.48) in delayed recall scores than in recognition scores (effect size = 0.98) between cognitively healthy individuals and individuals with aMCI, and also that recognition discriminability was the best predictor of progression to AD. Thus, if the distribution of recognition scores is closer to the control group used to build the regression equation, recognition scores should be expected to provide better estimates of practice effects compared to delayed recall scores.

The present work was developed to analyze the role of verbal memory and practice effects for the prediction of AD. Specifically, our objectives were 1) to analyze whether practice effects for delayed recall scores and recognition scores are similarly useful for predicting the AD-risk, 2) to compare simple and complex RCI to identify the AD-risk, and 3) to compare practice effects with APOE-ε4 and brain metabolism in the identification of progressors to AD. Additionally, we tested whether being categorized as not showing practice effects is related to amyloidosis and Tau levels obtained from cerebrospinal fluid (CSF). Previous works have shown that individuals showing low practice effects are more likely to have amyloid positivity in PET scans (Duff et al., 2017b, 2014; Mormino et al., 2014), with more amyloidosis being associated with lower practice effects. According to these data, we hypothesized that 1) practice effects on recognition scores would provide better estimates of the AD-risk than delayed recall scores for individuals with aMCI, and that 2) complex RCI_{RB} would be better than simple RCI_{RB} at identifying individuals at the greatest AD-risk. We also hypothesized that individuals categorized as not showing practice effects would show lower levels of amyloid-beta (Aβ) and higher levels of Tau and phosphorylated Tau (pTAU) in CSF.

2. Materials and methods

Data from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) were used in this study. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of the ADNI is to test whether serial magnetic resonance imaging, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

The normal cognition group (NC) included 394 participants (48.2% females) aged 56–89 years with no depression or metabolic diseases, no cognitive complaints, a Clinical Dementia Rating scale (CDR) score = 0, Mini-Mental State Examination (MMSE) score equal or higher than 24, normal education-corrected Logical Memory (LM) delayed recall scores, and no significant impairments in activities of daily living. None of the individuals in the NC group progressed to AD during a 6-year follow-up.

In the aMCI group (Petersen et al., 1999), 816 participants (40.8% females) aged 55–91 years with no metabolic diseases had subjective cognitive complaints, MMSE score $\geq$24, CDR score = 0.5 (mandatory memory box score $\geq$0.5), abnormal education-corrected LM delayed recall scores, general cognition and functional performance largely intact, and did not meet criteria for dementia. Five participants (0.6%) had mild depressive symptoms. All participants underwent physical and neurological examinations, screening laboratory tests, and provided blood samples for DNA and APOE testing. The ethical committee at each participating site approved the project, and all ADNI participants provided written consent before enrollment at each site.

2.1. Procedure

2.1.1. Auditory verbal learning test (AVLT)

The AVLT includes 15 words in List A that are read aloud over five trials. After the last presentation, an interference list (List B) with 15 different words is presented for a single trial, followed by free recall of the 15 words from List A. A delayed free recall of words from List A is obtained, followed by a recognition list including all the words from Lists A and B. As per the ADNI protocol, two equivalent alternate forms of the AVLT test were used across sessions.

2.1.2. Auditory verbal learning test – delayed recall scores

The AVLT delayed free recall trial (variable AVDEL30MIN for replication purposes) is administered 30 minutes after the last immediate recall trial. Baseline data from the 394 NC participants were used to predict AVLT delayed recall scores at six months (i.e., retest AVLT scores).

To analyze practice effects, the regression-based Reliable Change Index (RCI_{RB}) was used. The RCI_{RB} compares observed retest (Time 2) scores with predicted retest scores obtained with a linear regression. In the present work, the term *simple model* will be applied to the model in which retest scores were predicted using test (Time 1) scores alone, whereas the term *complex model* will be applied to the model in which retest scores were predicted using test scores in combination with a set of predictors. The RCI_{RB} calculates practice effects controlling for test-retest reliability, regression to the mean, floor and ceiling effects, and variability in both test and retest scores (Duff, 2012). The selection of the statistical model to calculate practice effects is not trivial, as complex regressions might identify more change than simpler methods (Duff et al., 2017a), and the statistics used to identify individuals with low practice effects might be related to Type I error (Crawford and Garthwaite, 2012).

In the simple RCI_{RB} model, retest AVLT delayed free recall scores at six months were regressed on baseline AVLT delayed free recall scores only, whereas in the complex RCI_{RB} model retest AVLT delayed free recall scores at six months were regressed on baseline AVLT delayed recall scores, age, sex and years of education. AVLT delayed recall scores at six months were predicted in each aMCI participant using the intercept, observed scores on each predictor and their associated beta coefficients from the simple and complex model separately. In both the simple and complex RCI_{RB} models, the discrepancy between predicted and observed scores was divided by the standard error of the prediction for a new case (Crawford et al., 2012; Crawford and Garthwaite, 2007; Crawford and Howell, 1998), and the standardized discrepancy was compared against a t_{n-k-1} degrees of freedom distribution (where k is the number of predictors and n is the sample size used to build the regression equation). The t_{n-k-1} distribution is preferred over a normal distribution because it treats the sample used to build the regression equation as a sample and not as a population, and it has a lower rate of type I error compared to a z distribution (Crawford et al., 2012; Crawford and Garthwaite, 2005). The RegBuild_MR program for multiple regressions (Crawford et al., 2012) was used to obtain the standard error of the prediction for each case and the p-value associated with the standardized discrepancy.

2.1.3. Auditory verbal learning test – recognition scores

The AVLT recognition trial is administered after the delayed free recall trial. Two measures are derived: the number of hits (i.e., items correctly identified as words from List A; variable AVDELTOT) and the number of false alarms (i.e., words erroneously identified as included in List A; variable AVDELERR2). Practice effects were calculated for the number of hits with the procedure detailed in section 2.1.2. To analyze whether the parameter estimates from the regression equations were trustworthy, the four regressions were repeated using 10,000 bootstrap replications.

As in previous research analyzing recognition scores on aMCI (De Simone et al., 2019), two measures related to recognition were calculated. D-prime (*d'*) was used as a discrimination index (Stanislaw and Todorov, 1999), with higher values of *d'* indicating a greater ability to distinguish between true positive and true negative responses. To calculate *d'*, a correction was applied whereby 0.5 was added to each frequency and the result was divided by the number of items + 1 (Snodgrass and Corwin, 1988). The C-index was used as a measure of response bias (Stanislaw and Todorov, 1999), with negative values indicating a bias towards responding *yes* and positive values indicating a bias towards responding *no*.

2.1.4. Practice effects groups

In both the simple and complex RCI_{RB} models, the p-value associated with the standardized discrepancy was interpreted as the percentage of individuals in the sample used to build the RCI regression equation showing an equal or larger discrepancy (Crawford and Garthwaite, 2007). Based on statistical cut-off points used in research on practice effects, the bottom 5% of the NC group was used to define low practice effects, which corresponds to scores from a t-distribution of approximately -1.64 for a one-sided test (Crawford and Garthwaite, 2012) and a sample of size $n = 394$[1]. Participants with aMCI showing a negative discrepancy found in 5% or less of the NC group were labeled as showing low practice effects (Low PE), whereas participants showing a discrepancy higher than the 5% of the sample used to build the equation were labeled as showing normal practice effects (Normal PE). Although a group of individuals with aMCI could show practice effects higher than NC individuals used to build the regression equation (e.g., >95%), we did not analyze these individuals separately because we were interested in analyzing individuals showing low practice effects. Although we acknowledge that the 5% cut-off point is arbitrary, it has been used to identify low practice effects (Crawford and Garthwaite, 2012). Thus, our interest is placed on "impairment" rather than on positive abnormality. The use of a one-tailed *t*-score of -1.64 to define abnormality of negative discrepancies has been suggested as the appropriate procedure when the RCI is used to analyze practice effects (Crawford and Garthwaite, 2012), especially when testing a directional hypothesis (Crawford and Garthwaite, 2007) as is the case in the present study. However, recent works have used the standardized discrepancy as a continuous variable to analyze whether practice effects are associated with the risk of cognitive decline and with cerebral biomarkers (Duff et al., 2017b, 2014). For this reason, we included the standardized discrepancy to analyze whether the continuous variable provide any benefit over and beyond the categorical variable use to identify PE.

