

- National Population Health Survey. *Am J Epidemiol* 2012; 176: 1095–100.
35. Chan YK, Lee RP. Network size, social support and happiness in later life: a comparative study of Beijing and Hong Kong. *J Happiness Stud* 2006; 7: 87–112.
36. Shiroma EJ, Lee IM. Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. *Circulation* 2010; 122: 743–52.
37. Thun MJ, Peto R, Lopez AD *et al.* Alcohol consumption and mortality among middle-aged and elderly U.S. adults. *N Engl J Med* 1997; 337: 1705–14.
38. Shye D, Mullooly JP, Freeborn DK, Pope CR. Gender differences in the relationship between social network support and mortality: a longitudinal study of an elderly cohort. *Soc Sci Med* 1995; 41: 935–47.
39. Wong HZ, Lim WY, Ma SS, Chua LA, Heng DM. Health screening behaviour among Singaporeans. *Ann Acad Med Singapore* 2015; 44: 326–34.

Received 10 November 2017; editorial decision 8 June 2018

Age and Ageing 2018; **47**: 866–871
doi: 10.1093/ageing/afy136
Published electronically 17 August 2018

© The Author(s) 2018. Published by Oxford University Press on behalf of the British Geriatrics Society. All rights reserved. For permissions, please email: journals.permissions@oup.com

Do people with Alzheimer’s disease improve with repeated testing? Unpacking the role of content and context in retest effects

ALDEN L. GROSS^{1,2}, NADIA CHU^{1,2}, LORETTA ANDERSON³, M. MARIA GLYMOUR⁴, RICHARD N. JONES⁵, FOR THE COALITION AGAINST MAJOR DISEASES*

¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

²Johns Hopkins University Center on Aging and Health, Baltimore, MD 21205, USA

³Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

⁴Departments of Epidemiology and Biostatistics, University of California, San Francisco, CA 94158, USA

⁵Departments of Psychiatry and Human Behavior and Neurology, Warren Alpert Medical School, Brown University, Providence, RI 02903, USA

Address correspondence to: Dr Alden Gross, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 2024 E. Monument Street, Baltimore, MD 21205, USA. E-mail: agross14@jhu.edu

*Data used in the preparation of this article were obtained from the Coalition Against Major Diseases (CAMD) database. In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the Coalition Against Major Diseases (CAMD). The Coalition brings together patient groups, biopharmaceutical companies, and scientists from academia, the US Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA). The Coalition Against Major Diseases (CAMD) includes over 200 scientists from member and non-member organizations. The data available in the CAMD database has been volunteered by CAMD member companies and non-member organizations.

Abstract

Objective: retest effects may be attributed to ‘repeated content’ in neuropsychological tests such as words in word list-learning tests, or the ‘testing context’ which involves procedural memory and reduced test anxiety following repeated administration. Alzheimer’s Disease (AD) severely impairs episodic memory, so longitudinal cognitive testing among people with dementia may reveal the relative contributions of content versus context to retest effects in neuropsychological testing.

Method: we used data from the Critical Path Institute’s repository of placebo arm data from randomized controlled trials (RCTs) of dementia conducted by participating pharmaceutical companies ($N = 990$ people, 4,170 study visits, up to 2.4 years of follow-up). To estimate retest effects on the Mini-Mental State Examination (MMSE), we used linear regressions with random effects for people and time, adjusting for age, sex and race, and longitudinal quantile regressions.

Results: average MMSE score (16.6 points, SD = 5.5, range 1, 27) declined by 2.0 points/year (95% confidence interval, CI: -2.3, -1.8). Mean retest effect was 0.6 points (95% CI: 0.4, 0.8) at second assessment (average 4 months after baseline). Retest effects were similar among participants with and without any recall on the short-delay word recall subscale score at baseline, and at the 30th, 50th and 70th percentiles of the MMSE distribution, suggesting similar retest effects across the spectrum from mild to severe cases of dementia.

