



Avoid or Embrace? Practice Effects in Alzheimer's Disease Prevention Trials

Andrew J. Aschenbrenner^{1*}, Jason Hassenstab¹, Guoqiao Wang¹, Yan Li¹, Chengjie Xiong¹, Eric McDade¹, David B. Clifford¹, Stephen Salloway², Martin Farlow³, Roy Yaari⁴, Eden Y. J. Cheng⁴, Karen C. Holdridge⁴, Catherine J. Mummery⁵, Colin L. Masters⁶, Ging-Yuek Hsiung⁷, Ghulam Surti⁸, Gregory S. Day⁹, Sandra Weintraub¹⁰, Lawrence S. Honig¹¹, James E. Galvin¹², John M. Ringman¹³, William S. Brooks¹⁴, Nick C. Fox¹⁵, Peter J. Snyder⁸, Kazushi Suzuki¹⁶, Hiroyuki Shimada¹⁷, Susanne Gräber¹⁸ and Randall J. Bateman¹ for the Dominantly Inherited Alzheimer Network Trials Unit (DIAN-TU)

OPEN ACCESS

Edited by:

Daniel Nation,
University of California, Irvine,
United States

Reviewed by:

David Morgan,
Michigan State University,
United States
Sheila Black,
The University of Alabama,
United States
Daniel Gillen,
University of California, Irvine,
United States

*Correspondence:

Andrew J. Aschenbrenner
a.aschenbrenner@wustl.edu

Specialty section:

This article was submitted to
Alzheimer's Disease and Related
Dementias,
a section of the journal
Frontiers in Aging Neuroscience

Received: 24 February 2022

Accepted: 19 May 2022

Published: 16 June 2022

Citation:

Aschenbrenner AJ, Hassenstab J, Wang G, Li Y, Xiong C, McDade E, Clifford DB, Salloway S, Farlow M, Yaari R, Cheng EYJ, Holdridge KC, Mummery CJ, Masters CL, Hsiung G-Y, Surti G, Day GS, Weintraub S, Honig LS, Galvin JE, Ringman JM, Brooks WS, Fox NC, Snyder PJ, Suzuki K, Shimada H, Gräber S and Bateman RJ (2022) Avoid or Embrace? Practice Effects in Alzheimer's Disease Prevention Trials. *Front. Aging Neurosci.* 14:883131. doi: 10.3389/fnagi.2022.883131

¹ Washington University in St. Louis School of Medicine, St. Louis, MO, United States, ² Warren Alpert Medical School of Brown University, Providence, RI, United States, ³ Indiana University School of Medicine, Indianapolis, IN, United States, ⁴ Eli Lilly and Company, Indianapolis, IN, United States, ⁵ University College London, London, United Kingdom, ⁶ University of Melbourne, Melbourne, VIC, Australia, ⁷ The University of British Columbia, Vancouver, BC, Canada, ⁸ The University of Rhode Island, Kingston, RI, United States, ⁹ Mayo Clinic, Jacksonville, FL, United States, ¹⁰ Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, ¹¹ Columbia University Irving Medical Center, New York, NY, United States, ¹² Miller School of Medicine, University of Miami, Miami, FL, United States, ¹³ University of Southern California, Los Angeles, CA, United States, ¹⁴ Neuroscience Research Australia, University of New South Wales Medicine, Randwick, NSW, Australia, ¹⁵ Dementia Research Center, University College London, London, United Kingdom, ¹⁶ The University of Tokyo, Tokyo, Japan, ¹⁷ Osaka City University, Osaka, Japan, ¹⁸ German Center for Neurodegenerative Disease (DZNE), Tübingen, Germany

Demonstrating a slowing in the rate of cognitive decline is a common outcome measure in clinical trials in Alzheimer's disease (AD). Selection of cognitive endpoints typically includes modeling candidate outcome measures in the many, richly phenotyped observational cohort studies available. An important part of choosing cognitive endpoints is a consideration of improvements in performance due to repeated cognitive testing (termed "practice effects"). As primary and secondary AD prevention trials are comprised predominantly of cognitively unimpaired participants, practice effects may be substantial and may have considerable impact on detecting cognitive change. The extent to which practice effects in AD prevention trials are similar to those from observational studies and how these potential differences impact trials is unknown. In the current study, we analyzed data from the recently completed DIAN-TU-001 clinical trial (TU) and the associated DIAN-Observational (OBS) study. Results indicated that asymptomatic mutation carriers in the TU exhibited persistent practice effects on several key outcomes spanning the entire trial duration. Critically, these practice related improvements were larger on certain tests in the TU relative to matched participants from the OBS study. Our results suggest that the magnitude of practice effects may not be captured by modeling potential endpoints in observational studies where assessments are typically less frequent and drug expectancy effects are absent. Using alternate instrument forms (represented in our study by computerized tasks) may partly mitigate

practice effects in clinical trials but incorporating practice effects as outcomes may also be viable. Thus, investigators must carefully consider practice effects (either by minimizing them or modeling them directly) when designing cognitive endpoint AD prevention trials by utilizing trial data with similar assessment frequencies.

Keywords: practice effects, Alzheimer's disease, clinical trials, learning, assessment frequency, alternative forms

INTRODUCTION

Phase 3 secondary prevention clinical trials in Alzheimer's disease aim to demonstrate the efficacy of drug or other interventions in preserving or improving cognitive function in at-risk individuals. Such trials typically use the slowing of the rate of cognitive decline between a treatment arm and a placebo group as their primary efficacy endpoint (Sperling et al., 2014; Bateman et al., 2017; Cummings et al., 2020). Comprehensive neuropsychological test batteries are administered at regular intervals (e.g., every 6–12 months) to best characterize cognitive change across the course of the trial and to monitor for adverse events such as unexpected drops in performance. However, these repeated administrations may have unanticipated consequences for trial outcomes. Specifically, it is well-known that healthy adults typically improve in performance (termed “practice effects” or “PEs”) with repeated cognitive testing (Calamia et al., 2012). These PEs can be attributed to several factors including increased familiarity with task procedures, development of testing strategies, or memorization of specific stimuli. These gains are not limited to short time intervals and can persist for as long as 7 years (Salthouse et al., 2004) after just one exposure, a longer time span than a typical AD prevention trial. It is also important to consider that in symptomatic AD populations, where active neurodegenerative processes drive worsening cognitive performance, practice effects do not always translate to better performance from visit to visit. Rather, the competing forces of disease and PEs can manifest as attenuations of decline such that PEs may be observable as flat or simply less negative slopes.

For these reasons, potential PEs must be taken into consideration when planning a clinical trial. The two primary analytical models used in AD trials either analyze change from baseline to final test (e.g., mixed models for repeated measures or MMRM) or conceptualize change as linear from baseline to end of study (random intercept and slope models). When PEs are present but unaccounted for in statistical models, the magnitude of decline over the course of the trial can be drastically underestimated (Hassenstab et al., 2015; Jacobs et al., 2017) reducing the power to detect a treatment effect. Therefore, it may be desirable to minimize the influence of PEs in a clinical trial. One way to do so would be to include multiple “screening” sessions (Goldberg et al., 2015) which give participants experience with the cognitive battery prior to the initiation of treatment, as PEs tend to be largest after the first or second retest (Collie et al., 2003; Bartels et al., 2010). Other methods for minimizing PEs include the use of alternate forms, although this presents the additional challenge of verifying that the different forms are truly psychometrically

equivalent (Gross et al., 2012), and yet still limit PEs due to familiarity. Computerized cognitive assessments, depending on the test paradigms, can protect against form-related PEs by randomly selecting stimuli for each test administration, creating an essentially endless number of alternate forms. But of course, this requires additional equipment and study management that can be costly and may not suit all trial protocols. Importantly, none of these approaches are entirely successful at eliminating practice effects (Beglinger et al., 2005; Falsetti et al., 2006). Given the difficulties with eliminating PEs in cognitive studies, some studies have turned away from efforts at avoiding PEs opting instead to determine if incorporating PEs as outcomes themselves may reveal meaningful information about cognitive status. For example, several studies have shown that the attenuation of PEs in clinically healthy older adults can predict important outcomes such as biomarker status or risk of progression to symptomatic AD (Duff et al., 2011; Hassenstab et al., 2015; Machulda et al., 2017; Oltra-Cucarella et al., 2018; Samaroo et al., 2020). PEs may therefore serve as a subtle marker of early disease even if average cognitive trajectories are relatively flat. It is critical, therefore, to have a comprehensive understanding of factors that produce or exaggerate practice effects and to develop statistical tools to appropriately model them. Ultimately, the magnitude of PEs may serve as an alternative or supplementary endpoint for trials.