2.2. FDG-PET, Aβ and TAU measures

For information about neuroimaging and biomarker data acquisition see http://adni.loni.usc.edu/methods/. The variable FDG from the ADNIMERGE file was analyzed, which indicates the baseline average FDG uptake of angular, temporal and posterior cingulate gyri, with higher values of FDG-PET indicating higher cerebral metabolism (Landau et al., 2011). FDG-PET values were multiplied by 100 for values to show the difference in the AD-risk for one

[1] Scores from a t-distribution with n = 394 approximate scores from a normal distribution and are comparable to z-scores

unit increase in FDG-PET metabolism. Using this scale is also easier to interpret than using exponentiated values. One-hundred and eighty-nine individuals with aMCI had missing FDG-PET values, reducing the sample available to 623 participants.

For data regarding CSF samples and the Luminex platform with Innogenetics (INNO-BIA AlzBio3; Ghent, Belgium; for research use–only reagents) immunoassay kit–based reagents see Shaw et al., (2009). $A\beta_{1\text{-}42}$, total-Tau and p-Tau from 503 participants with aMCI were included in the analyses after excluding participants with missing values in any of the CSF biomarkers. The variables ABETA, TAU and PTAU from the ADNIMERGE file were analyzed, which provide the number of pg/mL for each biomarker. Lower levels of $A\beta_{1\text{-}42}$ in CSF indicate a higher level of cerebral amyloid concentration, whereas higher levels of total-Tau and p-Tau in CSF indicate higher levels of neuronal damage.

2.3. Outcome

The primary outcome was the difference in the hazard of progressing to AD (McKhann et al., 2011, 1984) during a 6-year follow-up period. Secondary outcome included differences in CSF biomarkers between PE groups.

2.4. Statistical analysis

Continuous and categorical demographics were compared between groups with t-tests and χ-squared test respectively. For continuous variables, parametric tests were used whenever sample sizes were higher than 85, as means and standard deviations from samples this large are unbiased irrespective of skewness (Piovesana and Senior, 2018)**.** The effect size of the differences was calculated with Hedge's *g* and the square root of the average of the square standard deviations, with values of 0.20, 0.50 and 0.80 indicating small, medium and large effect sizes respectively (Fritz et al., 2012). The risk of having at least one copy of the APOE-ε4 allele was compared between groups using odds ratios (OR). The differential risk of being labeled as Low PE with either the simple or the complex RCI_{RB} was analyzed with ORs and the phi statistic for contingency tables, with values of 0.10, 0.30 and 0.50 indicating small, medium and large effects respectively (Cohen, 1992)**.**

The AD-risk in the aMCI group was compared with hazard ratios (HR) from a series of backward stepwise Cox proportional survival models. Univariate models included AVLT raw scores or PE groups as a binary variable. Multivariate models included the variables from the univariate models plus age, sex, education, MMSE, APOE-ε4, and FDG-PET. For the delayed recall scores, the multivariate model included also the standardized discrepancies between predicted and observed scores, and baseline AVLT as covariates, which allowed controlling whether the AD-risk for practice effects was above and beyond baseline AVLT delayed recall scores (Duff et al., 2017b, 2015, 2011; Gavett et al., 2016; Hassenstab et al., 2015). For recognition scores, the multivariate model included also standardized discrepancies, baseline recognition hits and false alarms, recognition discrimination index (*d'*), and recognition response bias (C-index).

The survival models including simple RCI_{RB} analyzed the effects of age, sex, and education on the progression to AD, but no association between demographic variables and practice effects. In the survival models including complex RCI_{RB}, the effects of demographics were analyzed both for practice effects and for the AD-risk. The assumption of proportionality was checked using log-minus-log plots (Vittinghoff et al., 2005).

To analyze the association among memory scores, practice effects and CSF biomarkers, bivariate Pearson's correlations were calculated between the standardized discrepancies, $A\beta$ and Tau values, and delayed recall and recognition AVLT scores. All statistical analyses were performed using SPSS v.26, with alpha level set at 0.05.

2.5. Comparison of the risk of AD according to clinical profile

We categorized participants with aMCI into one of four groups according to the clinical profile: group 1) participants showing low practice effects but no APOE-ε4 alleles [i.e., Low PE/APOE-ε4-], group 2) participants having at least one APOE-ε4 allele and normal practice effects [i.e., Normal PE/ APOE-ε4+], 3) participants showing low practice effects and having at least one APOE-ε4 allele [i.e., Low PE/APOE-ε4+], and group 4) participants showing practice effects with no APOE-ε4 alleles [i.e., Normal PE/APOE-ε4-]. This latter group was used as the reference group for comparisons. The AD-risk was compared among the four groups using a Cox regression with age, sex, education, and MMSE baseline scores.

3. Results

Table 1 shows demographics, FDG-PET, APOE-ε4, $A\beta$, Tau, and AVLT delayed recall and recognition scores for the NC and aMCI groups. The NC group was slightly older and more educated than the aMCI group. There was a higher proportion of women in the NC group. Participants in the NC group had higher MMSE scores, higher AVLT delayed free recall scores and hits at baseline and after 6 months, fewer false alarms at baseline and after 6 months, higher brain metabolism, higher $A\beta$ values, and lower Tau and p-Tau values compared to the aMCI group. According to effect sizes, differences between NC and aMCI groups were negligible for education, small for age and Tau, medium for $A\beta$ and pTau, and large for MMSE scores. The rate of progression from aMCI to AD during 6-years of follow-up was 21.4%, similar to other studies using the ADNI database (Russo et al., 2017).

3.1. AVLT delayed recall scores

Hedge's *g* showed that differences between NC and aMCI groups were large for both test and retest AVLT delayed recall scores. Retest AVLT delayed recall scores were significantly lower than baseline AVLT delayed recall scores both for the NC group ($t_{393} = 5.03$, $p < 0.001$) and for the aMCI group ($t_{815} = 7.98$, $p < 0.001$). The distribution of baseline raw AVLT delayed recall scores showed that 56.6% and 37.1% of the aMCI group scored lower than 1SD and 1.5SD of the NC group respectively**.**

As expected, the number of cases used to build the regression equation showing a standardized discrepancy equal to or more extreme than -1.64 was close to the nominal 5% for both the simple (4.8%) and the complex (4.8%) RCI_{RB} models. The average discrepancy between predicted and observed delayed recall scores in the aMCI group was -1.26 (SD = 2.98). The simple RCI_{RB} and complex RCI_{RB} identified 32 (3.9%) and 40 (4.9%) individuals with aMCI showing a discrepancy between observed and predicted retest AVLT delayed recall shown by fewer than 5% of the NC group. Bootstrap replications showed trustworthiness of confidence intervals in the regression equations (Supplemental material 1).

The Low PE category was highly correlated between the simple and complex RCI_{RB} model (phi = 0.80), so that the probability of being labeled as Low PE with either model was similar (OR = 1.27, SE = 0.24, $p = 0.832$). For the simple RCI_{RB} model, 15.6% (n = 5) of Low PE individuals progressed to AD, compared to 22% (n = 170) Normal PE individuals. For the complex RCI_{RB} model, 20% (n = 8) Low PE individuals progressed to AD, compared to 21.8% (n = 167) Normal PE individuals.

