ELSEVIER

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring ■ (2018) 1-10

Alzheimer's & Dementia

Diagnostic Assessment & Prognosis

# Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects

Jeremy A. Elman[a,b,*], Amy J. Jak[a,c], Matthew S. Panizzon[a,b], Xin M. Tu[d], Tian Chen[e], Chandra A. Reynolds[f], Daniel E. Gustavson[a,b], Carol E. Franz[a,b], Sean N. Hatton[a,b], Kristen C. Jacobson[g], Rosemary Toomey[h], Ruth McKenzie[h], Hong Xian[i,j], Michael J. Lyons[h], William S. Kremen[a,b]

[a]*Department of Psychiatry, University of California San Diego, La Jolla, CA, USA*
[b]*Center for Behavior Genetics of Aging, University of California San Diego, La Jolla, CA, USA*
[c]*Psychology Service, VA San Diego Healthcare System, La Jolla, CA, USA*
[d]*Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA*
[e]*Department of Mathematics and Statistics, University of Toledo, Toledo, OH, USA*
[f]*Department of Psychology, University of California Irvine, Irvine, CA, USA*
[g]*Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA*
[h]*Department of Psychology, Boston University, Boston, MA, USA*
[i]*Department of Statistics, St Louis University, St Louis, MO, USA*
[j]*Research Service, VA St Louis Healthcare System, St Louis, MO, USA*

**Abstract**

**Introduction:** Longitudinal testing is necessary to accurately measure cognitive change. However, repeated testing is susceptible to practice effects, which may obscure true cognitive decline and delay detection of mild cognitive impairment (MCI).

**Methods:** We retested 995 late-middle-aged men in a ~6-year follow-up of the Vietnam Era Twin Study of Aging. In addition, 170 age-matched replacements were tested for the first time at study wave 2. Group differences were used to calculate the difference between practice effects and attrition effects. MCI diagnoses were generated from practice-adjusted scores.

**Results:** There were significant practice effects on most cognitive domains. Conversion to MCI doubled after correcting for practice effects, from 4.5% to 9%. Importantly, practice effects were present although there were declines in uncorrected scores.

**Discussion:** Accounting for practice effects is critical to early detection of MCI. Declines, when lower than expected, can still indicate practice effects. Replacement participants are needed for accurately assessing disease progression.



*Keywords:* Practice effects; Repeat testing; Serial testing; Longitudinal testing; Mild cognitive impairment; Cognitive change

## 1. Introduction

Longitudinal assessments are necessary for directly measuring cognitive change over time to track disease progression from cognitively normal to mild cognitive impairment (MCI) or MCI to Alzheimer's disease (AD), and for assessing efficacy of therapeutic interventions [1,2]. Because the pathological process begins decades before the onset of AD, it is widely agreed that early identification is of enormous importance [3]. However, repeat testing is susceptible to practice effects [4,5], and failure to account for practice effects may obscure cognitive declines and delay detection of conversion to MCI or AD [6,7].

The authors J.A.E. and A.J.J. contributed as joint first authors.
The authors have declared that no conflict of interest exists.
*Corresponding author. Tel.: (858) 534-6842.
E-mail address: jaelman@ucsd.edu

https://doi.org/10.1016/j.dadm.2018.04.003

Here, we assessed practice effects on neuropsychological testing and their impact on diagnosis of MCI in late-middle-aged adults. Practice effects are typically defined as improvements in performance due to prior exposure to a test as opposed to frank cognitive improvements [5,8]. However, in longitudinal studies, the absence of practice effects over time could signal cognitive decline as opposed to cognitive stability [9]. In midlife and later life, when normative declines are expected, stable performance, and even declines, could still reflect the contribution of practice effects. This situation is particularly problematic for studies that use age-based norms to diagnose individuals with cognitive impairment. Consider, for example, two individuals with similar characteristics who have identical cognitive test scores just above threshold for an MCI diagnosis, the only difference is that one individual is being tested for the first time, whereas the other has taken these tests before. We can infer that the second individual may actually have more impairment, but the effects of practice are artificially increasing their test scores, keeping them above threshold. This scenario would suggest that the individual may have dipped below the norm-based threshold and would have been diagnosed as having MCI had the test been taken for the first time. Clinically, this scenario is becoming more relevant as AD drug intervention trials shift toward secondary prevention strategies that rely on early identification. Failure to correct for practice effects may result in underdiagnosis or delays in detecting MCI.