Similar to clinical trials, observational studies of AD provide systematic and longitudinal assessment of clinical, cognitive and pathological progression of the disease, albeit in the absence of a specific intervention. Although PEs have been relatively well studied in community-based observational studies of sporadic AD, to date, we are unaware of any systematic evaluation of PEs in the context of a clinical trial. One might expect that PEs would be attenuated in clinical trials if the study protocol includes a comprehensive screening assessment, which may provide exposure to the testing materials (Goldberg et al., 2015). Alternatively, in some cases, trial participants might be recruited from ongoing observational studies and hence are already familiar with the process of cognitive testing and may have exposure to the same test materials. Another important difference from observational studies is the role of participant expectations in clinical trials. Trial participants may exhibit enhanced PEs due to a type of placebo effect, wherein motivation and engagement may be higher in the trial compared to observational studies where expectations and motivations for participation may be different. As many trials rely on data from observational studies to select appropriate cognitive measures as endpoints and conduct power analyses to determine the requisite sample sizes needed to detect a hypothetical treatment effect, it is critical to test the assumption that participants in observational studies will perform similarly to those engaged in clinical trial

research. If these two populations differ in terms of PEs or overall cognitive trajectories, pre-specified cognitive endpoints selected based on observational study data may not be suitable for a clinical trial and sample sizes may be underestimated, among other concerns.

To address these issues, we present analyzes from the recently completed DIAN-TU 001 (TU) clinical trial (Mills et al., 2013) and the associated DIAN Observational study (OBS, Bateman et al., 2012). The DIAN-TU is a phase 2/3, double blind, placebo controlled study of disease modifying therapies in autosomal dominant AD (ADAD), a rare form of AD due to specific genetic mutations that has similar pathological and clinical presentations, other than in age at onset, as sporadic AD (Bateman et al., 2011). These genetic mutations cause AD with virtually 100% penetrance and onset of clinical symptoms begin at a predictable and typically much younger age than sporadic AD (Ryman et al., 2014). The expected number of years to symptom onset (EYO) can be calculated based on the participant's age and the historical average age-at-symptomatic onset of gene-carriers with the same mutation or from the same family. The predictability of expected symptom onset as well as pathological similarities to the more common sporadic form of AD, makes ADAD a critical population in which to understand and build a model of cognitive, clinical, and pathological disease progression (McDade et al., 2018). To maintain participant blinding to their mutation status, ADAD mutation carriers (MCs) and non-carriers (NMCs) were enrolled in the trial, with all NMCs being assigned to placebo in a double blinded manner. The DIAN Observational study was launched in 2008 to provide natural history data on the progression of clinical, cognitive, and pathological changes in this population. Several participants who enrolled in the OBS study later enrolled in the TU study. We utilized the data from these two studies to answer the following questions: (1) Do PEs in ADAD vary as a function of mutation status or clinical status? (2) Do alternate forms that vary the stimuli across repeated administration (computerized battery vs. pen and paper) moderate the size of PEs? and (3) Do cognitive trajectories in clinical trials differ from those in observational only studies?

MATERIALS AND METHODS

A total of 384 participants were included in our analyzes. One-hundred ninety-three participants from the TU cohort and 191 from the OBS cohort. Both studies recruited a population of ADAD mutation carriers and non-carriers to determine the natural history (OBS) and to implement safe, efficient, and effective clinical trials that have the highest likelihood of success in advancing overall treatment (TU). Although the TU study was not powered to determine cognitive effects at the higher treatment doses that were ultimately used (5% power to detect a 30% slowing in the rate of cognitive decline), we have previously shown the absence of a treatment effect on cognitive outcomes in the TU (Salloway et al., 2021). Thus, given the relatively small group differences between treatment and placebo arms, for the present analyzes, all participants were combined and treatment arm [e.g., drug (solanezumab/gantenerumab) vs. placebo] was

not considered. A small number of NMCs had clinical evidence for impairment (3 in the TU and 7 in OBS), these participants were removed prior to analysis due to small sample size, leaving a total of 374 participants available for analysis.

Clinical/Cognitive Evaluation

Participants in both the TU and OBS studies underwent comprehensive clinical and cognitive evaluations. Presence and severity of dementia symptoms was ascertained using the Clinical Dementia Rating® (CDR) scale (Morris, 1993). A global rating of 0 on the CDR reflects no dementia, while scores of 0.5, 1, 2, and 3 reflect very mild, mild, moderate, and severe dementia, respectively. The Mini-Mental State Exam (Folstein et al., 1975) (MMSE) was also given as a measure of general cognitive function.

The cognitive batteries were largely similar across the two studies. Neuropsychological tests that were given in common across the two cohorts have been described elsewhere (Storandt et al., 2014) and include Wechsler Memory Scale-Revised Logical Memory Immediate and Delayed Recall (Wechsler, 1987) and Digit Span, Trail making Parts A and B (Armitage, 1945), Category Fluency for Animals and Vegetables (Goodglass and Kaplan, 1983), and Digit Symbol Substitution from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981). In the TU, participants were also administered the Cogstate computerized battery which included Identification, Detection, One-Back, One Card Learning, and the International Shopping List test. These measures have been described extensively elsewhere (Hammers et al., 2011; Lim et al., 2012). In the TU, most of these tests were administered every 6 months except for category fluency and the MMSE which were measured annually. All tests utilized the same versions at each testing occasion with the exception of the Cogstate tests which produced randomly generated stimuli at each occasion. Assessment frequency in the OBS study ranged from every 1–3 years depending on clinical status and when the participant entered the study. The OBS study has enrolled over 575 participants to date, but for the purpose of these analyzes, we selected participants that matched the enrollment criterion for the TU. We included as many participants as possible who met the following criteria: baseline global CDR score of 1 or less and estimated years to EYO range from –15 to +10 years (Salloway et al., 2021; See **Table 1** for full demographics). For the purposes of these analyzes, participants who were initially enrolled in the OBS study and then transitioned to the TU (41% of the TU CDR 0 carriers, 32% of the TU CDR > 0 carriers and 33% of the TU non-carriers started in the OBS study) were included in the TU cohort but were excluded from analyzes in the OBS cohort.

Statistical Analysis

Our analyzes proceeded in several steps. We first compared cognitive trajectories in the TU battery between NMCs, CDR 0 MCs and CDR > 0 MCs. We constructed linear mixed effects (LME) models for each cognitive test and predicted cognition from baseline EYO, time-in-study (hereafter referred to as “time”), group and the group by time interaction. A random intercept and random slope of time was also included in

TABLE 1 | Demographic characteristics of the clinical trial (TU) and observational (OBS) study cohorts.

	DIAN-TU			DIAN Obs		
	NMC	MC CDR 0	MC CDR > 0	NMC	MC CDR 0	MC CDR > 0
N	46	85	59	115	35	34
Age	42.0 (9.2)	40.9 (8.5)	49.2 (10.1)	41.3 (8.9)	38.7 (9.5)	46.0 (8.3)
EYO	-4.5 (6.3)	-5.8 (6.3)	2.7 (4.8)	-6.1 (6.8)	-8.3 (6.0)	1.2 (3.9)
Sex (% female)	20 (43%)	45 (53%)	28 (47%)	70 (61%)	25 (71%)	22 (65%)
Education	15.5 (3.2)	15.6 (3.2)	14.1 (2.6)	14.9 (2.8)	14.3 (2.7)	13.2 (3.2)
Number of assessments	7.3 (3.6)	9.5 (2.2)	8.1 (2.5)	2.2 (1.3)	2.7 (1.1)	3.3 (1.2)
Length of follow-up	3.1 (1.8)	4.2 (1.1)	3.6 (1.3)	2.5 (2.5)	3.6 (1.9)	2.9 (1.6)

Results are reported as mean (SD) where appropriate.

all models with an unstructured covariance matrix. Follow-up contrasts were constructed to compare slopes on each test between the NMCs and the CDR 0 MCs, and between the CDR 0 MCs and the CDR > 0 MCs. For ease of comparison across tests, all outcomes were z-scored to the baseline mean and standard deviation of the CDR 0 non-carriers so that a score of “0” represents the score of a relatively cognitively normal participant. Scores were oriented such that a positive slope indicates an improvement over time and a negative slope indicates decline.

