Title

Potential implications of practice effects in Alzheimer's disease prevention trials

Journal

Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3(4)

ISSN

2352-8737

Authors

Jacobs, Diane M

Ard, M Colin

Salmon, David P

et al.

Publication Date

2017-11-01

DOI

10.1016/j.trci.2017.08.010

Peer reviewed

Alzheimer's & Dementia: Translational Research & Clinical Interventions 3 (2017) 531-535

Alzheimer's & Dementia

Short Report

Potential implications of practice effects in Alzheimer's disease prevention trials

Diane M. Jacobs[a,b,*], M. Colin Ard[b,c,1], David P. Salmon[a,b], Douglas R. Galasko[a,b], Mark W. Bondi[b,d,e], Steven D. Edland[a,b,c]

[a]*Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA*
[b]*Shiley-Marcos Alzheimer's Disease Research Center, La Jolla, CA, USA*
[c]*Division of Biostatistics, Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA, USA*
[d]*Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA*
[e]*Veteran Affairs San Diego Healthcare System, San Diego, CA, USA*

Abstract

Introduction: Practice effects (PEs) present a potential confound in clinical trials with cognitive outcomes. A single-blind placebo run-in design, with repeated cognitive outcome assessments before randomization to treatment, can minimize effects of practice on trial outcome.

Methods: We investigated the potential implications of PEs in Alzheimer's disease prevention trials using placebo arm data from the Alzheimer's Disease Cooperative Study donepezil/vitamin E trial in mild cognitive impairment. Frequent ADAS-Cog measurements early in the trial allowed us to compare two competing trial designs: a 19-month trial with randomization after initial assessment, versus a 15-month trial with a 4-month single-blind placebo run-in and randomization after the second administration of the ADAS-Cog. Standard power calculations assuming a mixed-model repeated-measure analysis plan were used to calculate sample size requirements for a hypothetical future trial designed to detect a 50% slowing of cognitive decline.

Results: On average, ADAS-Cog 13 scores improved at first follow-up, consistent with a PE and progressively worsened thereafter. The observed change for a 19-month trial (1.18 points) was substantively smaller than that for a 15-month trial with 4-month run-in (1.79 points). To detect a 50% slowing in progression under the standard design (i.e., a 0.59 point slowing), a future trial would require 3.4 times more subjects than would be required to detect the comparable percent slowing (i.e., 0.90 points) with the run-in design.

Discussion: Assuming the improvement at first follow-up observed in this trial represents PEs, the rate of change from the second assessment forward is a more accurate representation of symptom progression in this population and is the appropriate reference point for describing treatment effects characterized as percent slowing of symptom progression; failure to accommodate this leads to an oversized clinical trial. We conclude that PEs are an important potential consideration when planning future trials.

Keywords: Alzheimer's disease; Practice effects; Clinical trial; Single-blind run-in design; Prevention; Preclinical

1. Introduction

Practice effects (PEs) are improvements in cognitive test performance over serial assessments attributed to repeated exposure to test stimuli or procedures. Clinically, PEs can provide valuable information about level of cognitive

[1]Current address: Micron Technology Inc., 8000 S. Federal Way, Boise ID 83716, USA.
*Corresponding author. Tel.: +1-858-246-2543.
E-mail address: djacobs@ucsd.edu

http://dx.doi.org/10.1016/j.trci.2017.08.010

functioning, vis-à-vis ability to benefit from repeated exposure [1,2]; however, in randomized controlled trials, they introduce a source of external signal that may confound observation of the target outcome [3].

Various methods have been proposed to address PEs, including statistical corrections and use of alternate test forms [3–5]. Although alternate forms may minimize memory for specific test items, they do not account for improvements that arise from increased familiarity with test procedures in general [6–8], and equivalent alternate forms are not available for many neurocognitive measures.

Another method to accommodate for PEs in clinical trials is to use a test run-in or "dual baseline" wherein the cognitive outcome measure(s) are administered twice before randomization and scores from the second testing are used as the baseline reference. This approach helps to account for the initial, rapid improvements that occur with repeated testing, which are typically most pronounced between the first and second test administration [5,8]. In a variant of this approach, often referred to as a single-blind placebo run-in design, participants are randomized to treatment or placebo, but all receive placebo during the run-in period between dual baseline assessments and only receive the treatment to which they had been randomized (i.e., active or placebo) after the second assessment. Dual baseline or run-in designs have been used to reduce the influence of practice and placebo effects on clinical trials with neuropsychological outcomes in a variety of diseases and interventions [9–11].