Table 1
Demographics, cognitive scores and biomarker levels for NC and MCI groups

	NC (n = 394)	MCI (n = 816)	*p*	Cohen's *d*
Sex (Female)	190 (48.2%)	333 (40.8%)	0.015	–
Age	74.83 (5.73)	73.06 (7.48)	<0.001	0.27
Education	16.30 (2.73)	15.92 (2.86)	0.026	0.14
MMSE	29.06 (1.14)	27.56 (1.82)	<0.001	0.98
Test AVLT-DR	7.54 (3.8)	3.79 (3.83)	<0.001	0.98
Retest AVLT-DR	6.72 (3.75)	3.11 (3.63)	<0.001	0.98
Test AVLT-H	12.82 (2.42)	10.49 (3.48)	<0.001	0.78
Test AVLT-F	0.78 (1.16)	1.75 (2.02)	<0.001	-0.59
Retest AVLT-H	12.66 (2.44)	10.06 (3.65)	<0.001	0.84
Retest AVLT-F	0.62 (1.29)	1.68 (2.09)	<0.001	-0.61
d'	1.83 (1.61)	1.91 (1.08)	0.271	-0.06
C-index	0.33 (0.46)	0.31 (0.47)	0.505	0.04
APOE-ε4 (1+)	107 (27.2%)	419 (51.3%)	<0.001	–
FDG-PET	130.70 (11.47)	124.43 (13.44)	<0.001	0.50
Amyloid-β	1024.90 (386.92)	837.77 (341.25)	<0.001	0.51
Tau	237.36 (87.29)	288.81 (129.53)	<0.001	-0.47
p-Tau	21.84 (8.81)	28.04 (14.69)	<0.001	-0.51

Key: AVLT-DR, auditory verbal learning test delayed recall scores; AVLT-F, auditory verbal learning test recognition false positive errors; AVLT-H, auditory verbal learning test recognition true positive scores; C-index, response bias index at baseline; d', d-prime discriminability index at baseline; MCI, mild cognitive impairment group; MMSE, mini-mental state examination; NC, cognitively normal group.

Table 2
Results for univariate and multivariate Cox survival regressions

AVLT score	RCI Model	Cox regression	Variables	OR (95% CI)	*p*
Delayed recall	Simple RCI_{RB}	Univariate	AVLT-DR	0.76 (0.71–0.81)	<0.001
		Univariate	PE	0.69 (0.29–1.69)	0.428
		Multivariate	AVLT-DR	0.82 (0.74–0.92)	<0.001
			PE	1.48 (0.29–7.63)	0.638
			t-scores	0.75 (0.49–1.14)	0.179
	Complex RCI_{RB}	Univariate	PE	0.92 (0.45–1.86)	0.809
		Multivariate	AVLT-DR	0.82 (0.74–0.90)	<0.001
			PE	1.87 (0.48–7.38)	0.369
			t-scores	0.76 (0.51–1.14)	0.183
Recognition hits	Simple RCI_{RB}	Univariate	Hits	0.85 (0.82–0.89)	<0.001
		Univariate	PE	3.03 (2.24–4.08)	<0.001
		Multivariate	AVLT-Hits	0.91 (0.86–0.97)	<0.001
			PE	2.02 (1.28–3.19)	0.003
			t-scores	0.92 (0.70–1.21)	0.544
	Complex RCI_{RB}	Univariate	PE	2.87 (2.13–3.86)	<0.001
		Multivariate	AVLT-Hits	0.91 (0.86–0.97)	0.004
			PE	2.17 (1.37–3.43)	0.001
			t-scores	0.82 (0.56–1.18)	0.287

Complex RCI_{RB} model: retest scores regressed on test scores, age, sex, and education. See section 2.4 for variables included in the multivariate Cox regressions. Key: AVLT, auditory verbal learning test; DR, delayed recall; OR, odds ratio; PE, practice effects group; RCI_{RB}, regression-based reliable change index; Simple RCI_{RB} model, retest scores regressed on test scores.

Univariate survival models in the aMCI group showed that the AD-risk was statistically associated with baseline AVLT delayed recall scores, but not with PE using either the simple or the complex RCI_{RB} model (Table 2). Multivariate survival analyses for the simple and complex RCI_{RB} model showed that the risk-AD was related to age (HR = 1.05, $p = 0.008$), FDG-PET (HR = 0.95, $p = 0.000$), MMSE scores (HR = 0.88, $p = 0.045$) and baseline AVLT delayed recall scores, but not with having one or more APOE-ε4 allele (HR = 1.52, $p = 0.083$). Categorizing individuals with aMCI as Low PE was not statistically significant for either the simple or the complex RCI_{RB} models, nor were the z-scores associated with the discrepancy between predicted and observed retest AVLT delayed recall scores (Table 2). The same results were found when both simple and complex RCI models were entered into the survival regression. Log-minus-log plots showed no evidence of non-proportionality (Supplemental material 2).

3.2. AVLT recognition scores

Hedge's *g* showed that differences between NC and aMCI groups were medium to large for baseline AVLT hits and false alarms, and non-significant for discrimination index and response bias. AVLT recognition hits were similar at baseline and after 6 months in the NC group ($t_{393} = 1.28$, $p = 0.202$), but were smaller at retest in the aMCI group ($t_{815} = 4.07$, $p < 0.001$). The distribution of test AVLT recognition hits showed that 45.1% and 27.8% of the aMCI group scored lower than 1SD and 1.5SD of the NC group respectively. These data indicate that the overlap of raw scores between the NC and the aMCI groups was larger for AVLT recognition scores than for AVLT delayed recall scores.

As expected, the number of cases used to build the regression equation showing a standardized discrepancy equal to or more extreme than -1.64 was close to 5% for both the simple (6.3%) and

the complex (6.6%) RCI_{RB} models. The average discrepancy between predicted and observed recognition hits in the aMCI group was -1.69 (SD = 3.83). The number of individuals categorized as Low PE was 210 (24%) for the simple RCI_{RB} and 210 (25.7%) for the complex RCI_{RB}. Bootstrap replications showed trustworthiness of confidence intervals (Supplemental material 1).

As for AVLT delayed recall scores, the Low PE category was highly correlated between the simple and complex RCI_{RB} model (phi = 0.89), and the probability of being labeled as Low PE was similar between both models (OR = 1.10, SE = 0.11, $p = 0.788$). Of the 606 individuals categorized as Normal PE, only 6 individuals (0.7%) showed practice effects higher than cognitively normal individuals, which suggest that aMCI subjects tend not to improve more than healthy controls. There were no statistically significant differences between Low PE and Normal PE groups in age, sex or education (p values >0.286).

Individuals in the Low PE group were more likely to have at least one copy of the APOE-ε4 allele (OR = 2.00, 95%CI = 1.45, 2.77, $p < 0.001$) than participants in the Normal PE group. Overall, 174 (21.5%) individuals progressed to AD. The percentage of progressors to AD was 39.3% in the Low PE group and 15.8% in the Normal PE group for the simple RCI_{RB}, and 37.6% in the Low PE group and 15.8% in the Normal PE group for the complex RCI_{RB}.

Univariate survival models showed that the AD-risk was statistically associated with baseline AVLT hit scores, and with Low PE using the simple and the complex RCI_{RB} models (Table 2). In the multivariate survival model for the simple RCI_{RB} model, the AD-risk was associated with age (HR = 1.04, $p = 0.037$), MMSE scores (HR = 0.87, $p = 0.039$), FDG-PET (HR = 0.96, $p < 0.001$), having at least one APOE-ε4 allele (HR = 1.64, $p = 0.039$), baseline recognition hits (HR = 0.91, $p = 0.003$), the discrimination index *d'* (HR = 0.73, $p = 0.002$), and with being labeled as Low PE (Table 2; Supplemental material 3).