A second set of LMEs were constructed to compare performance in the TU vs. the OBS study. Specifically, we analyzed performance on each cognitive test as a function of time, group (NMC, CDR 0 MC, and CDR > 0 MCs) and cohort (TU vs. OBS), and included all of the two and three-way interactions while also controlling for baseline EYO. All models were fit in the R statistical computing software (version 4.0.5, R Core Team, 2021) using the lme4 package (version 1.1.27.1, Bates et al., 2015). *P*-values were obtained using the lmerTest (version 3.1.3, Kuznetsova et al., 2017) package. To ensure that no influential, outlying data points were unduly biasing our results, we used the influence.ME package (Nieuwenhuis et al., 2012) to iteratively remove a single participant from each model and re-run the statistical analysis. We checked for a change in statistical significance in key model parameters (specifically, the group by time or group by cohort by time interactions) when a given participant was removed. Across all the analyses we conducted, none of those parameters changed significance suggesting no single person was exerting undue influence on these results. Finally, although a relatively large set of statistical comparisons were conducted in order to fully describe practice effects across a range of cognitive tests, no corrections for multiple comparisons were made.

RESULTS

Analysis 1: Clinical Trial Only

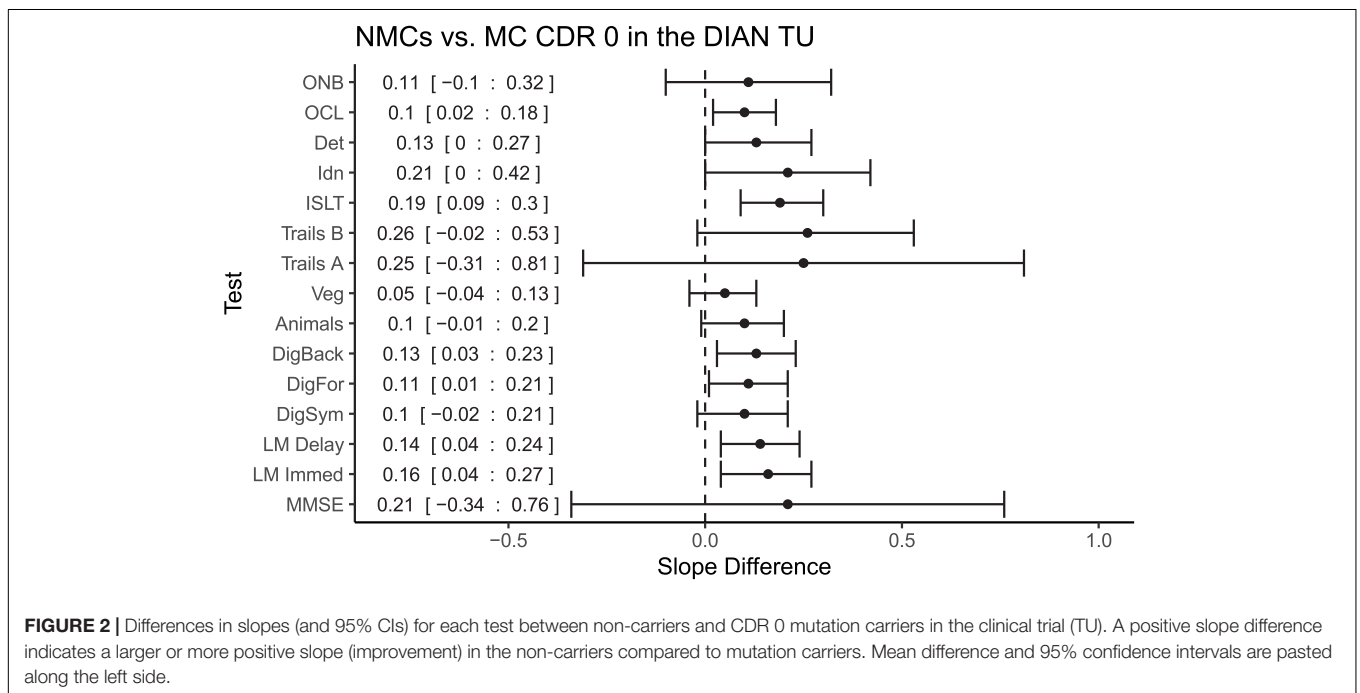
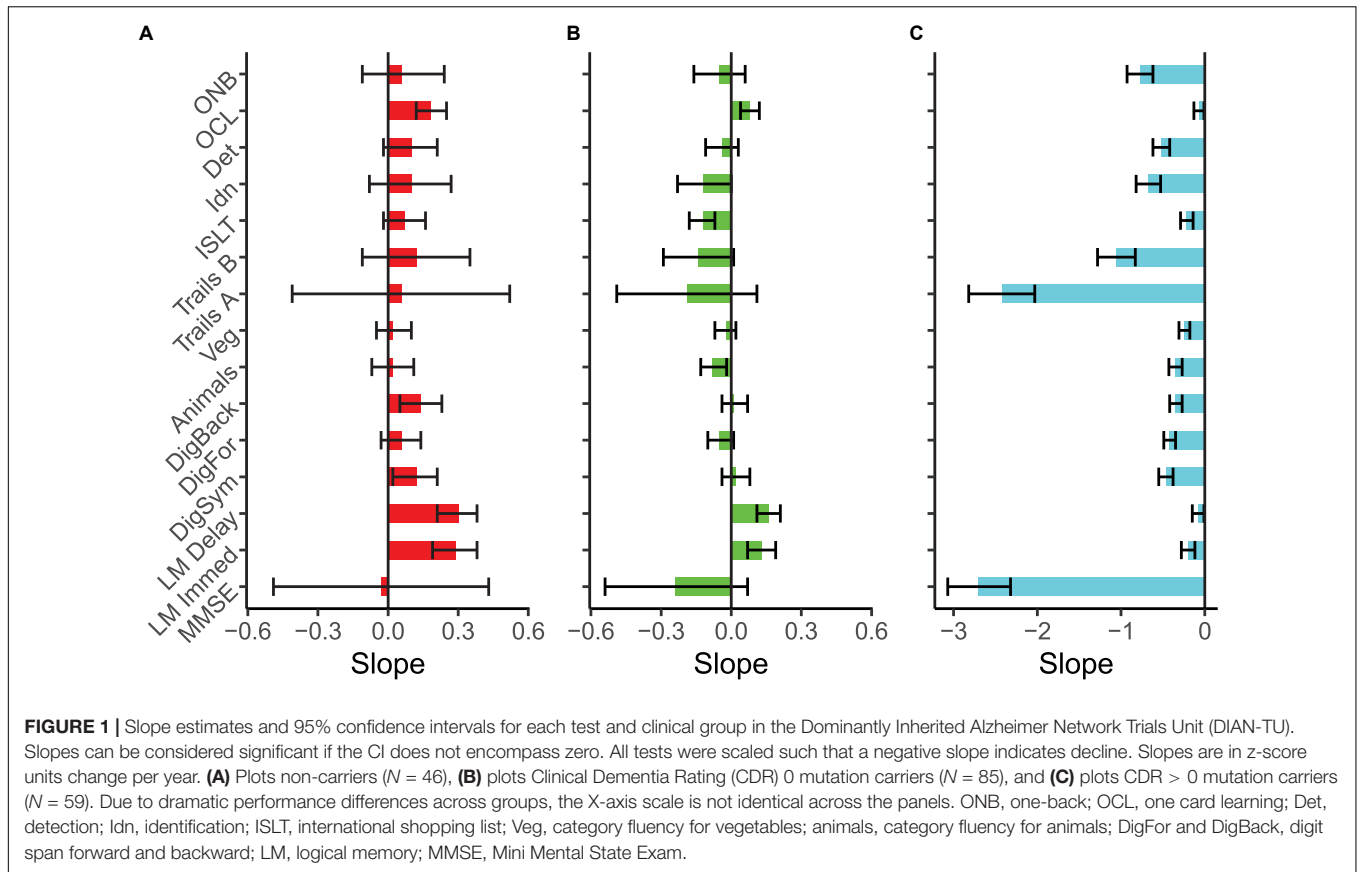
Slopes over time for each cognitive test and each group are illustrated in **Figure 1**. Intercepts and slope scores for each test can also be found in **Supplementary Table 1**. Not surprisingly, the MC CDR > 0 group evinced significant decline on all cognitive measures with some of the largest effects occurring on tests of perceptual speed and attention (Cogstate Detection,