We investigated the impact of a cognitive test run-in design on magnitude of potential effect size and power calculations by examining the performance of participants in the placebo arm of a secondary prevention trial to delay progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD) dementia.

2. Methods

2.1. Overview

We conducted retrospective analyses of placebo arm data from a multicenter, randomized, double-blind, placebo-controlled trial of vitamin E and donepezil HCL to delay clinical progression from MCI to AD dementia; design and results of the trial are described elsewhere [12].

2.2. Participants

Data were obtained from participants in the placebo arm of the donepezil/vitamin E study. All participants were between the age of 55 to 90 years and met diagnostic criteria for amnestic MCI [13]. The placebo group comprised 259 participants with a mean age of 72.9 years (standard deviation [SD] = 7.6), and an average of 14.7 years of education (SD = 3.1); 47% were female, 53% were *APOE* ε4 carriers, and the mean score on the MMSE at screening was 27.35 (SD = 1.8). Data from only the first 18 months of the 36-month trial were used for these analyses because converters to AD dementia were offered open-label donepezil, precluding the ability to look at PEs separate from potential treatment effects in subjects who converted.

2.3. Procedure

The modified 13-item Alzheimer's Disease Assessment Scale, cognitive subscale (ADAS-Cog 13), was administered at the screening visit (1 month before randomization), 3 and 6 months after randomization, and semiannually thereafter. The ADAS-Cog 13 includes all items from the original ADAS-Cog (i.e., word list recall and recognition; measures of language, orientation, constructional and ideational praxis), plus a number cancellation task and a delayed free recall task for a total of 85 points, with higher scores indicating greater cognitive impairment [14]. Three alternate forms of the word-recall word list component were used in the trial: list 1 was administered at screening and 12 months, list 2 at 3 and 18 months, and list 3 at 6 months.

2.4. Data analyses

Sample size calculations informed by placebo arm data from the MCI trial were performed assuming a mixed-model repeated-measures (MMRM) analysis using standard methods we have described [15] and implemented in the R statistical programming language package longpower [16] using a type-I error rate of 5%, power of 80%, and assuming equal allocation to arms. The mean and covariance matrix of repeated ADAS-Cog measures were supplied to the power.mmrm function within the longpower package. To simplify presentation, we assumed no covariate adjustment and no loss to follow-up in power calculations. MMRM, as used in contemporary secondary prevention trials, compares change from randomization to final visit in the treatment arm versus change in control [17]. Mean and SD at each assessment are reported, as is the mean and SD of change from treatment randomization to month-18 visit. We compare the relative sample size required for the two trial designs by example, calculating sample size required to detect a 50% slowing of decline. Under our assumptions, the relative sample size required when effect size is expressed as percent slowing of decline is solely a function of the mean and covariance structure of the pilot data for this analysis plan [18]. Hence, relative sample size for our reported findings for the 50% slowing of decline generalize to any effect size expressed as percent slowing of decline.

3. Results

Participant mean scores on the ADAS-Cog 13 are shown in Fig. 1. At screening, the group mean score was 17.40 (SD = 6.0). At 3-month follow-up, the group mean score improved slightly to 16.79 (SD = 7.0). At 6-month follow-up, the group mean returned to the baseline level (mean = 17.38; SD = 7.0), and performance progressively declined thereafter. Between screening and 18-month