Conclusions: retest effects are apparent in people with dementia despite reduced episodic memory, suggesting a prominent role of the testing context in RCTs and cohort studies.

Keywords: *retest effects, practice effects, neuropsychological testing, ageing, dementia*

Introduction

Retest or practice effects, the extent to which performance improves after repeated cognitive testing [1], are well-documented clinically and in longitudinal studies of cognitive ageing [2–4]. Retest effects due to repeated neuropsychological testing can obscure findings from studies for which cognitive decline is an outcome [3, 5]. The manner in which these effects are accounted for, whether statistically or through study design, depends on their presumed sources. Retest effects have been attributed to the ‘content’ of neuropsychological tests, as well as to the ‘context’ in which they are administered.

Retest effects attributable to memory for test content implies that respondents remember content from earlier testing on subsequent testing occasions. To minimize or eliminate retest effects related to content of neuropsychological testing, many studies use alternate test forms at different testing occasions (e.g. 6–7). The implicit motivation behind alternate forms is based on the notion that retest effects are attributable to episodic memory for test content [8]. The RBANS, for example, features presumably equivalent alternate forms to avoid confounding changes in cognitive performance with content-related retest effects [9]. However, previous research has also demonstrated that alternate forms that are parallel, but not necessarily equivalent, not only fail to eliminate retest effects in cognitively healthy and cognitively impaired older adults, but also further contaminate the observed cognitive trajectory with form effects [6, 10].

In contrast to the episodic memory explanation, the role of contextual factors on retest effects has been acknowledged but are underappreciated in research. Contextual factors in cognitive testing include task proficiency or familiarity, test anxiety, procedural memory and other environmental cues; these are multi-faceted processes that involve cultural and socioeconomic dynamics. Thorndike [11] reviewed reports of retest effects on the Army Alpha test battery, concluding that priming test-takers with an explanation of a test before administration it could eliminate retest effects. More recent recommendations involve dual baseline assessments in which an entire battery is repeated prior to the start of an intervention or analysis of trajectory [12]. This treatment of retest is consistent with

the notion that retest effects are attributable to contextual factors.

To our knowledge, the contribution of context to retest effects isolated from content-related retest has not been hitherto evaluated. Because AD is characterized by severely impaired episodic memory, but procedural memory remains largely intact, people with dementia provide an ideal opportunity in which to examine the contribution of procedural memory to retest effects. Thus, we examined the role of context-related retest effects in neuropsychological testing using prospectively collected longitudinal data on a large sample of patients with a clinical diagnosis of AD. Our assumption is that for cognitively impaired people, any observed retest effects are largely attributable to context or procedural memory. We hypothesized we would observe retest effects among people with dementia, even among persons with very poor recall ability (e.g. scoring 0 on a three-item short-delay recall subtest), which would indicate the key role of procedural familiarity in retest effects. We further hypothesized that subtests of the MMSE most susceptible to contextual cues, namely orientation to time and place, would show the largest retest effects.

Methods

Participants

We used data from the Critical Path Institute’s repository of data from randomized trials of AD conducted by member pharmaceutical companies. This repository includes prospectively collected data from the placebo arms of clinical trials of people with clinically diagnosed AD that administered the Mini-Mental State Examination (MMSE) repeatedly [13]. A review of the characteristics of the database has been published, and study procedures were approved by IRBs at companies and institutions where the studies were conducted consistent with the Helsinki Declaration [14].

Each trial had between 57 and 700 control participants, for a total of $N = 4,113$ participants and 19,560 study visits. To minimize variability in the test-retest interval which can affect retest effects [4], we excluded trials in which the cognitive testing interval was less than three months, leaving five trials ($N = 990$ participants over 4,170 visits).