For the complex RCI_{RB} model, the AD-risk was associated with age (HR = 1.04, $p = 0.012$), FDG-PET (HR = 0.96, $p < 0.001$), having at least one APOE-ε4 allele (HR = 1.67, $p = 0.033$), baseline recognition hits (HR = 0.91, $p = 0.004$), the discrimination index *d'* (HR = 0.72, $p = 0.001$), and with being labeled as Low PE (Table 2).

The continuous standardized discrepancy between predicted and observed scores were not associated with the AD-risk either for the simple or for the complex RCI_{RB} models (Table 2), nor were recognition false alarms (HR = 1.07, $p = 0.185$) and recognition response bias (HR = 0.94, $p = 0.794$). The same results were found when both the simple and the complex RCI_{RB} models were entered into the survival regression, with the simple RCI_{RB} model becoming statistically non-significant.

3.3. Association between practice effects and CSF biomarkers

Table 3 shows the bivariate correlations between memory scores, standardized discrepancies and CSF biomarkers in the aMCI group. Delayed free recall scores correlated with recognition scores, with higher delayed recall associated with higher recognition hits (38.07% shared variance) and lower recognition false alarms (10.89% shared variance). The continuous standardized discrepancies between predicted and observed scores correlated with delayed free recall scores and recognition scores, but not with discrimination index *d'* and response bias. CSF biomarkers correlated with delayed recall scores, recognition scores and standardized discrepancies. These correlations were higher for delayed recall (8.41%–9% shared variance) than for recognition scores (3.17%–5.81% shared variance) or for standardized discrepancies (1%–4.75% shared variance). Lastly, lower levels of Aβ were associated with higher levels of Tau and pTau in CSF. Most correlations showed a low to medium effect size. The only correlations showing a large effect size were the correlations between delayed recall scores and recognition hits.

3.4. Comparison of the AD-risk according to clinical profile

The survival regression showed that the AD-risk was higher for individuals in the Low PE/APOE-ε4- group (HR = 2.49, $p = 0.001$), individuals in the Normal PE/APOE-ε4+ group (HR = 1.97, $p = 0.002$), and individuals in the Low PE/APOE-ε4+ group (HR = 3.41, $p < 0.001$) compared to those in the Normal PE/APOE-ε4- group.

Figure 1 shows how the absolute AD-risk for participants with aMCI over a 6-year follow-up when APOE-ε4 and 6-months practice effects are combined. As shown, for individuals meeting standard criteria for aMCI at baseline assessment, the expected AD-risk at six years without any additional information is 21.4%. The AD-risk increases for individuals with one or more APOE-ε4 alleles, with a risk estimate twice as high as that for APOE-ε4 negative individuals. Adding data on practice effects in recognition hits six months after baseline again modifies the risk estimates. The AD-risk among APOE-ε4 negative individuals is three times as high for Low PE individuals, and also higher than the absolute risk for aMCI at baseline. Among individuals with one or more APOE-ε4 alleles, Low PE individuals have the greatest AD-risk, with a risk estimate twice as high as that for aMCI diagnosis at baseline. Interestingly, the AD-risk for APOE-ε4 negative and Low PE individuals is higher than the AD-risk for APOE-ε4 positive and normal PE individuals, which suggests that cognitive variables might be more useful than APOE genotype to identify a greater risk of progression from MCI to AD.

4. Discussion

After the original paper on PE and the AD-risk was published (Oltra-Cucarella et al., 2018b), we identified a coding error of the AVLT delayed recall variable. This unintentional error could confuse the readers and researchers, because delayed recall and recognition are two different stages in memory and differ both theoretically and in the distribution of scores. To avoid confusion and misinterpretation of our original results, the present study was conducted to correct the coding error and aimed 1) to analyze whether the RCI_{RB} estimates differ for delayed recall scores and for recognition scores, 2) to analyze whether complex RCI_{RB} would provide additional value over simple RCI_{RB} to identify individuals with aMCI at the greatest AD-risk and, additionally, 3) to analyze the association between practice effects and CSF biomarkers. Two main results must be highlighted. First, the complex RCI_{RB} calculated with age, sex, education and baseline scores was superior to the simple RCI_{RB} model for predicting the AD-risk. Second, the superiority of the complex RCI_{RB} model was found when practice effects were calculated for AVLT recognition hits, with no significant association between the AD-risk and practice effects for AVLT delayed recall scores. Thus, the present work replicates the results of the previous report, and also shows that the AD-risk was associated to recognition scores rather than to delayed recall scores.

The results reported here show that individuals with aMCI who showed low PE on recognition task across two successive assessments had a significantly higher AD-risk, and that the risk estimate for PE was higher than that for FDG-PET values and APOE-ε4 genotype. Although PE were related to genetic data as previously reported (Duff et al., 2017b; Machulda et al., 2013), with the Low PE group being more likely to have at least one copy of the APOE-ε4 allele, the Low PE category outperformed FDG-PET in the identification of individuals at the greatest AD-risk, and showed a similar and more precise risk estimate than APOE-ε4

Table 3
Correlations among cognitive scores, practice effects and CSF biomarkers

	2	3	4	5	6	7	8	9	10	11	12
(1) AVLT DR	0.617[b]	-0.335[b]	0.052	0.007	0.192[b]	0.184[b]	0.359[b]	0.336[b]	0.320[b]	-0.297[b]	-0.308[b]
(2) AVLT H	1	-0.090[b]	0.062	-0.061	0.184[b]	0.186[b]	0.259[b]	0.276[b]	0.241[b]	-0.191[b]	-0.199[b]
(3) AVLT F		1	-0.008	0.005	-0.133[b]	-0.132[b]	-0.091[b]	-0.083[a]	-0.185[b]	0.178[b]	0.184[b]
(4) d'			1	-0.356[b]	-0.005	-0.031	-0.001	-0.017	0.034	0.016	0.007
(5) C-index				1	0.011	0.021	0.020	0.021	-0.008	-0.002	-0.002
(6) DR t-values SR					1	0.966[b]	0.359[b]	0.340[b]	0.218[b]	-0.105[a]	-0.122[b]
(7) DR t-values MR						1	0.355[b]	0.374[b]	0.196[b]	-0.111[b]	-0.127[b]
(8) Rec t-values SR							1	0.985[b]	0.206[b]	-0.183[b]	-0.187[b]
(9) Rec t-values MR								1	0.185[b]	-0.181[b]	-0.183[b]
(10) Ab									1	-0.268[b]	-0.308[b]
(11) Tau										1	0.982[b]
(12) pTau											1

Key: AVLT-DR, auditory verbal learning test delayed recall scores; AVLT-F, auditory verbal learning test recognition false positive errors; AVLT-H, auditory verbal learning test recognition true positive scores; C-index, response bias index at baseline; d', d-prime discriminability index at baseline; MR, complex RCI model; SR, simple RCI model.
[a] <0.05. N = 816. N for Aβ = 503. N for Tau = 583.
[b] <0.01.