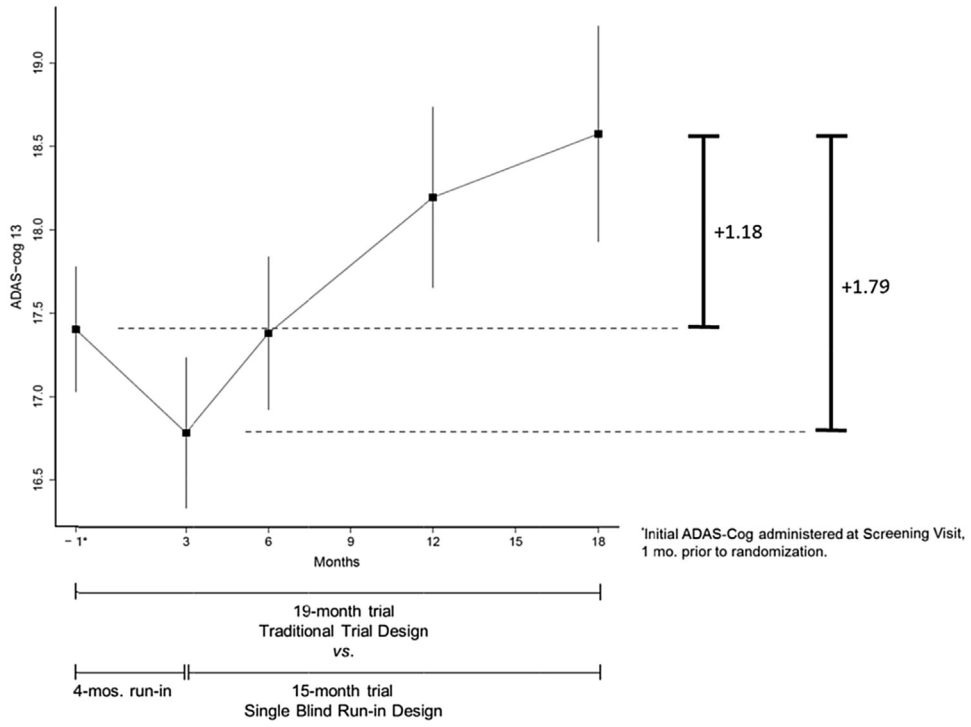

Fig. 1. ADAS-Cog 13 scores of participants with MCI in the placebo arm of the ADCS donepezil/vitamin E study. Abbreviations: ADAS-Cog 13, modified 13-item Alzheimer's Disease Assessment Scale; ADCS, Alzheimer's Disease Cooperative Study; MCI, mild cognitive impairment.

follow-up, the mean change was +1.18 (SD = 6.2); the change between the 3- and 18-month visits was +1.79 (SD = 5.2).

Using data from the screening visit (1 month before baseline) to 18-month follow-up in power calculations, a standard 19-month trial design is estimated to require 1764 subjects per arm to detect a 50% slowing of decline. Using data from 3- to 18-month follow-up in power calculations, a 15-month study with 4-month placebo run-in would require 521 subjects per arm to detect a 50% slowing of decline, 70% less subjects than required by the standard design. Stated another way, the standard design trial would require 3.4 times more subjects for comparable power. This is a general finding when treatment effect size is expressed as percentage slowing of mean rate of decline (see Section 2); that is, regardless of the percentage effect size powered for, when effect size is expressed as percentage slowing of decline, the standard design trial will require 3.4 times more subjects than the run-in design trial.

4. Discussion

We used the placebo arm data from a completed trial to demonstrate the impact of PEs on magnitude of potential effect size and sample size projections for two study designs. Regarding treatment effect size, clearly in the presence of PEs, the rate of change after washout of PEs (from the 3-month visit forward in our example) is the most correct characterization of rate of disease progression on a given instrument [5]. It follows that treatment effect sizes characterized relative to this rate are more accurate and meaningful. Moreover, characterizing treatment effect size relative to change from first assessment results in oversized trials in the presence of PEs.

Results demonstrate that two investigators, informed by the pattern of progression observed in a previous trial and with similar directives in terms of treatment effect size, can come to dramatically differently sized trials depending on the design. This is explained by the difference in effect size when effect size is expressed in units of ADAS-Cog 13. For example, for the standard design with 19-month treatment, a 50% reduction in change corresponds to half of 1.18 (Fig. 1) or 0.59 units. For the run-in design, a 50% reduction in change corresponds to half of 1.79 (Fig. 1) or 0.90 units. In units of ADAS-Cog 13, the effect size powered for in the run-in design is much larger and requires a smaller projected sample size.

There are limitations to this analysis. Alternate versions of the ADAS-Cog word list memory test were used in the trial, so it is possible that improved performance observed at the second administration of the ADAS-Cog reflects version differences rather than a PE. The possibility that the word list used at the 3-month visit was easier for participants seems less likely, however, given that a similar improvement relative to previous measurements was not observed at 18-month follow-up, when the same word list was again used. Furthermore, the run-in design, as proposed, does not consider the effects of any additional PEs that may occur beyond the second test administration. Finally, instruments that are less vulnerable to PEs will be less prone to the issues outlined in this report. Replication of these findings in other cohort and trial data, with other commonly used outcome measures, will

further confirm the potential impact of the single-blind placebo cognitive test run-in design on effect size characterization.

Consideration of PEs will be increasingly important as the target population for AD trials moves earlier in the disease course, where measurable treatment effects may be subtle and PEs more robust [5,19,20]. Trials of treatments intended to slow the underlying AD neurodegenerative process, where no acute treatment effects are anticipated, may be particularly vulnerable in this regard. We conclude that the presence of PEs is an important consideration when planning future trials.