Variables

The MMSE is a brief mental status instrument comprised of 19 questions that assess 11 domains: attention and calculation, comprehension, drawing, naming, orientation to place, orientation to time, reading, registration, recall, repetition and writing. Cues are not given in clinical administrations [13]. For attention and calculation, three of the five studies administered both serial 7s and DLROW while the other two administered only DLROW; for consistency we used the DLROW item available from all studies. Each of the subtests of the MMSE have administration histories as far back as the original Stanford–Binet intelligence exam [15]. The MMSE is typically inappropriate as a measure of longitudinal change in community-living samples because of pronounced ceiling effects and non-linear measurement properties [6]. However, in our sample of people with AD, ceiling effects are negligible (Supplemental Figure 1, available at Age and Ageing online). Moreover, we addressed non-linear scaling properties of the MMSE by estimating a composite score from a factor analysis on the 11 categorical subtests of the MMSE. The model is consistent with graded-response item response theory [16], and resulting scores theoretically have interval-level scaling properties in the study sample. We scaled the IRT-equated MMSE score to have the same theoretical range as the original MMSE (0–30 points). No inferences regarding retest effects in this study changed after comparing results using MMSE raw scores with analyses using MMSE factor scores (Supplemental Table 1, available at Age and Ageing online).

Analysis plan

We first described the sample using means for continuous variables and proportions for categorical variables. We next modelled longitudinal change in the MMSE total scores and subscales of the MMSE using random effects models with individual level random effects for intercept and time in the full sample. To further evaluate whether retest effects are diminished among participants with more severe dementia, we estimated retest effects separately among participants who were unable to recall any words at baseline in the three-word recall task, and also among those who recalled any words at baseline. While three-word recall may be seen as a relatively easy test that is not a precise measure of episodic memory among cognitively normal older adults, our sample was comprised of older adults with dementia. We then modelled longitudinal changes in each subtest of the MMSE to evaluate subtest-specific retest effects.

Each model included fix effects for time, the retest effect and covariates. The coefficient for time is interpretable as the annual rate of cognitive change after the second testing occasion. The retest effect in each model was parameterized as the difference between the first and subsequent testing occasions [17, 18]. The retest effect is thus interpretable as the jump in cognitive performance between the first visit and later visits, after accounting for the anticipated changes due to ageing [19]. Previous research has

suggested gains following the second testing occasion are negligible [18, 20], and more importantly that different ways of parameterizing retest effects do not affect inferences about covariate effects [21].

Covariates included age, sex, indicators for trial and race (white vs. non-white). The models use all observed data and parameter estimates are obtained assuming missing data are missing at random conditional on variables in the model [22].

Additional analyses. To evaluate the robustness of the results, we conducted two additional analyses. First, we estimated the random effects model in the largest single trial ($N = 406$ participants) to verify the estimated retest effect was not biased by the combination of multiple trials with different follow-up schedules. Our second sensitivity analysis was intended to address the possibility that regression to the mean would incorrectly inflate apparent practice effects in people with low baseline test scores. We therefore evaluated the robustness of the estimated retest effect across the range of baseline cognitive performance in the sample, using quantile regressions to avoid regression to the mean. This model evaluates changes in the 30th, 50th and 70th quantiles of MMSE across repeated testing, using a clustered bootstrap to estimate confidence intervals. Because the quantiles need not be defined by the same individuals at successive assessments, this method is not vulnerable to regression to the mean.

Results

The MMSE data contained 4,170 visits for $N = 990$ participants. The sample was mostly female (59%), white (92%), and the average age was 76 years (range 48, 95) (Table 1). Participants across the trials completed on average 4.2 study visits (range 1, 7) during up to 2.4 years of follow-up. The mean MMSE score at baseline was 16.5 (range 1, 27); the