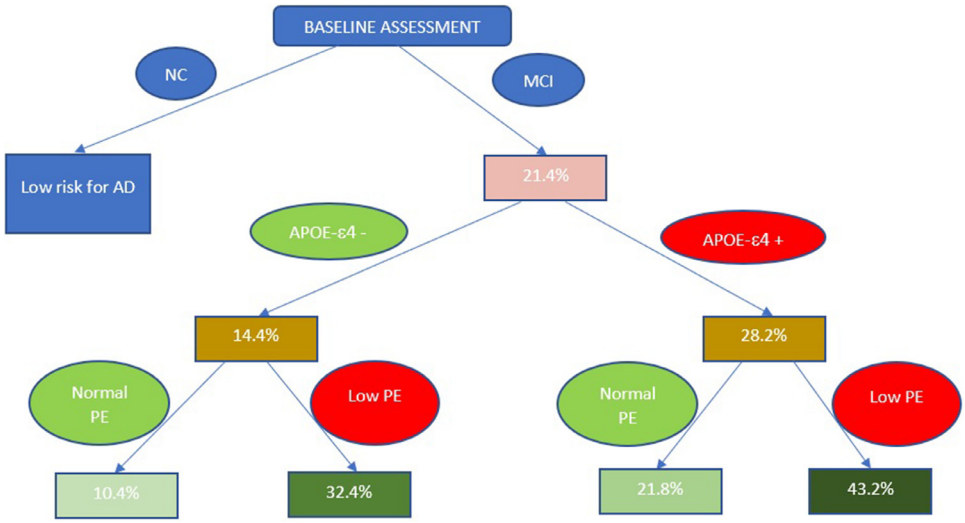

Fig. 1. Estimates of the absolute AD-risk at six years according to clinical profile. NC: normal cognition. MCI: mild cognitive impairment. APOE-ε4+: individuals with one or more APOE-ε4 allele. APOE-ε4-: individuals with no APOE-ε4 allele. Normal PE: individuals showing practice effects on the Auditory Verbal Learning Test recognition hits. Low PE: individuals not showing practice effects on the Auditory Verbal Learning Test recognition hits. "(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)"

genotype. These results are in line with other reports analyzing practice effects and biomarkers (Nation et al., 2019). For example, Hassenstab et al., (2015) found that APOE was not significant to predict worsening of clinical symptoms of cognitive impairment. Thus, our results cannot support that FDG-PET are more sensitive than cognitive scores for predicting AD in aMCI, as previously suggested (Herholz et al., 2011). Duff et al., (2015) suggested that PE can be a proxy of certain biomarkers, and our data add that practice effects and biomarkers might be a useful combination to identify individuals at the greatest AD-risk during a 6-year follow-up. However, risk estimates for APOE-ε4 and FDG-PET may be biased due to the association of these two variables in the ADNI database (Landau et al., 2013). Our results are also in line with the findings reported by Machulda et al., (2013), who found that APOE carriers' performance was similar to baseline after an average follow-up period of 6 years. The association between practice effects and biomarkers is still under debate (Jutten et al., 2020), so further works are needed to replicate whether PE are similar or even superior to genetic and biomarker data for predicting progression to AD in different follow-up periods.

One of the most important data in our work is related to the importance of AVLT recognition scores for the prediction of AD, in line with previous reports. Russo, Campos, Vázquez, Sevlever and Allegri (2017) found that the interaction between AVLT delayed recall and recognition scores were significant to predict progression from aMCI to AD in the ADNI database. Recently, De Simone et al., (2019) found that individuals with aMCI recalled fewer items than did cognitively healthy individuals on a list-learning test, and were worse in recognizing true positive items. Interestingly, the discrimination index *d'* was the variable that best predicted progression to AD. Although *d'* was statistically significant in our model, being labeled as Low PE with the RCI_{RB} was the variable that best predicted progression to AD. Of note is that the statistical analyses used in the present work are not comparable to the ones used by De Simone et al., (2019), who analyzed progression to AD with ROC curves, and also that we applied the correction to *d'* suggested

by Snodgrass and Corwin (1988). The main result is that among measures related to verbal memory, recognition on a list-learning test may be more sensitive than delayed recall for the prediction of progression from aMCI to AD.

Previous studies comparing simple and complex RCI_{RB} models reported that multiple regression models might identify more change than simple regression models (Duff et al., 2017a). Our results support these conclusions, as complex RCI_{RB} remained significant to predict the AD-risk, with both simple and complex RCI_{RB} models showing a higher AD-risk estimate when compared to APOE-ε4 and FDG-PET. However, the high correlation between the simple and complex RCI_{RB} might cause collinearity in the analyses, so the superior ability of the complex RCI_{RB} must be replicated in future works. One of the main characteristics of the RCI_{RB} models are that they can be calculated as long as raw data are reported, because regression parameters can be computed using summary statistics (Crawford et al., 2012; Crawford and Garthwaite, 2007). In the case of multiple regressions, a k x k matrix of correlations is also needed (Crawford et al., 2012). Due to the findings reported in this and previous works (Duff et al., 2017a), we encourage researchers to report correlations among variables that allow calculating the complex RCI_{RB} for the individual case (Table 2). However, when it is not possible to calculate complex models, the simple RCI_{RB} model is expected to provide similar results.

It is important to highlight that our results apply only to Crawford and Howell's model for RCI_{RB}. In contrast to other models for calculating practice effects, Crawford and Howell (1998) and Crawford and Garthwaite (Crawford et al., 2012; Crawford and Garthwaite, 2007) recommend using a t-distribution for standardized discrepancies between observed and predicted scores. According to these authors, the t-distribution treats the statistics from the regression equation as sample statistics and not as population statistics with known means and SDs. Additionally, the p-value associated to t-scores tests whether the discrepancy between observed and predicted scores is an observation from the sample used to build the equation. In Crawford and Garthwaite's words, "The p value used to test significance is also a point estimate of the proportion of the population with the same value on the predictor variable (i.e., X) as the patient who would obtain a discrepancy more extreme than that which was observed for the patient" (Crawford and Garthwaite, 2007, p. 613). Thus, the question remains of whether the RCI_{RB} provides any benefit over raw scores when the case's score on the predictor is likely to fall outside the range of scores in the sample used to build the equation. In this case, the range of raw scores would discriminate between individuals (e.g., aMCI vs. cognitively healthy individuals) and the case under study could not be considered as an observation of the sample used to build the regression equation even before calculating practice effects.

Our results seem to support this hypothesis. The RCI_{RB} for delayed recall scores, which showed a lower overlap between the aMCI and NC groups, identified a very small number of individuals showing low PE and were not significant for predicting AD. Relatedly, raw delayed recall scores outperformed the dichotomous PE variable for identifying progression to AD. Conversely, recognition scores for the MCI group were on average 1SD below the NC group as found in previous research comparing cognitively normal and cognitively impaired individuals (Spaan et al., 2005). Thus, recognition scores had a higher overlap between aMCI and NC groups than did delayed recall scores, and proved to be more effective than APOE-ε4 and FDG-PET values to predict the AD-risk. Indeed, the dichotomous practice effects variable for recognition hits proved to be the best predictor of progression to AD, better than discrimination and response bias indices from the recognition task (De Simone et al., 2019) and also better than the continuous practice effects variable. As shown by the small overlap on the delayed recall scores, this could be an artifact of using a regression formula for values (i.e., values from individuals with MCI) that are outside the range of the values used to build the regression equation (i.e., values from cognitively normal individuals).

An alternative, but related explanation is that the floor effect on delayed recall scores could prevent individuals with aMCI from showing a test-retest discrepancy in the range of that found in NC individuals, which could explain that the RCI_{RB} for delayed recall scores identified a small percentage of individuals with aMCI showing low practice effects (i.e., <5%). The distribution of residuals in the NC group showed that the 5th percentile corresponds to discrepancies equal to or higher than -4.74 items. The small average discrepancy between predicted and observed delayed recall scores in the aMCI group, along with the floor effects on observed delayed recall scores, might cause the difficulty in identifying individuals with extreme discrepancies. Nation et al., (2019) built a regression line including nondemented individuals from the ADNI database, which eliminates this limitation as the range of scores for the regression equation includes scores from individuals with MCI. However, Nation et al., (2019) combined both immediate and delayed recall scores, and did not use recognition hits in their calculations. Thus, no comparison between Nation et al.'s results and the results reported here is possible. In contrast, the residual distribution of recognition hits in the NC group showed that the 5th percentile corresponds to discrepancies equal to or higher than -3.58 items. The small average discrepancy between predicted and observed recognition hits in the context of higher average raw scores, may increase the possibility of showing extreme discrepancies and being classified as showing low PE.