Identification and One back, Digit Symbol Substitution and Trail Making Part A). In contrast, cognitive trajectories for the MC CDR 0 group were relatively flat with a few notable exceptions. There was significant decline on Category Fluency for Animals, the ISLT and the Identification test, suggesting that measures of semantic fluency, episodic memory and attention are sensitive to preclinical cognitive decline. Interestingly, the Logical Memory immediate and delayed recall tests showed significant improvement over time in this population as did Cogstate One Card Learning, a test of visual learning ability. NMCs did not decline on any measure, which was expected in a relatively young and cognitively healthy cohort. Showing the classic pattern of practice effects, NMCs exhibited significant improvement over time compared to a zero slope on several measures including Logical Memory Immediate and Delayed Recall, Digit Symbol Substitution, Digit Span Backward and One Card Learning.

In order to determine disease effects on learning and decline, we next compared slopes between the NMCs and the MC CDR 0 group (shown in **Figure 2**) to determine if differences in rate of change distinguished the groups. Slopes (reflecting change per year in z-score units) were significantly different between these two groups on the following measures: One Card Learning (Difference = 0.10, *p* = 0.01, *CI* = 0.02:0.18), Logical Memory Immediate (Difference = 0.16, *p* = 0.009, *CI* = 0.04:0.27), Logical Memory Delayed (Difference = 0.14, *p* = 0.007, *CI* = 0.04:0.24), Digit Span Forward (Difference = 0.11, *p* = 0.04, *CI* = 0.006:0.21), Digit Span Backward (Difference = 0.13, *p* = 0.02, *CI* = 0.03:0.23) and the ISLT (Difference = 0.19, *p* < 0.001, *CI* = 0.08:0.30). These results indicate that while both MCs and NMCs exhibited PEs (see **Figure 1**) on the Logical Memory and One Card Learning tests, practice-related gains were significantly larger in the NMCs. Moreover, NMCs improved over time on the Digit Span Backward test whereas the MCs showed no significant change. Finally, the NMCs did not show improvement or decline on ISLT whereas the MCs significantly declined.

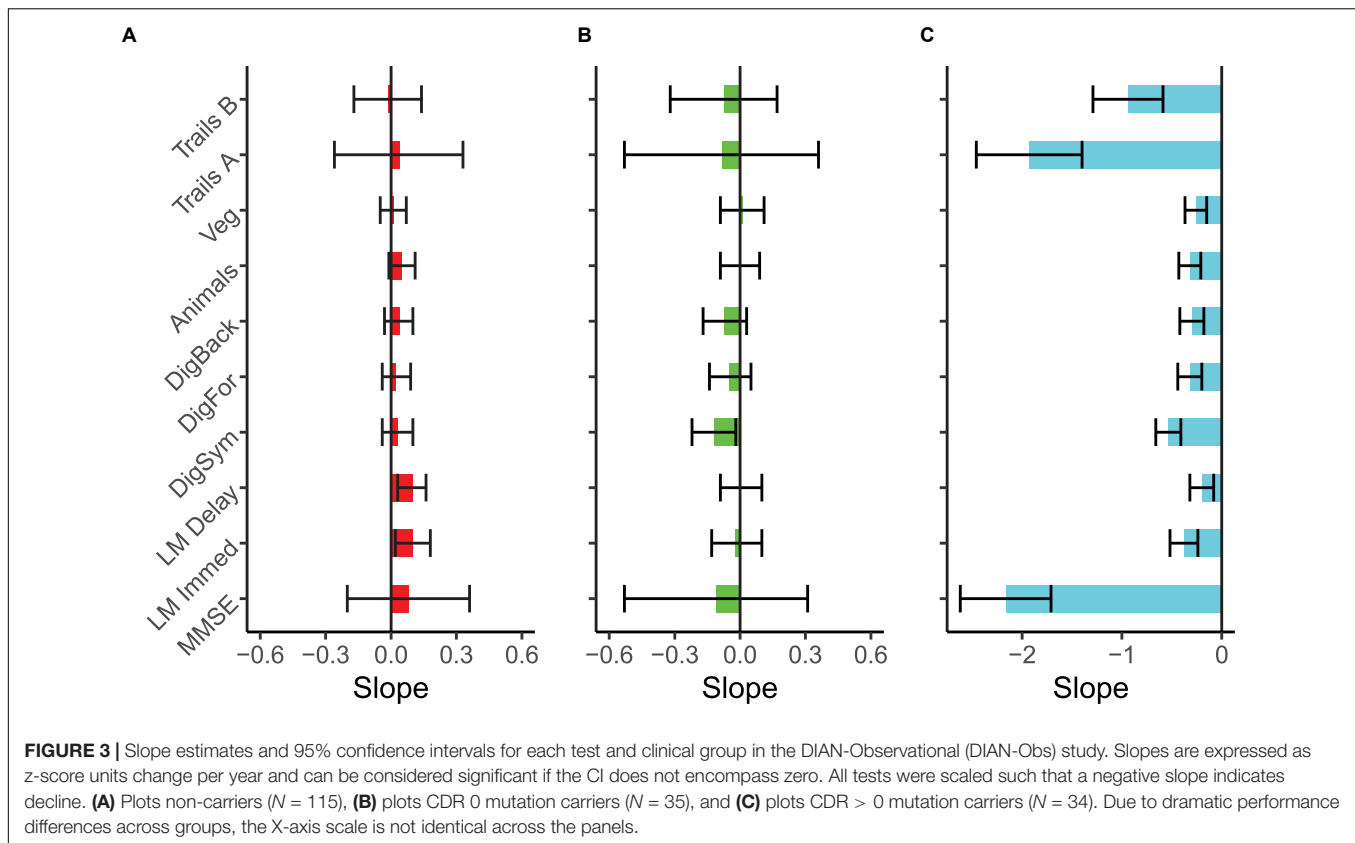
Analysis 2: Observational Versus Clinical Trial

Intercepts and slopes for the eligible participants in the OBS study are provided in **Supplementary Table 2**, and slopes for each test and group are plotted in **Figure 3**, showing time-dependent changes. First, similar to the TU, the MC CDR > 0



group in the observational study declined significantly on all measures. Second, the MC CDR 0 group again showed relatively flat cognitive trajectories with the notable exception of the Digit

Symbol Substitution test which significantly declined by 0.12 z-score units per year. Most importantly, there was no hint of practice related improvements in the MC CDR 0s, with lack



of evidence of positive slope estimates, on any of the cognitive measures. Finally, the NMC group significantly improved on the Logical Memory Immediate and Delayed Recall tests but the slopes for the other measures were relatively flat and not significantly different from zero.

Direct comparisons between the symptomatic MCs in the OBS and TU cohorts (**Figure 4**), revealed no significant differences in slopes between the cohorts on any measure with the exception of Logical Memory Immediate Recall (Difference = 0.18, $p = 0.03$, $CI = 0.02:0.34$), in which participants in the TU showed slightly less decline than in OBS. Interestingly, a number of differences emerged when comparing the asymptomatic MCs across TU and OBS (**Figure 5**). Specifically, on the Digit Symbol Substitution test (Difference = 0.14, $p = 0.02$, $CI = 0.02:0.25$), Logical Memory Immediate (Difference = 0.15, $p = 0.03$, $CI = 0.02:0.28$) and Delayed recall (Difference = 0.15, $p = 0.007$, $CI = 0.04:0.26$) slopes were markedly less negative in the TU as compared to the OBS study. Finally, in the comparison of NMCs (**Figure 6**), the OBS participants improved less on Logical Memory Immediate (Difference = 0.19, $p = 0.003$, $CI = 0.07:0.31$) and Delayed recall (Difference = 0.20, $p < 0.001$, $CI = 0.10:0.30$) compared to the TU participants.