Table 1. Characteristics of the sample ($N = 990$)

| Characteristic | Statistic | Observed range |
|--|------------|----------------|
| Age at baseline, Mean (SD) | 75.5 (8.1) | (48.0, 95.0) |
| White, non-Hispanic, n (%) | 910 (91.9) | |
| Sex, female, n (%) | 581 (58.7) | |
| Number of study visits, Mean (SD) | 4.2 (2.2) | (1.0, 7.0) |
| Years of follow-up, Mean (SD) | 0.9 (0.6) | (0.0, 2.4) |
| MMSE score, Mean (SD) | 16.5 (5.5) | (1.0, 27.0) |
| MMSE subscale scores (higher is better), Mean (SD) | | |
| Attention and Calculation | 2.7 (2.0) | (0.0, 5.0) |
| Comprehension | 2.4 (0.8) | (0.0, 3.0) |
| Drawing | 0.4 (0.5) | (0.0, 1.0) |
| Naming | 1.8 (0.5) | (0.0, 2.0) |
| Orientation to Place | 2.9 (1.6) | (0.0, 5.0) |
| Orientation to Time | 2.0 (1.7) | (0.0, 5.0) |
| Reading | 0.8 (0.4) | (0.0, 1.0) |
| Short-delay recall | 0.5 (0.8) | (0.0, 3.0) |
| Registration | 2.6 (0.8) | (0.0, 3.0) |
| Repetition | 0.6 (0.5) | (0.0, 1.0) |
| Writing | 0.7 (0.5) | (0.0, 1.0) |

SD: Standard deviation.

Table 2. Results from random effects regression models of MMSE in patients with Alzheimer’s disease ($N = 990$)

| Predictor | Full sample ($N = 990$) | | No baseline short-delay recall ($N = 678$) | | Any baseline short-delay recall ($N = 312$) | |
|-----------------------------|---------------------------|----------------|---|----------------|--|----------------|
| | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI |
| Intercept | 16.55* | (16.20, 16.90) | 15.07* | (14.64, 15.50) | 19.69* | (19.16, 20.22) |
| Annual rate of change | -2.02* | (-2.29, -1.75) | -2.05* | (-2.38, -1.72) | -2.01* | (-2.50, -1.52) |
| Retest effect | 0.61* | (0.43, 0.79) | 0.51* | (0.29, 0.73) | 0.83* | (0.50, 1.16) |
| Female sex | -1.59* | (-2.33, -0.85) | -1.76* | (-2.58, -0.94) | -1.1 | (1.06, -3.26) |
| Baseline age (per 10 years) | -0.42* | (-0.83, -0.01) | -0.11 | (-0.64, 0.42) | -0.14 | (0.13, -0.41) |
| Non-white race | 0.01 | (-1.26, 1.28) | -0.54 | (-2.09, 1.01) | 1.7 | (-1.63, 5.03) |

* $P < 0.05$.

Table 3. Retest effects in specific subscale scores of the MMSE in patients with Alzheimer’s disease from random effects regression models ($N = 990$)

| MMSE subscale scores | Retest effect estimate (Cohen’s d) | 95% CI |
|---------------------------|---------------------------------------|---------------|
| Orientation to Place | 0.18* (0.12) | (0.10, 0.26) |
| Short-delay recall | 0.09* (0.10) | (0.03, 0.15) |
| Registration | 0.07* (0.10) | (0.01, 0.13) |
| Orientation to Time | 0.10* (0.06) | (0.00, 0.20) |
| Attention and Calculation | 0.10 (0.05) | (-0.02, 0.22) |
| Writing | -0.02 (-0.05) | (-0.04, 0.00) |
| Comprehension | 0.03 (0.04) | (-0.03, 0.09) |
| Drawing | 0.02 (0.04) | (-0.02, 0.06) |
| Naming | -0.02 (-0.04) | (-0.04, 0.00) |
| Reading | 0.01 (0.03) | (-0.01, 0.03) |
| Repetition | 0.01 (0.02) | (-0.03, 0.05) |

* $P < 0.05$.

Subtests are sorted by the absolute value of the size of the standardized retest effect.

IRT-equated MMSE score at the first study visit demonstrated minimal ceiling effects (Supplemental Figure 1, available at Age and Ageing online).