Our results showed that low PE were associated with lower Aβ levels and higher Tau and p-Tau levels in CSF, supporting the data indicating that more CSF Aβ concentration and neural damage is associated with reduced practice effects (Duff et al., 2017b, 2014; Jutten et al., 2020; Mormino et al., 2014; Nation et al., 2019), although the size of the association was small. However, notable differences between studies must be highlighted. Mormino et al., (2014) analyzed the association between practice effects and Aβ in cognitively normal individuals, and both Mormino et al., (2014) and Duff et al., (2017b, 2014) assessed Aβ concentration with PET techniques, whereas we used Aβ and Tau levels in CSF.

One of the most important differences between the present and previous works analyzing the association between RCI and Aβ concentration is the sample size, which might be the reason for the differences in the effect size of the correlations, and also for the absence of statistical significance of standardized discrepancies in the survival regression. On the one hand, after calculating the RCI_{RB} in a large sample we found that the dichotomized variable PE outperformed the continuous standardized discrepancy variable related to PE, which was not significant for identifying progression to AD in either simple or complex models. On the other hand, using z-scores from small samples is likely to bias the results by inflating the Type I error (Crawford and Garthwaite, 2012). For example, a standardized discrepancy of -1.645 is associated with a 4.99% probability when a z-score is used. When a t-distribution is used, a standardized discrepancy of -1.645 is associated with an 8% probability for a sample of size n = 5, and equals the probability associated with z-scores when the size of the sample is greater than 300. Crawford and Garthwaite (2005) reported that the rate of possible misclassifications increases with the use of z-scores compared to t-values even for samples of size equal to or higher than 100. Indeed, stable means and standard deviations (and, thus, the possibility of using z-scores) are found when the size of the sample is greater than 85 regardless of the level of skewness (Piovesana and Senior, 2018).

The size of the sample used in the present study to calculate PE is larger than the minimum sample needed to run a multiple regression with 4 predictors (Green, 1991; Tabachnick and Fidell, 2013), and reduces the probability that the distribution of residuals was a concern (Lumley et al., 2002; Williams et al., 2013). The similarity in confidence intervals for the sample estimates and bootstrap estimates supports this conclusion. Relatedly, concerns regarding the distribution of residuals for regressions with small samples (Lumley et al., 2002), and the fact that small samples are only powered to identify large effects that become smaller as the size of the sample increases (Button et al., 2013), both have been related to unreliable results. These reasons seem to support the use of a t-distribution for the calculation of PE (Crawford and Garthwaite, 2012; Crawford and Howell, 1998), with the use of a dichotomous PE variable rather than a continuous standardized discrepancy.

Our results have limitations that must be highlighted. The findings reported here are applicable to individuals who are administered two alternate forms of the same test in a 6 months follow-up period. This wide follow-up period may preclude the utility of these findings for inclusion of individuals with aMCI in clinical trials. Practice effects calculated over shorter periods (Duff et al., 2017b, 2014) may provide a more feasible way of identifying individuals with aMCI at a higher AD-risk who can be rapidly included in prevention or intervention trials.

The procedures used in the ADNI may also affect the results. Regarding the assessment of memory abilities, the ADNI includes two alternate versions of the AVLT. Although practice effects have been reported even for alternate versions (Beglinger et al., 2005), the second version of the AVLT may be harder than the first and may explain in part the lower performance at follow-up, mostly for individuals with lower memory ability levels (Crane et al., 2012) as is the case in individuals with aMCI. Also, the use of one single memory test for the aMCI diagnosis in the ADNI has been related to a number of false positive MCI diagnoses, which in turn have been associated with low rates of progression to AD (Bondi et al., 2014; Edmonds et al., 2015; Oltra-Cucarella et al., 2018c). Thus, the utility of PE on either delayed recall or recognition scores from a verbal memory test must be replicated in samples of individuals diagnosed with MCI using other MCI criteria. When diagnostic criteria other than those developed by Petersen have been used to analyze PE in the ADNI database, delayed recall scores have been statistically significant to predict the AD-risk (Nation et al., 2019). However, the use of composite scores of both immediate and delayed recall scores precludes comparisons with our methodology.

Rather than being considered as a source of error, practice effects may provide valuable information for the identification of individuals with aMCI at the greatest risk of progressing to AD. Our results showed that RCI_{RB} models may be useful for estimating PE when large samples are used to build the regression equations, with risk estimates similar to those obtained with APOE-ε4 genotype. Our data highlight the need to investigate whether PE are more suitable for variables in which individual case's scores fall within the distribution of scores in the control sample. Whether complex models including cognitive, demographic and clinical variables provide more accurate risk estimates of the progression to AD than simple RCI models also needs to be analyzed in future works, especially for small samples and when the same version of the test is used in serial assessment. The findings reported here could also be useful for interpreting the results of clinical trials (Brooks and Loewenstein, 2010), as it has been shown that it is important not only to identify changes in raw scores over a 6 months period but also to identify whether negative discrepancy between observed and expected scores are uncommon in healthy individuals who do not progress to AD.

Author contributions

Javier Oltra-Cucarella: Conceptualization, Data curation, Formal analysis, Validation, Roles/Writing - original draft; Writing - review & editing. Miriam Sánchez-SanSegundo: Conceptualization, Validation, Writing - review & editing. Rosario Ferrer-Cascales: Conceptualization, Validation, Writing - review & editing.

Disclosure statement

The authors have no conflict of interest to report.

Acknowledgements

The authors thank the editors of Neurobiology of Aging for giving us the opportunity of correcting the coding error identified in the original report. The authors also thank the reviewers for improving the quality of our manuscript. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2021.12.011.

References

Beglinger, L.J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D.A., Crawford, J., Fastenau, P.S., Siemers, E.R., 2005. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol. 20, 517–529. doi:10.1016/j.acn.2004.12.003.

Bennett, I.J., Golob, E.J., Parker, E.S., Starr, A., 2006. Memory evaluation in mild cognitive impairment using recall and recognition tests. J. Clin. Exp. Neuropsychol. 28, 1408–1422. doi:10.1080/13803390500409583.

Bondi, M.W., Edmonds, E.C., Jak, A.J., Clark, L.R., Delano-Wood, L., McDonald, C.R., Nation, D.A., Libon, D.J., Au, R., Galasko, D., Salmon, D.P., 2014. Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates. J. Alzheimers Dis. 42, 275–289. doi:10.3233/JAD-140276.

Brooks, L.G., Loewenstein, D.A., 2010. Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. Alzheimers Res. Ther. 2. doi:10.1186/alzrt52.

Button, K.S., Ioannidis, J.Pa, Mokrysz, C., Nosek, B.a, Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. doi:10.1038/nrn3475.

Calamia, M., Markon, K., Tranel, D., 2012. Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. Clin. Neuropsychol. 26, 543–570. doi:10.1080/13854046.2012.680913.

Campos-Magdaleno, M., Facal, D., Lojo-Seoane, C., Pereiro, A.X., Juncos-Rabadán, O., 2017. Longitudinal assessment of verbal learning and memory in amnestic mild cognitive impairment: practice effects and meaningful changes. Front. Psychol. 8. doi:10.3389/fpsyg.2017.01231.

Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155–159. doi:10.1037/0033-2909.112.1.155.

Crane, P.K., Carle, A., Gibbons, L.E., Insel, P., Mackin, R.S., Gross, A., Jones, R.N., Mukherjee, S., Curtis, S.M., Harvey, D., Weiner, M., Mungas, D., 2012. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imaging Behav 6, 502–516. doi:10.1007/s11682-012-9186-z.