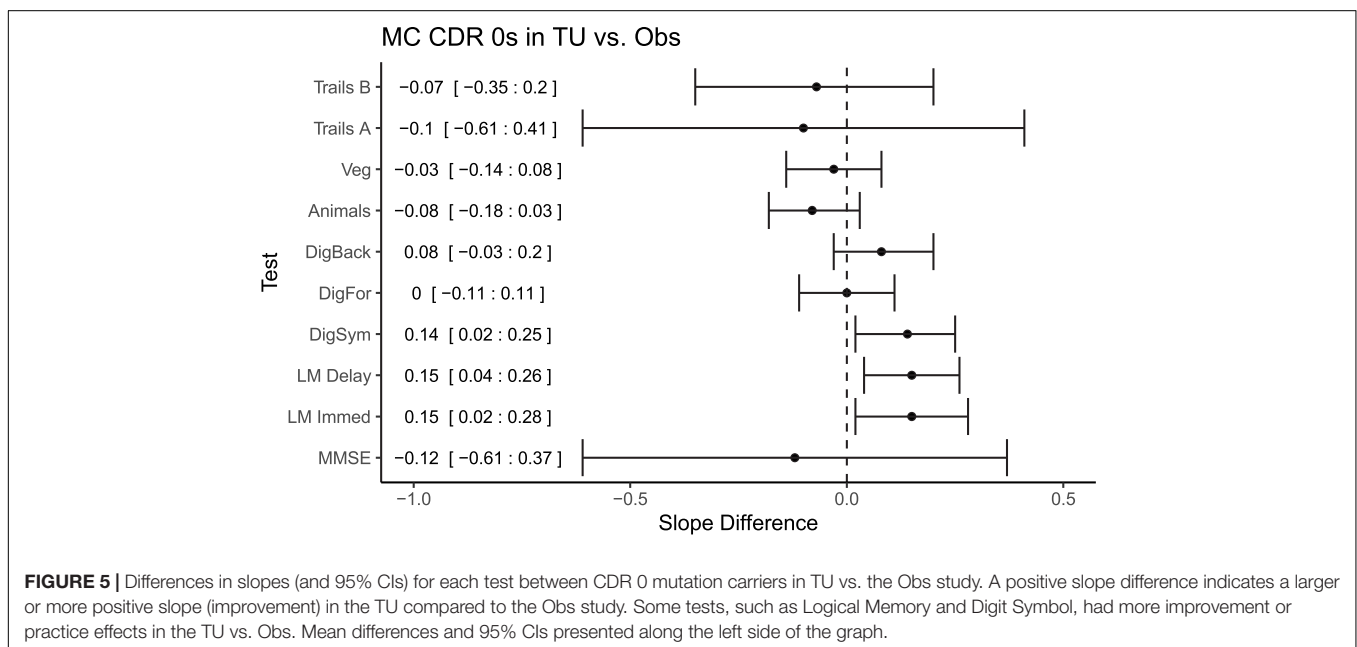
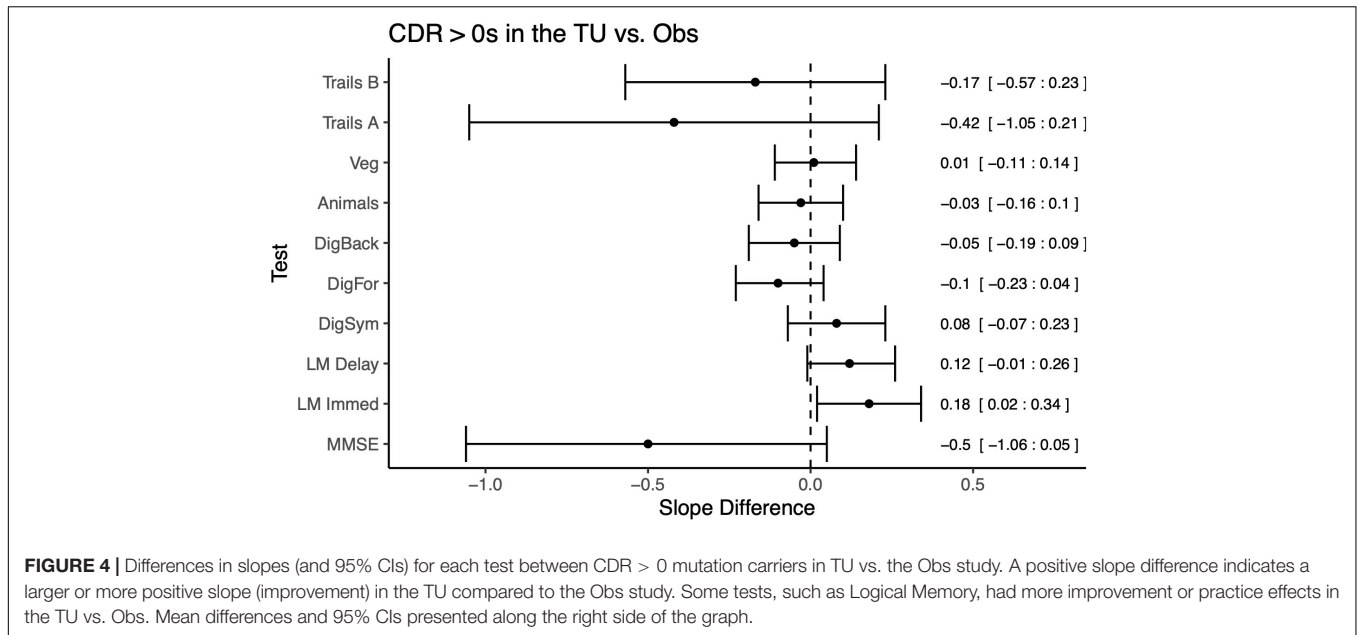
DISCUSSION

In this study, we compared performance on a comprehensive cognitive battery in two cohorts to answer several important

questions regarding practice related improvements in observational studies and clinical trials in AD populations.

Question 1: Does Mutation Status or Clinical Status Moderate Practice Effects in the Dominantly Inherited Alzheimer Network Trials Unit?