Acknowledgments

Data used in the preparation of this article were obtained from the Alzheimer's Disease Cooperative Study (ADCS) legacy database (NIA U19 5U19AG010483). The ADCS donepezil/vitamin E trial in MCI (ClinicalTrials.gov identifier NCT00000173) was supported by NIA UO1 AG10483 and by Pfizer and Eisai. The authors and the research herein were additionally supported by NIA P50 AG005131; NIA R03 AG047580; NIA R01 AG049810; NIA K24 AG026431; NIH UL1 TR001442; and the Shiley-Marcos Alzheimer's Disease Research Center at the University of California San Diego Medical Center.

RESEARCH IN CONTEXT

1. Systematic review: Practice effects on neuropsychological measures have been well documented and may mask treatment effects in clinical trials with cognitive end points. Single-blind run-in designs have been used to attenuate practice effects before randomization for other diseases but have not been applied consistently in Alzheimer's disease trials.
2. Interpretation: Archival analyses of placebo arm data from a secondary Alzheimer's disease (AD) prevention trial with mild cognitive impairment participants revealed that using a single-blind placebo cognitive test run-in design yielded greater change in cognitive outcome (ADAS-Cog 13) than a traditional design with randomization at the first assessment visit. The run-in design dramatically reduced the requisite sample size to achieve comparable statistical power.
3. Future directions: Replication of these findings using other outcome measures as well as in participants in the preclinical and asymptomatic stages of AD will further validate the utility of using a single-blind placebo cognitive test run-in for AD prevention trials.

References

[1] Duff K, Lyketsos CG, Beglinger LJ, Chelune G, Moser DJ, Arndt S, et al. Practice effects predict cognitive outcome in amnestic mild cognitive impairment. Am J Geriatr Psychiatry 2011; 19:932–9.
[2] Hassenstab J, Ruvolo D, Jasielec M, Xiong C, Grant E, Morris JC. Absence of practice effects in preclinical Alzheimer's disease. Neuropsychology 2015;29:940–8.
[3] 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: Diagnosis, Assessment & Disease Monitoring 2015; 1:103–11.
[4] 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.
[5] Vivot A, Power MC, Glymour MM, Mayeda ER, Benitez A, Spiro A III, 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.
[6] Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol 2005; 20:517–29.
[7] Zgaljardic DJ, Benedict RH. Evaluation of practice effects in language and spatial processing test performance. Appl Neuropsychol 2001; 8:218–23.
[8] 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.
[9] McCaffrey RJ, Ortega A, Orsillo SM, Haase RF, McCoy GC. Neuropsychological and physical side effects of metoprolol in essential hypertensives. Neuropsychology 1992;6:225–38.
[10] Schmitt FA, Bigley JW, McKinnis R, Logue PE, Evans RW, Drucker JL, et al. Neuropsychological outcome of Zidovudine (AZT) treatment of patients with AIDS and AIDS-related complex. N Engl J Med 1988;319:1573–8.
[11] Beglinger LJ, Adams WH, Langbehn D, Fiedorowicz JG, Jorge R, Biglan K, et al. Results of the citalopram to enhance cognition in Huntington disease trial. Mov Disord 2014;29:401–5.
[12] Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 2005;352:2379–88.
[13] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8 [Erratum, Arch Neurol 1999; 56:760.].
[14] Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord 1997; 11:S13–21.
[15] Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. J Alzheimers Dis 2011;26:369–77.
[16] Donohue MC, Gamst AC, Edland SD. Sample Size Calculations for Longitudinal Data. Longpower: Power and Sample Size Calculators for Longitudinal Data R Package Version 1.0–16. Available at: Https://cran.r-Project.org/web/packages/longpower/longpower.pdf; 2016. Accessed September 19, 2017.
[17] Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014; 71:961–70.

[18] Edland SD, Ard MC, Sridhar J, Cobia D, Martersteck A, Mesulam MM, et al. Proof of concept demonstration of optimal composite MRI endpoints for clinical trials. Alzheimers Dement: Translational Research & Clinical Interventions 2016;2:177–81.

[19] Machulda MM, Hagen CE, Wiste HJ, Mielke MM, Knopman DS, Roberts RO, et al. Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. Clin Neuropsychol 2017;31:99–117.

[20] Machulda MM, Pankratz VS, Christianson TJ, Ivnik RJ, Mielke MM, Roberts RO, et al. Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. Clin Neuropsychol 2013;27:1247–64.