Crawford, J.R., Garthwaite, P.H., 2012. Single-case research in neuropsychology: A comparison of five forms of t-test for comparing a case to controls. Cortex 48, 1009–1016. doi:10.1016/j.cortex.2011.06.021.

Crawford, J.R., Garthwaite, P.H., 2007. Using regression equations built from summary data in the neuropsychological assessment of the individual case. Neuropsychology 21, 611–620. doi:10.1037/0894-4105.21.5.611.

Crawford, J.R., Garthwaite, P.H., 2005. Evaluation of criteria for classical dissociations in single-case studies by monte carlo simulation. Neuropsychology 19, 664–678. doi:10.1037/0894-4105.19.5.664.

Crawford, J.R., Garthwaite, P.H., Denham, A.K., Chelune, G.J., 2012. Using regression equations built from summary data in the psychological assessment of the individual case: Extension to multiple regression. Psychol. Assess. 24, 801–814. doi:10.1037/a0027699.

Crawford, J.R., Howell, D.C., 1998. Comparing an individual's test score against norms derived from small samples. Clin. Neuropsychol. 12, 482–486. doi:10.1076/clin.12.4.482.7241.

Darby, D., Maruff, P., Collie, A., McStephen, M., 2002. Mild cognitive impairment can be detected by multiple assessments in a single day. Neurology 59, 1042–1046. doi:10.1212/WNL.59.7.1042.

De Simone, M.S., Perri, R., Fadda, L., Caltagirone, C., Carlesimo, G.A., 2019. Predicting progression to Alzheimer's disease in subjects with amnestic mild cognitive impairment using performance on recall and recognition tests. J. Neurol. 266, 102–111. doi:10.1007/s00415-018-9108-0.

Duff, K., 2012. Evidence-Based Indicators of Neuropsychological Change in the Individual Patient: Relevant Concepts and Methods. Arch. Clin. Neuropsychol. 27, 248–261. doi:10.1093/arclin/acr120.

Duff, K., Atkinson, T.J., Suhrie, K.R., Dalley, B.C.A., Schaefer, S.Y., Hammers, D.B., 2017a. Short-term practice effects in mild cognitive impairment: Evaluating different methods of change. J. Clin. Exp. Neuropsychol. 39, 396–407. doi:10.1080/13803395.2016.1230596.

Duff, K., Chelune, G., Dennett, K., 2012. Within-session practice effects in patients referred for suspected dementia. Dement. Geriatr. Cogn. Disord. 33, 245–249. doi:10.1159/000339268.

Duff, K., Foster, N.L., Hoffman, J.M., 2014. Practice effects and amyloid deposition: preliminary data on a method for enriching samples in clinical trials. Alzheimer Dis. Assoc. Disord. 28, 247–252. doi:10.1097/WAD.0000000000000021.

Duff, K., Hammers, D.B., Dalley, B.C.A., Suhrie, K.R., Atkinson, T.J., Rasmussen, K.M., Horn, K.P., Beardmore, B.E., Burrell, L.D., Foster, N.L., Hoffman, J.M., 2017b. Short-term practice effects and amyloid deposition: providing information above and beyond baseline cognition. J. Prev. Alzheimers Dis. 4, 87–92. doi:10.14283/jpad.2017.9.

Duff, K., Horn, K.P., Foster, N.L., Hoffman, J.M., 2015. Short-term practice effects and brain hypometabolism: preliminary data from an FDG PET study. Arch. Clin. Neuropsychol. 30, 264–270. doi:10.1093/arclin/acv018.

Duff, K., Lyketsos, C.G., Beglinger, L.J., Chelune, G., Moser, D.J., Arndt, S., Schultz, S.K., Paulsen, J.S., Petersen, R.C., McCaffrey, R.J., 2011. Practice effects predict cognitive outcome in amnestic mild cognitive impairment. Am. J. Geriatr. Psychiatry 19, 932–939. doi:10.1097/JGP.0b013e318209dd3a.

Edmonds, E.C., Delano-Wood, L., Clark, L.R., Jak, A.J., Nation, D.A., McDonald, C.R., Libon, D.J., Au, R., Galasko, D., Salmon, D.P., Bondi, M.W., 2015. Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. Alzheimers Dement 11, 415–424. doi:10.1016/j.jalz.2014.03.005.

Elias-Sonnenschein, L.S., Viechtbauer, W., Ramakers, I.H.G.B., Verhey, F.R.J., Visser, P.J., 2011. Predictive value of APOE- 4 allele for progression from MCI to AD-type dementia: a meta-analysis. J. Neurol. Neurosurg. Psychiatry 82, 1149–1156. doi:10.1136/jnnp.2010.231555.

Elman, J.A., Jak, A.J., Panizzon, M.S., Tu, X.M., Chen, T., Reynolds, C.A., Gustavson, D.E., Franz, C.E., Hatton, S.N., Jacobson, K.C., Toomey, R., McKenzie, R., Xian, H., Lyons, M.J., Kremen, W.S., 2018. Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. Alzheimers Dement. Diagn. Assess. Dis. Monit. 10, 372–381. doi:10.1016/j.dadm.2018.04.003.

Fritz, C.O., Morris, P.E., Richler, J.J., 2012. Effect size estimates: current use, calculations, and interpretation. J. Exp. Psychol. Gen. 141, 2–18. doi:10.1037/a0024338.

Gavett, B.E., Gurnani, A.S., Saurman, J.L., Chapman, K.R., Steinberg, E.G., Martin, B., Chaisson, C.E., Mez, J., Tripodis, Y., Stern, R.A., 2016. Practice effects on story memory and list learning tests in the neuropsychological assessment of older adults. PLOS ONE 11, e0164492. doi:10.1371/journal.pone.0164492.

Green, S.B., 1991. How many subjects does it take to do a regression analysis. Multivar. Behav. Res. 26, 499–510. doi:10.1207/s15327906mbr2603_7.

Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., Morris, J.C., 2015. Absence of practice effects in preclinical Alzheimer's disease. Neuropsychology 29, 940–948. doi:10.1037/neu0000208.

Herholz, K., Westwood, S., Haense, C., Dunn, G., 2011. Evaluation of a calibrated 18F-FDG PET score as a biomarker for progression in alzheimer disease and mild cognitive impairment. J. Nucl. Med. 52, 1218–1226. doi:10.2967/jnumed.111.090902.

Jutten, R.J., Grandoit, E., Foldi, N.S., Sikkes, S.A.M., Jones, R.N., Choi, S.-E., Lamar, M.L., Louden, D.K.N., Rich, J., Tommet, D., Crane, P.K., Rabin, L.A., 2020. Lower practice effects as a marker of cognitive performance and dementia risk: A literature review. Alzheimers Dement. Diagn. Assess. Dis. Monit. 12, e12055. doi:10.1002/dad2.12055.

Kremen, W.S., Sanderson-Cimino, M.E., Elman, J.A., Tu, X.M., Gross, A.L., Panizzon, M.S., Eglit, G.M.L., Jak, A.J., Edmonds, E.C., Thomas, K.R., Eppig, J.S., Williams, M.E., Bondi, M.W., Lyons, M.J., Franz, C.E., 2020. Accounting for cognitive practice effects results in earlier detection and more accurate diagnosis of MCI: Biomarker confirmation. Alzheimers Dement 16, e044883. doi:10.1002/alz.044883.

Landau, S.M., Harvey, D., Madison, C.M., Koeppe, R.A., Reiman, E.M., Foster, N.L., Weiner, M.W., Jagust, W.J., 2011. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol. Aging 32, 1207–1218. doi:10.1016/j.neurobiolaging.2009.07.002.