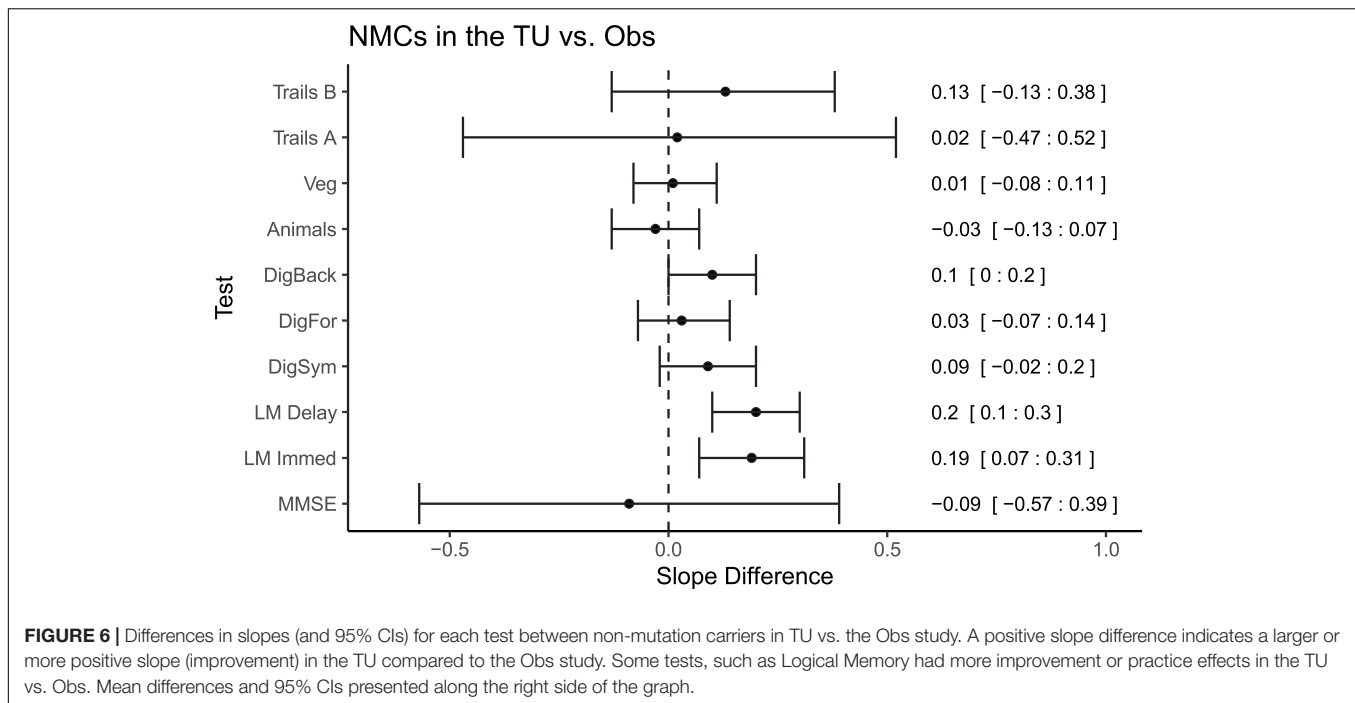
Clinical status was an important predictor of PEs in the DIAN-TU. Specifically, individuals who were CDR > 0 at entry significantly declined on all cognitive measures and therefore did not show practice-related gains. This is not to say that PEs were not present in this group, only that any gains associated with practice were overshadowed by the decline attributable to AD pathology. More importantly, mutation status in the CDR 0 groups also predicted magnitude of change in the TU. MC CDR 0s declined significantly over time on measures of attention, episodic memory, and semantic fluency whereas NMCs showed no change in these domains. Interestingly, differences in performance between MC CDR 0s and NMCs also emerged on the ISLT (list recall, MCs declined more than NMCs), Logical Memory (narrative recall, MCs improved less than NMCs), Digit Span (working memory, MCs improved less than NMCs), and One Card Learning (visual learning, MCs improved less than NMCs). Together these findings suggest that differences in the magnitude of practice related improvements in domains of memory and learning might serve as a sensitive and supplemental indicator of preclinical AD.



Question 2: Do Alternative Forms Influence Practice Effects?

We expected *a priori* that computerized measures from the Cogstate battery might show less practice effects due to the nature of randomized stimuli which generates essentially unlimited alternate forms. For example, many of these tasks use playing cards as stimuli presented in a newly randomized order at each administration. Such a design reduces the possibility of memorizing specific items which can be a contributor to PEs. This contrasts with Logical Memory in the DIAN studies, for example, which presents the same narrative each time the test is taken.

Our hypothesis was largely supported. Most of the computerized tests were resistant to practice effects in the NMCs or sensitive to decline in the MCs (e.g., ISLT and Identification tests). Practice related gains were apparent on the One Card Learning test and due to the nature of the randomized stimuli, it is assumed that participants are developing or learning some strategy besides rote memorization to improve over time. One possibility is that this test might be particularly amenable to visual strategies such as the method of loci (Gross et al., 2014). As the cards are shown one at a time, participants may over time learn to organize the items in a meaningful fashion (e.g., into poker hands or by suit) which might aid recall.



Question 3: Are Practice Effects Similar Across Clinical Trials and Observational Studies?

One of the most important questions addressed in this study was whether cognitive trajectories were similar across a clinical trial cohort and an observational study. For participants who were $CDR > 0$ at baseline, the answer was clearly “yes”. Regardless of the cohort, MC $CDR > 0$ s declined significantly over time and the magnitude of change did not differ significantly between the TU and OBS with the sole exception of Logical Memory Immediate Recall. This may reflect disease progression such that symptomatic MCs have declined to the extent that any practice related gains were outweighed by the task demands. An interesting but complex question for future studies is to determine the point at which PEs are effectively overwhelmed by disease related declines.

For the MC $CDR 0$ s, however, a few critical differences did emerge. Specifically, OBS participants declined at a faster rate than the TU participants on the Digit Symbol Substitution test and improved less on the Logical Memory Immediate and Delayed Recall tests. One obvious possible explanation for these differences is the assessment frequency across the two studies (every 6 months in the TU, \sim every 2 years in OBS). This explanation is likely for the Logical Memory tests, where participants will hear the same story at each testing occasion which reinforces encoding and aids in recall. It is less clear why Digit Symbol Substitution would show such enhanced practice effects in the TU when other measures of speed and executive function did not (e.g., the Trail Making tests). Studies of retest have shown performance gains on Digit Symbol Substitution, but this test typically demonstrates less gains than episodic memory

measures (Calamia et al., 2012). Thus, frequency of assessment needs to be carefully considered during trial design.

Another important possibility is an enhanced placebo effect in the DIAN-TU. Specifically, TU participants were randomized to treatment vs. placebo at a ratio of 3:1. Thus, there may have been a greater expectation of being on active drug which may have then impacted cognitive performance. Regardless of the underlying mechanisms, these differences in practice related gains are particularly noteworthy as the Logical Memory and Digit Symbol tests feature heavily in multiple cognitive composite endpoints (Sperling et al., 2014; Bateman et al., 2017). Investigators should keep in mind potential differences between observational and trial cohorts when planning their studies and conducting power analyses.

Using alternate instrument forms has been shown in some studies to be a viable strategy to reduce PEs. For example, a meta-analysis of test/retest effects found substantial reductions in performance gains when alternate forms were used for verbal list learning measures (Calamia et al., 2012). This finding is similar to the results shown here, in which the computerized tests were largely resistant to practice-related gains. The one exception we found was One Card Learning, a visual learning test that uses randomly generated sequences of cards such that there are essentially hundreds of alternate forms. This task produced the largest PEs in asymptomatic MCs enrolled in the DIAN-TU clinical trial. We could not, however, determine if this was due solely to clinical trial participation, as this measure was not collected with sufficient samples in the OBS study for comparison. In a recent study, the developers of the One Card Learning test made a shorter and less difficult version of the test (as evidenced by less floor effects in symptomatic AD participants) that demonstrated no PEs in young cognitively

normal participants across very short retest intervals (White et al., 2021). The authors argue that the increased difficulty and length of the longer version of the task may lead to participants forming strategies that in turn lead to more PEs.

Although our results indicate that rates of change on key cognitive outcomes may be underestimated in clinical trials due to the presence of these practice effects, it is important to highlight situations in which these practice effects might limit the ability to detect treatment effects. Specifically, in clinical trials that include a placebo arm in which participants undergo identical clinical and cognitive assessments as participants on active treatment, the negative impact of practice effects may be minimal, to the extent that practice effects manifest similarly in placebo vs. treated patients. However, this also assumes that the influence of improved cognition due to treatment is additive, rather than interactive, with improved cognition due to practice effects, which may not be the case. Moreover, the primary cognitive outcome is often a composite score formed of multiple tests. If some tests exhibit practice effects while others do not, as is the case in the present study, decline on a global composite score may be very small, limiting the power to detect any differences among groups.

It is unclear if attempts to avoid or reduce practice effects are futile. Completely avoiding practice effects does seem an impossible task. One of the most fundamental aspects of human behavior is adaptation, or learning. As we and others have previously shown, in the context of a cognitive assessment this learning is not just limited to familiarity with test materials but also to process factors like test strategies, effort, demand characteristics, and expectancy effects, among others (Beglinger et al., 2005; Hassenstab et al., 2015; Machulda et al., 2017). Instead of avoiding PEs, trials that enroll cognitively normal or mildly affected participants might consider designs and statistical models that anticipate and account for the influence of PEs. Such protocols might include extended baseline designs that allow cluster assessments prior to dosing in so-called “run-in” designs (Frost et al., 2008). Less emphasis might be placed on spreading assessments out at regular time intervals (e.g., one assessment every 6 months) in favor of clustering assessments at key read-out times and averaging across the clusters, which might not only minimize the effects of practice but also reduce individual variability in scores (Valdes et al., 2016). An alternative strategy is to incorporate PEs as outcomes themselves. Several recent studies have deliberately measured learning effects in cognitively normal older adults at risk for AD (Hassenstab et al., 2015; Baker et al., 2020; Lim et al., 2020; Samaroo et al., 2020). Effect sizes differentiating participants with biomarker-confirmed preclinical AD from those with normal biomarker levels are extraordinarily large for these paradigms, suggesting that PEs may be a highly sensitive indicator of disease progression.