Shorter test-retest intervals are significantly associated with an increased magnitude of practice effects [8,10,11]. However, practice effects have been found across intervals of over 5 years, and it has been estimated that it may take at least 7 years for practice effects to decrease to zero in adults aged 18–58 years [12,13]. Alternate test forms do not solve the problem because they do not fully remove practice effects [14,15] and they introduce test differences as yet another factor that affects performance. Selective attrition is an additional concern. Because returnees usually represent a healthier or higher performing subgroup, they would be expected to score higher at follow-up than the overall sample at baseline [1,16]. In case-control studies, the control group may be used to gauge practice effects, but that only allows for assessment of relative change in cases versus controls. If, however, there is no explicit control group and the goal is to determine when someone meets criteria for a diagnosis of MCI, cutoff scores for cognitive impairment at the point of diagnosis must be made. If there are practice effects, then those cutoff scores should be modified. The standard approaches for gauging practice effects do not provide a way to adjust the impairment threshold.

In contrast, inclusion of replacement participants—individuals who complete their baseline testing visit at the same time and age as the initial sample's follow-up visit—provides an optimal strategy to calculate practice and attrition effects [13]. Rönnlund et al. [13] used this approach to examine practice effects on tests of episodic and semantic memory. Here, we applied this approach to investigate practice effects across multiple cognitive domains over a 6-year interval during late midlife. We hypothesized that adjusting for practice effects would result in an increased rate of MCI at follow-up, suggesting that MCI cases are being missed when practice effects are not taken into account. In addition, because we previously found that different episodic memory tests showed different patterns of change across time [17], we examined whether practice effects contributed to these differences.

## 2. Methods

### 2.1. Participants

Participants were from waves 1 and 2 of the Vietnam Era Twin Study of Aging (VETSA) [18]. VETSA participants comprise a national, community-dwelling sample of male-male twins who are similar to American men in their age range with respect to health and lifestyle characteristics based on Center for Disease Control and Prevention data [19]. All served in the military sometime between 1965 and 1975, but nearly 80% reported no combat exposure. Detailed descriptions of the sample composition and method of ascertainment have been reported elsewhere [20,21].

The current analysis included 1220 individuals tested at wave 1 (mean age = 55.88 years, standard deviation [SD] = 2.5). Of these, 225 were dropouts and 995 (82%) returned for wave 2 (returnees; mean age = 61.54 years, SD = 2.4). At wave 2, 170 attrition replacements of a similar age as the returnees were tested for the first time (ARs; mean age = 62.67 years, SD = 2.3). ARs were recruited from the same twin registry as the other participants.

The study was approved by the institutional review boards at the participating institutions.

### 2.2. Demographics

The estimation of practice effects based on group means of returnees and ARs assumes the samples are well-matched on measures that may cause systematic differences in test performance. Although participants were recruited from a random sample, it is still possible that group differences exist. We found some small but significant differences between groups (see Table 1). The returnee group had a higher percentile score on the general cognitive ability (GCA) test that was administered at an average age of 20 years (61.4 vs. 54.3; $t_{(225.79)} = 3.765$, $P < .001$). These scores are approximately equivalent to IQ scores of 104.4 versus 101.6. Section 2.3 contains a description of the GCA test. The returnee group also had a higher average education (13.9 years vs. 13.4 years, $t_{(222.54)} = 3.765$, $P < .001$). The AR group was older than returnees (62.67 vs. 61.54; $t_{(238.98)} = 5.880$, $P < .001$). The percentage of apolipoprotein E ε4 (*APOE*-ε4) carriers was similar between groups (29.4% vs. 25.8%; $\chi^2(1) = 2.16$, $P = .142$).