Results from random effects models of the MMSE score, in the full sample and stratified by performance on the short-delay recall subtest of the MMSE, are in Table 2. From the full sample, the model-estimated MMSE score at baseline was 16.6 (95% CI: 16.2, 16.9), which agrees with the sample mean of 16.5 points. The model-estimated annual rate of decline in the MMSE was 2.0 points/year (95% CI: -2.3, -1.8 points/year). The retest effect among the entire sample was 0.6 MMSE points (95% CI: 0.4, 0.8), corresponding to a Cohen’s d effect size of 0.1 SD units. The estimated retest effect among the $N = 678$ (68%) participants with no baseline recall (0.5 points, 95% CI: 0.3, 0.7 points, $d = 0.09$) was similar to that among the $N = 312$ (32%) of individuals who recalled at least one word at baseline (0.8 points, 95% CI: 0.5, 1.2 points, $d = 0.14$) (Table 2). Although persons with any recall showed a larger retest effect than people who showed no recall, this difference was not statistically significant ($P = 0.07$).

Table 3 shows mean retest effects for subscales of the MMSE. Each score was standardized to its baseline standard deviation to facilitate comparisons of relative magnitudes across the subscales. Consistent with our a priori

hypothesis, significant and positive retest effects in the range of approximately 0.1 standardized units were observed for orientation to place (0.2, 95% CI: 0.1, 0.3) and time (0.1, 95% CI: 0.0, 0.2). Significant retest effects were also observed for registration (0.07, 95% CI: 0.01, 0.13) and short-delay recall (0.09, 95% CI: 0.03, 0.15).

Additional analyses. The model-estimated mean retest effect among participants from the largest dataset ($N = 406$) was 0.7 MMSE points (95% CI: 0.4, 0.9). Quantile regressions revealed similar magnitudes of retest at the 30th (0.42 MMSE points, 95% CI: -0.06, 0.9), 50th (0.46 MMSE points, 95% CI: -0.02, 0.9) and 70th percentiles (0.65 MMSE points, 95% CI: 0.05, 1.3), suggesting the presence of retest effects across the entire cognitive spectrum from severe to mild cases of AD dementia.

Discussion

We assessed retest effects by leveraging a large sample of individuals with AD dementia, including those who may have severe episodic memory impairments. This sample demonstrated a statistically significant improvement in MMSE scores, on average, between the first and subsequent visits, consistent with influences from aspects of the testing context on retest effects. By contrasting people with very poor recall (none of three words recalled on MMSE task) versus other people with at least one word recalled, we may infer the presence of some procedural memory and the possible effects of memory for, or familiarity with, the testing context as at least a partial explanation of the observed practice or retest effect observed in this sample. In fact, if people with no baseline short-term recall have zero capacity for episodic memory, then we might infer that at least 64% (0.09/0.14) of the retest effect in patients with dementia reflects procedural memory and the remaining third reflects episodic memory. Our results are consistent with the hypothesis that episodic memory for test content is not solely responsible for retest effects, but rather procedural memory and memory for context contributes to the observed retest effects in longitudinal studies.

Unlike most studies of retest effects, which have focused on cognitively healthy older adults, in the present study we examined effects of retesting in older adults with dementia

[8]. Older adults with dementia may have limited capacity to learn and synthesize information after repeated testing. Our study demonstrated that retest effects in samples of people with dementia are non-trivial, even though they may be smaller (0.11 SD units) than in groups of people without dementia (0.60 SD units) [19]. Despite, a relative lack of attention for retest effects in people with AD, they have been documented in early trials of tacrine in AD [23] and clinical trials of cholinesterase inhibitors in which even the control group improved during the trial [24, 25]—although in such settings retest might be attributable to effects of the medication rather than content or context of the battery. Other studies using the MMSE in people with dementia have observed robust practice effects [26], but other studies that used other, often more difficult, tests have not observed practice effects in people with dementia [27–29]. Our estimated retest effect among people with dementia may even be an underestimate: in another sample of patients with dementia, Abner and colleagues [2] reported a baseline to follow-up change of about 0.2 SD units among participants, many of whom had dementia. One potential explanation for the lower retest effect observed in the present study relative to other studies of retest effects in people with dementia is that we cannot be certain whether the sample was subject to some pre-screening testing not included in the Clinical Path institute data.