There are many strengths to this study including use of a comprehensive cognitive battery on very well-characterized clinical cohorts, designed comparability between an observational study and clinical trial, enrolling the same population for both studies, and frequent assessments over many years. However, some limitations need to be noted. First, because

this is a study of ADAD, a very rare form of AD, the sample sizes included here could be considered small. Moreover, it is unclear whether differences in practice related gains will translate to the more common sporadic form of the disease. Second, some participants in these studies may become aware of their mutation status and this might alter their cognitive outcomes (Aschenbrenner et al., 2020). It is unknown whether the number of participants who did and did not learn their status were similar across the two studies. Third, we did not have data from the Cogstate testing battery in the DIAN-OBS study which precluded a comparison of PEs between the trial and observational study on these measures. Finally, we conducted many statistical tests due to the large cognitive battery that was administered and although many effects could have been predicted *a priori* this could be seen as an additional limitation.

Nevertheless, these results highlight three important points. (1) Practice effects were highly evident in the DIAN-TU-001 clinical trial in asymptomatic mutation carriers and non-carriers. (2) Alternate forms may have attenuated practice effects, but not for all measures. (3) The magnitudes of practice effects were larger in the DIAN-TU-001 clinical trial than seen in a well-matched sample from the DIAN Observational study, suggesting that more frequent assessments and placebo effects in clinical trials may drive increases in practice effects. Clinical trials that utilize a cognitive endpoint should carefully consider the potential for practice effects and select statistical modeling strategies that can incorporate them directly.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because risk of identifying individual participants and/or risk to ongoing trial activities. Requests to access the datasets should be directed to <https://dian.wustl.edu/our-research/for-investigators/diantu-investigator-resources>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Local Ethics Committees at TU sites. The patients/participants provided their written informed consent to participate in this study.

DOMINANTLY INHERITED ALZHEIMER NETWORK TRIALS UNIT (DIAN-TU)

Data used in the preparation of this article were obtained from the Dominantly Inherited Alzheimer Network Trials Unit (DIAN-TU). As such, the study team members within the DIAN-TU contributed to the design and implementation of DIAN-TU and/or provided data but may not have participated in the analysis or writing of this report. A complete listing of the DIAN-TU Study Team Members can be found at dian.wustl.edu, DIAN-TU Study Team.

AUTHOR CONTRIBUTIONS

AA statistical analysis, development of study hypotheses, and initial draft of the manuscript. JH development of study hypotheses, acquisition of study data, and revised manuscript for intellectual content. GW, YL, CX, EC, and KH revised manuscript for intellectual content. EM, DC, SS, MF, RY, CJM, CLM, G-YH, GS, GD, SW, LH, JG, JR, WB, NE, PS, KS, HS, SG, and RB acquisition of study data and revised manuscript for intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under award numbers U01AG042791 and U01AG042791-S1 (FNIH and Accelerating Medicines Partnership), R1AG046179 and R01AG053267-S1. The research for the DIAN-TU-001 trial, solanezumab and gantenerumab drug arms was also supported by the Alzheimer's Association, Eli Lilly and Company, F. Hoffman-LaRoche Ltd., Avid Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly and Company), GHR Foundation, an anonymous organization, Cogstate, and Signant. The DIAN-TU has received funding from the DIAN-TU Pharma Consortium. We acknowledged the altruism of the participants and their

REFERENCES

- Armitage, S. G. (1945). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychol. Monogr.* 60, 1–48.
- Aschenbrenner, A. J., James, B. D., McDade, E., Wang, G., Lim, Y. Y., Benzinger, T. L. S., et al. (2020). Awareness of genetic risk in the Dominantly Inherited Alzheimer Network (DIAN). *Alzheimers Dement.* 16, 219–228. doi: 10.1002/alz.12010
- Baker, J. E., Bruns, L., Hassenstab, J., Masters, C. L., Maruff, P., and Lim, Y. Y. (2020). Use of an experimental language acquisition paradigm for standardized neuropsychological assessment of learning: a pilot study in young and older adults. *J. Clin. Exp. Neuropsychol.* 42, 55–65. doi: 10.1080/13803395.2019.1665626
- Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., and Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci.* 11:118. doi: 10.1186/1471-2202-11-118
- Bateman, R. J., Aisen, P. S., De Strooper, B., Fox, N. C., Lemere, C. A., Ringman, J. M., et al. (2011). Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimers Res. Ther.* 3:1. doi: 10.1186/alzrt59
- Bateman, R. J., Benzinger, T. L., Berry, S., Clifford, D. B., Duggan, C., Fagan, A. M., et al. (2017). The DIAN-TU Next Generation Alzheimer's prevention trial: adaptive design and disease progression model. *Alzheimers Dement.* 13, 8–19. doi: 10.1016/j.jalz.2016.07.005
- Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., et al. (2012). Clinical and biomarker changes in Dominantly Inherited Alzheimer's disease. *N. Engl. J. Med.* 367, 795–804. doi: 10.1056/NEJMoa1202753
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Beglinger, L., Gaydos, B., Tangphaodaniels, O., Duff, K., Kareken, D., Crawford, J., et al. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Arch. Clin. Neuropsychol.* 20, 517–529. doi: 10.1016/j.acn.2004.12.003

families and contributions of the DIAN, DIAN Expanded Registry, and DIAN-TU research and support staff at each of the participating sites (see DIAN-TU Study Team) for their contributions to this study.

ACKNOWLEDGMENTS

This manuscript has been reviewed by DIAN-TU Study investigators for scientific content and consistency of data interpretation with previous DIAN-TU Study publications. David Holtzman, former Department Head of Neurology where the research was conducted, an inventor on patents for one of the treatments (solanezumab), which has been tested in the DIAN-TU clinical trials. If solanezumab is approved as a treatment for Alzheimer's disease or Dominantly Inherited Alzheimer's Disease, Washington University and Holtzman will receive part of the net sales of solanezumab from Eli Lilly, which has licensed the patents related to solanezumab from Washington University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.883131/full#supplementary-material>

- Calamia, M., Markon, K., and 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
- Collie, A., Maruff, P., Darby, D. G., and McSTEPHEN, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test–retest intervals. *J. Int. Neuropsychol. Soc.* 9, 419–428. doi: 10.1017/S1355617703930074
- Cummings, J., Lee, G., Ritter, A., Sabbagh, M., and Zhong, K. (2020). Alzheimer's disease drug development pipeline: 2020. *Alzheimers Dement. Transl. Res. Clin. Interv.* 6:e12050. doi: 10.1002/trc2.12050
- Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., et al. (2011). Practice Effects Predict Cognitive Outcome in Amnesic Mild Cognitive Impairment. *Am. J. Geriatr. Psychiatry* 19, 932–939. doi: 10.1097/JGP.0b013e318209dd3a
- Falletti, M. G., Maruff, P., Collie, A., and Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the cogstate battery at 10-minute, one week and one month test-retest intervals. *J. Clin. Exp. Neuropsychol.* 28, 1095–1112. doi: 10.1080/13803390500205718
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198.
- Frost, C., Kenward, M. G., and Fox, N. C. (2008). Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? *Stat. Med.* 27, 3717–3731. doi: 10.1002/sim.3280
- Goldberg, T. E., Harvey, P. D., Wesnes, K. A., Snyder, P. J., and Schneider, L. S. (2015). Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimers Dement.* 1, 103–111. doi: 10.1016/j.dadm.2014.11.003
- Goodglass, H., and Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination Booklet, III, ORAL EXPRESSION*, *J. Animal Naming (Fluency in Controlled Association)*. Philadelphia: Lea & Febiger.
- Gross, A. L., Brandt, J., Bandeen-Roche, K., Carlson, M. C., Stuart, E. A., Marsiske, M., et al. (2014). Do older adults use the method of loci? Results from the