Landau, S.M., Lu, M., Joshi, A.D., Pontecorvo, M., Mintun, M.A., Trojanowski, J.Q., Shaw, L.M., Jagust, W.J., 2013. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of β-amyloid: CSF and Amyloid PET Imaging. Ann. Neurol. 74, 826–836. doi:10.1002/ana.23908.

Loring, D.W., Goldstein, F.C., Chen, C., Drane, D.L., Lah, J.J., Zhao, L., Larrabee, G.J., 2016. False-positive error rates for reliable digit span and auditory verbal learning test performance validity measures in amnestic mild cognitive impairment and early alzheimer disease. Arch. Clin. Neuropsychol. 31, 313–331. doi:10.1093/arclin/acw014.

Lumley, T., Diehr, P., Emerson, S., Chen, L., 2002. The importance of the normality assumption in large public health data sets. Annu. Rev. Public Health 23, 151–169. doi:10.1146/annurev.publhealth.23.100901.140546.

Machulda, M.M., Pankratz, V.S., Christianson, T.J., Ivnik, R.J., Mielke, M.M., Roberts, R.O., Knopman, D.S., Boeve, B.F., Petersen, R.C., 2013. Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the mayo clinic study of aging. Clin. Neuropsychol. 27, 1247–1264. doi:10.1080/13854046.2013.836567.

McKhann, G.M., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group* under the auspices of department of health and human services task force on Alzheimer's disease. Neurology 34, 939. doi:10.1212/WNL.34.7.939.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7, 263–269. doi:10.1016/j.jalz.2011.03.005.

Mitchell, A., Shiri-Feshki, M., 2009. Rate of progression of mild cognitive impairment to dementia - Meta-analysis of 41 robust inception cohort studies. Acta Psychiatr. Scand. 119, 252–265. doi:10.1111/j.1600-0447.2008.01326.x.

Mormino, E.C., Betensky, R.A., Hedden, T., Schultz, A.P., Amariglio, R.E., Rentz, D.M., Johnson, K.A., Sperling, R.A., 2014. Synergistic effect of β-Amyloid and neurodegeneration on cognitive decline in clinically normal individuals. JAMA Neurol 71, 1379. doi:10.1001/jamaneurol.2014.2031.

Nation, D.A., Ho, J.K., Dutt, S., Han, S.D., Lai, M.H.C., 2019. Neuropsychological decline improves prediction of dementia beyond Alzheimer's disease biomarker and mild cognitive impairment diagnoses. J. Alzheimers Dis. 69, 1171–1182. doi:10.3233/JAD-180525.

Oltra-Cucarella, J., Ferrer-Cascales, R., Alegret, M., Gasparini, R., Díaz-Ortiz, L.M., Ríos, R., Martínez-Nogueras, Á.L., Onandia, I., Pérez-Vicente, J.A., Cabello-Rodríguez, L., Sánchez-SanSegundo, M., 2018a. Risk of progression to AD for different neuropsychological Mild Cognitive Impairment subtypes. A hierarchical meta-analysis of longitudinal studies. Psychol. Aging 33, 1007–1021. doi:10.1037/pag0000294.

Oltra-Cucarella, J., Sánchez-SanSegundo, M., Ferrer-Cascales, R., 2018b. Cognition or genetics? Predicting Alzheimer's disease with practice effects, APOE genotype, and brain metabolism. Neurobiol. Aging 71, 234–240. doi:10.1016/j.neurobiolaging.2018.08.004.

Oltra-Cucarella, J., Sánchez-SanSegundo, M., Lipnicki, D.M., Sachdev, P.S., Crawford, J.D., Pérez-Vicente, J.A., Cabello, L., Ferrer-Cascales, R., 2018c. Using the base rate of low scores helps to identify progression from amnestic MCI to AD. J. Am. Geriatr. Soc. 66, 1360–1366. doi:10.1111/jgs.15412.

Petersen, R., Smith, G., Waring, S., Ivnik, R., Tangalos, T., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308. doi:10.1001/archneur.56.3.303.

Piovesana, A., Senior, G., 2018. How Small Is Big: Sample Size and Skewness. Assessment 25, 793–800. doi:10.1177/1073191116669784.

Qian, J., Wolters, F.J., Beiser, A., Haan, M., Ikram, M.A., Karlawish, J., Langbaum, J.B., Neuhaus, J.M., Reiman, E.M., Roberts, J.S., Seshadri, S., Tariot, P.N., Woods, B.M., Betensky, R.A., Blacker, D., 2017. APOE-related risk of mild cognitive impairment

and dementia for prevention trials: An analysis of four cohorts. PLOS Med 14, e1002254. doi:10.1371/journal.pmed.1002254.

Russo, M.J., Campos, J., Vázquez, S., Sevlever, G., Allegri, R.F., 2017. Adding recognition discriminability index to the delayed recall is useful to predict conversion from mild cognitive impairment to Alzheimer's disease in the Alzheimer's disease neuroimaging initiative. Front. Aging Neurosci. 9. doi:10.3389/fnagi.2017.00046.

Schrijnemaekers, A.M.C., de Jager, C.A., Hogervorst, E., Budge, M.M., 2006. Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls. J. Clin. Exp. Neuropsychol. 28, 438–455. doi:10.1080/13803390590935462.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M.-Y., Trojanowski, J.Q., 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann. Neurol. 65, 403–413. doi:10.1002/ana.21610.

Smailagic, N., Vacante, M., Hyde, C., Martin, S., Ukoumunne, O., Sachpekidis, C., 2015. 18F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst. Rev. doi:10.1002/14651858.CD010632.pub2.

Snodgrass, J.G., Corwin, J., 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. J. Exp. Psychol. Gen. 117, 34–50. doi:10.1037/0096-3445.117.1.34.

Spaan, P.E.J., Raaijmakers, J.G.W., Jonker, C., 2005. Early assessment of dementia: the contribution of different memory components. Neuropsychology 19, 629–640. doi:10.1037/0894-4105.19.5.629.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. 31, 137–149. doi:10.3758/BF03207704.

Tabachnick, B.G., Fidell, L.S., 2013. Using Multivariate Statistics, Sixth Ed. Pearson Education Inc, New Jersey.

Tversky, B., 1973. Encoding processes in recognition and recall. Cognit. Psychol. 5, 275–287. doi:10.1016/0010-0285(73)90037-6.

Van Liew, C., Santoro, M.S., Goldstein, J., Gluhm, S., Gilbert, P.E., Corey-Bloom, J., 2016. Evaluating recall and recognition memory using the montreal cognitive assessment: applicability for Alzheimer's and Huntington's diseases. Am. J. Alzheimers Dis. Dementias® 31, 658–663. doi:10.1177/1533317516668573.

Vittinghoff, E., Glidden, D.V., Shiboski, S.C., McCulloch, C.E., 2005. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, Statistics for Biology and Health. Springer, New York.

Williams, M.N., Gómez Grajales, C.A., Kurkiewicz, D., 2013. Assumptions of multiple regression: correcting two misconceptions. Pract. Assess. Res. Eval. 18, 1–14.

Yonelinas, A.P., Aly, M., Wang, W.-C., Koen, J.D., 2010. Recollection and familiarity: Examining controversial assumptions and new directions. Hippocampus 20, 1178–1194. doi:10.1002/hipo.20864.

Yu, J.-T., Tan, L., Hardy, J., 2014. Apolipoprotein E in Alzheimer's Disease: An Update. Annu. Rev. Neurosci. 37, 79–100. doi:10.1146/annurev-neuro-071013-014300.

Zehnder, A.E., Bläsi, S., Berres, M., Monsch, A.U., Stähelin, H.B., Spiegel, R., 2009. Impact of APOE status on cognitive maintenance in healthy elderly persons. Int. J. Geriatr. Psychiatry 24, 132–141. doi:10.1002/gps.2080.