Our study has several limitations. First, we were unable to ascertain whether there were form differences across studies in the MMSE; this would add measurement error especially in analyses of retest effects on subscales of the MMSE. A second limitation is that we did not adjust for complicating diseases, medications, or education because such information was unavailable across all datasets. This limitation is only relevant, however, if some covariate is related to the size of the retest effect and also related to retesting occasion number. Third, the MMSE was not originally designed as a dementia monitoring tool, and demonstrates ceiling effects which can mask retest effects especially in community-living populations [30]. In this study, however, we did not observe strong ceiling effects on the MMSE because all participants had dementia (Supplemental Figure 1, available at *Age and Ageing* online). Further, we used the MMSE to represent cognitive trajectories amount people with dementia, not as a tool for diagnosing or monitoring dementia. A final limitation is that it is impossible to be certain that persons who score 0 on the short-delay recall have no capacity to develop memories for test content from one testing occasion to the next; there is no available external criterion available for impaired episodic memory. Although impaired episodic memory is likely the case for most participants with clinical dementia scoring at that level, we must be open to the possibility that observed practice effects are not entirely context-related and that other factors may have interrupted performance (e.g. aphasia, deafness, inattention, environmental disturbance, etc.). A related caveat is that one of the many impediments to success in treatment trials of AD is the

misdiagnosis of other forms of dementia as AD. Although all participants in the present study have a diagnosis of AD, we cannot be certain of the pathophysiology which may affect retest effects.

Retest effects are robust in people with AD despite reduced episodic memory, suggesting a prominent role of the testing context in RCTs and cohort studies. This has implications for studies and trials that longitudinally follow people with dementia. Previous reports have called for novel strategies, such as development of alternative forms, to eliminate changes in cognitive test performance that are attributable to retest effects (e.g. 8). Given the heterogeneity in retest across types of cognitive tests [4], the variability in retest intervals across different studies, and the inadequacy of commonly accepted alternate forms for traditional tests [19], we suggest retest effects need not be eliminated. Instead, analytic recommendations that have been documented in detail elsewhere (e.g. 19, 21) are readily available to accommodate and characterize retest effects, and adjust estimates in the presence of retest effects.

Key points

- Retest effects may be attributed to ‘repeated content’ or the ‘testing context’ in cognitive tests.
- Retest effects are apparent in people with AD despite impaired episodic memory, suggesting a role of context.
- Previous reports have called for use of alternative forms to eliminate retest effects. Given the heterogeneity in retest effects across different types of cognitive tests, the variability in retest intervals across different studies, and the inadequacy of many commonly accepted alternate forms of traditional tests, we suggest retest effects need not be eliminated. Instead, analytic recommendations are readily available to adjust estimates in the presence of retest effects.

Supplementary Data

Supplementary data mentioned in the text are available to subscribers in *Age and Ageing* online.

Acknowledgements

Data used in the preparation of this article were obtained from the Coalition Against Major Diseases database (CAMD). As such, the investigators within CAMD contributed to the design and implementation of the CAMD database and/or provided data, but did not participate in the analysis of the data or the writing of this report.

Funding

Dr Gross was supported by K01-AG050699 from the National Institute on Aging. Dr Jones was supported in

part by National Institute on Aging grant AG029672. This work was supported by National Institute on Aging grant R13 AG030995 (PI: Mungas).

Conflicts of interest

None.