- ACTIVE study. *Exp. Aging Res.* 40, 140–163. doi: 10.1080/0361073X.2014.882204
- Gross, A. L., Inouye, S., Rebok, G., Brandt, J., Crane, P. K., Parisi, J., et al. (2012). Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. *J. Clin. Exp. Neuropsychol.* 34, 758–772. doi: 10.1080/13803395.2012.681628
- Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Heidebrink, J., Barbas, N., et al. (2011). Reliability of repeated cognitive assessment of dementia using a brief computerized battery. *Am. J. Alzheimers Dis. Dementiasr.* 26, 326–333. doi: 10.1177/1533317511411907
- Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., and Morris, J. C. (2015). Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology* 29, 940–948. doi: 10.1037/neu0000208
- Jacobs, D. M., Ard, M. C., Salmon, D. P., Galasko, D. R., Bondi, M. W., and Edland, S. D. (2017). Potential implications of practice effects in Alzheimer's disease prevention trials. *Alzheimers Dement. Transl. Res. Clin. Interv.* 3, 531–535. doi: 10.1016/j.trci.2017.08.010
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). **lmerTest** Package: tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lim, Y. Y., Baker, J. E., Bruns, L., Mills, A., Fowler, C., Fripp, J., et al. (2020). Association of deficits in short-term learning and A β and hippocampal volume in cognitively normal adults. *Neurology* 95, e2577–e2585. doi: 10.1212/WNL.0000000000010728
- Lim, Y. Y., Harrington, K., Ames, D., Ellis, K. A., Lachovitzki, R., Snyder, P. J., et al. (2012). Short term stability of verbal memory impairment in mild cognitive impairment and Alzheimer's disease measured using the International Shopping List Test. *J. Clin. Exp. Neuropsychol.* 34, 853–863. doi: 10.1080/13803395.2012.689815
- Machulda, M. M., Hagen, C. E., Wiste, H. J., Mielke, M. M., Knopman, D. S., Roberts, R. O., et al. (2017). Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *Clin. Neuropsychol.* 31, 99–117. doi: 10.1080/13854046.2016.1241303
- McDade, E., Wang, G., Gordon, B. A., Hassenstab, J., Benzinger, T. L. S., Buckles, V., et al. (2018). Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology* 91, e1295–e1306. doi: 10.1212/WNL.0000000000006277
- Mills, S. M., Mallmann, J., Santacruz, A. M., Fuqua, A., Carril, M., Aisen, P. S., et al. (2013). Preclinical trials in autosomal dominant AD: implementation of the DIAN-TU trial. *Rev. Neurol.* 169, 737–743. doi: 10.1016/j.neurol.2013.07.017
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 43, 2412–2414. doi: 10.1212/wnl.43.11.2412-a
- Nieuwenhuis, R., Te Grotenhuis, M., and Pelzer, B. (2012). influence.ME: tools for detecting influential data in mixed effects models. *R J.* 4, 38–47.
- Oltra-Cucarella, J., Sánchez-SanSegundo, M., and Ferrer-Cascales, R. (2018). 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
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ryman, D. C., Acosta-Baena, N., Aisen, P. S., Bird, T., Danek, A., Fox, N. C., et al. (2014). Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology* 83, 253–260. doi: 10.1212/WNL.0000000000000596
- Salloway, S., Farlow, M., McDade, E., Clifford, D. B., Wang, G., Llibre-Guerra, J. J., et al. (2021). A trial of gantenerumab or solanezumab in dominantly inherited Alzheimer's disease. *Nat. Med.* 27, 1187–1196. doi: 10.1038/s41591-021-01369-8
- Salthouse, T. A., Schroeder, D. H., and Ferrer, E. (2004). Estimating Retest Effects in Longitudinal Assessments of Cognitive Functioning in Adults Between 18 and 60 Years of Age. *Dev. Psychol.* 40, 813–822. doi: 10.1037/0012-1649.40.5.813
- Samaroo, A., Amariglio, R. E., Burnham, S., Sparks, P., Properzi, M., Schultz, A. P., et al. (2020). Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* 12:e12132. doi: 10.1002/dad2.12132
- Sperling, R. A., Rentz, D. M., Johnson, K. A., Karlawish, J., Donohue, M., Salmon, D. P., et al. (2014). The A4 Study: Stopping AD Before Symptoms Begin? *Sci. Transl. Med.* 6, fs13–fs228. doi: 10.1126/scitranslmed.3007941
- Storandt, M., Balota, D. A., Aschenbrenner, A. J., and Morris, J. C. (2014). Clinical and psychological characteristics of the initial cohort of the dominantly inherited Alzheimer Network (DIAN). *Neuropsychology* 28:19. doi: 10.1037/neu0000030
- Valdes, E. G., Sadeq, N. A., Harrison Bush, A. L., Morgan, D., and Andel, R. (2016). Regular cognitive self-monitoring in community-dwelling older adults using an internet-based tool. *J. Clin. Exp. Neuropsychol.* 38, 1026–1037. doi: 10.1080/13803395.2016.1186155
- Wechsler, D. (1981). *Manual: Wechsler Adult Intelligence Scale - Revised*. New York, NY: Psychological Corporation.
- Wechsler, D. (1987). *Manual: Wechsler Memory Scale - Revised*. San Antonio, TX: Psychological Corporation.
- White, J. P., Schembri, A., Edgar, C. J., Lim, Y. Y., Masters, C. L., and Maruff, P. (2021). A paradox in digital memory assessment: increased sensitivity with reduced difficulty. *Front. Digit. Health* 3:780303. doi: 10.3389/fdgh.2021.780303

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: EC, KH, and RY are full-time employees and/or stockholders of Eli Lilly. RB has received funding from Avid Radiopharmaceuticals, Jansse, Hoffman La-Roche/Genentech, Eli Lilly & Co., Eisai, Biogen, AbbVie, and Bristol Meyer Squibbs, has royalties/licenses from C2N Diagnostics, consulting fees from Eisai, Amgen, and Hoffman La-Roche, honoraria from the Korean Dementia Association and the American Neurological Association, and is on the data safety monitoring board or advisory board for Roche/Genentech and Biogen. Unrelated to this article, RB serves as the principal investigator of the DIAN-TU, which is supported by the Alzheimer's Association, GHR Foundation, an anonymous organization, and the DIAN-TU Pharma Consortium (Active: Eli Lilly and Company/Avid Radiopharmaceuticals, F. Hoffman-La Roche/Genentech, Biogen, Eisai, and Janssen. Previous: Abbvie, Amgen, AstraZeneca, Forum, Mithridion, Novartis, Pfizer, Sanofi, and United Neuroscience). In addition, in-kind support has been received from CogState and Signant Health. G-YH has received research support as a clinical trials site investigator from Anavax, Biogen, Eli Lilly, and Roche, has received research grants from the CIHR, Alzheimer Society of Canada, and NIA/NIH, is supported by the Ralph Fisher Professorship in dementia research from the Alzheimer Society of British Columbia, and has participated in expert advisory committees sponsored by Biogen and Roche. GD's research is supported by NIH (K23AG064029), the Alzheimer's Association, and Chan Zuckerberg Initiative, and he serves as a consultant for Parabon Nanolabs Inc., as a Topic Editor (Dementia) for DynaMed (EBSCO), and as the Clinical Director of the Anti-NMDA Receptor Encephalitis Foundation (Inc, Canada; uncompensated) and owns stock in ANI pharmaceuticals. DH, former Department Head of Neurology where the research was conducted, is an inventor on patents for one of the treatments (solanezumab), which has been tested in the DIAN-TU clinical trials. If solanezumab is approved as a treatment for Alzheimer's disease or Dominantly Inherited Alzheimer's Disease, Washington University, and will receive part of the net sales of solanezumab from Eli Lilly, which has licensed the patents related to solanezumab from Washington University.

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Aschenbrenne, Hassensta, Wang, Li, Xiong, McDade, Clifford, Salloway, Farlow, Yaari, Cheng, Holdridge, Mummery, Masters, Hsiung, Surti, Day, Weintraub, Honig, Galvin, Ringman, Brooks, Fox, Snyder, Suzuki, Shimada, Gräber and Bateman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.