References

- Horton AM Jr. Neuropsychological practice effects x age: a brief note. *Percept Mot Skills* 1992; 75: 257–8.
- Abner EL, Dennis BC, Mathews MJ, Mendiondo MS, Caban-Holt A, Kryscio RJ, SELECT Investigators. Practice effects in a longitudinal, multi-center Alzheimer's disease prevention clinical trial. *Trials* 2012; 13: 217.
- Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol* 1998; 20: 339–52.
- Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol* 2012; 26: 543–70.
- Salthouse TA. Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology* 2010; 24: 563–72.
- Crane PK, Carle A, Gibbons LE, Insel P, Mackin RS, Gross AL, Alzheimer's Disease Neuroimaging Initiative. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav* 2012; 6: 502–16.
- Jobe JB, Smith DM, Ball K *et al.* ACTIVE: a cognitive intervention trial to promote independence in older adults. *Control Clin Trials* 2001; 22: 453–79.
- Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimers Dement (Amst)* 2015; 1: 103–11.
- Randolph C. Repeatable battery for the assessment of neuropsychological status (RBANS). San Antonio, TX: Psychological Corporation, 1998.
- Gross AL, Inouye SK, Rebok GW *et al.* Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. *J Clin Exp Neuropsychol* 2012; 34: 758–72.
- Thorndike EL. Practice effects in intelligence tests. *J Exp Psychol* 1922; 5: 101–7.
- McCaffrey RJ, Westervelt HJ. Issues associated with repeated neuropsychological assessments. *Neuropsychol Rev* 1995; 5: 203–21.
- Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–98.
- Neville J, Kopko S, Broadbent S *et al.* Development of a unified clinical trial database for Alzheimer's disease. *Alzheimers Dement* 2015; 11: 1212–21.
- Terman LM. *The Measurement Of Intelligence*. Boston: Houghton Mifflin Company, 1916.
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 1969; 34: 1–17.
- Ivnik RJ, Smith GE, Lucas JA *et al.* Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance. *Neuropsychology* 1999; 13: 121–7.
- Rabbitt P, Diggle P, Holland F, McInnes L. Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *J Gerontol B Psychol Sci Soc Sci* 2004; 59: 84–97.
- Gross AL, Benitez A, Shih R *et al.* Predictors of retest effects in a longitudinal study of cognitive aging in a diverse community-based sample. *J Int Neuropsychol Soc* 2015; 21: 506–18.
- Kausler DH. *Learning and memory in normal aging*. San Diego: Academic Press, 1994.
- Vivot A, Power MC, Glymour MM *et al.* Jump, hop, or skip: modeling practice effects in studies of determinants of cognitive change in older adults. *Am J Epidemiol* 2016; 183: 302–14.
- Little RJA, Rubin DB. *Statistical analysis with missing data*. New York: John Wiley & Sons, 1987.
- Eagger S, Morant N, Levy R, Sahakian B. Tacrine in Alzheimer's disease. Time course of changes in cognitive function and practice effects. *Br J Psychiatry* 1992; 160: 36–40.
- Birks J, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst Rev* 2006; CD001190.
- Rogers SL, Doody RS, Mohs RC, Friedhoff LT. Donepezil improves cognition and global function in Alzheimer disease: a 15-week, double-blind, placebo-controlled study. Donepezil study group. *Arch Intern Med* 1998; 158: 1021–31.
- Galasko D, Abramson I, Corey-Bloom J, Thal LJ. Repeated exposure to the Mini-Mental State Examination and the Information-Memory-Concentration Test results in a practice effect in Alzheimer's disease. *Neurology* 1993; 43: 1559–63.
- Cooper DB, Epker M, Lacritz L *et al.* Effects of practice on category fluency in Alzheimer's disease. *Clin Neuropsychol* 2001; 15: 125–8.
- Hassenstab J, Ruvolo D, Jaszec M, Xiong C, Grant E, Morris JC. Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology* 2015; 29: 940–8.
- Helkala EL, Kivipelto M, Hallikainen M *et al.* Usefulness of repeated presentation of Mini-mental State Examination as a diagnostic procedure in a population-based study. *Acta Neurol Scand* 2002; 106: 341–6.
- Salthouse TA. When does age-related cognitive decline begin? *Neurobiol Aging* 2009; 30: 507–14.

Received 26 January 2018; editorial decision 6 June 2018