Featured Article

Sex-specific composite scales for longitudinal studies of incipient Alzheimer's disease

Sarah J. Banks[a,*], Benjamin Shifflett[b], Jody-Lynn Berg[a], Erin Sundermann[c], Guerry Peavy[a], Mark W. Bondi[c,d], Steven D. Edland[b], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA
[b]Department of Family Medicine and Public Health, UCSD, La Jolla, CA
[c]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA
[d]Psychology Service, VA San Diego Healthcare System, San Diego, CA, USA

Abstract

Introduction: The impact of Alzheimer's disease (AD) on cognitive decline differs by sex. Composite scores are useful as singular outcomes in clinical trials, yet to date these have not been developed to measure sex-specific change.

Method: We derived optimal composites from component scales available in the AD Neuroimaging Initiative (ADNI) database among cognitively normal and mild cognitively impaired subjects who are cerebrospinal fluid amyloid-β positive for early AD. Maximally sensitive composites were constructed separately for men and women using standard formulas. We compared the statistical power of the composites with the ADNI Prodromal Alzheimer's Cognitive Composite.

Results: Among 9 cognitive measures and clinical dementia rating sum of boxes, the optimal sex-specific composites included 5 measures, including the clinical dementia rating and 4 distinct cognitive measures. The sex-specific composites consistently outperformed sex-agnostic composites and the ADNI Prodromal Alzheimer's Cognitive Composite.

Discussion: Sex-specific composite scales may improve the power of longitudinal studies of early AD and clinical trials.

© 2019 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Sex differences; Cognition; Clinical trial outcomes; Alzheimer's disease; Composites

1. Introduction

Sex is an important influence on cognition throughout the life span: Women are generally found to have stronger verbal abilities, particularly in memory, and men have stronger visuospatial abilities [1]. There is concern that this verbal memory reserve may prevent the timely diagnosis of Alzheimer's disease (AD) in women [2] and complicate the interpretation of decline in studies where data from men and women are expected to show similar rates of decline.

*Corresponding author. Tel.: (858) 246-1264; Fax: (858) 246-1263.
E-mail address: sbanks@ucsd.edu

In the context of clinical trials, the use of composite scores has become increasingly popular. These composite scores combine the results of numerous tests usually across different domains. This allows for a singular outcome score that takes into account performance in several aspects of cognition [3,4]. Composites can be theoretically or empirically derived, or a combination of both the approaches. However, averaging across several domains can dilute some effects and reduce ability to capture treatment effects that can be domain specific [4,5]. Thus, it is important to only include tests where meaningful change with progression of the disease is likely. Composite score construction ranges from simply averaging across all scores in a battery [6,7] to more complex weighting of the components. Empirically derived weightings of components based on longitudinal data enhance signal-to-noise ratio, creating a statistically powerful outcome that can better track disease progression [8]. Despite sex differences in how men and women respond cognitively to AD and in performance on composite scales, for example, in the Prodromal Alzheimer's Cognitive Composite (PACC) [9], sex-specific composite scores have not yet been developed. Rather, sex is usually treated as a covariate instead of a focus of analyses in most studies of AD. In the present study, we examine on sex differences in a prodromal AD sample from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and derive sex-specific composites from these data.

The ADNI battery is loaded to verbal tests, especially of memory. This is clinically and experimentally important because many of these tests have been shown to have robust sex differences. The ADNI cognitive battery tests, and reported differences in performance on these tests by sex, are summarized in Table 1. Performance varies by sex for several of the instruments in this battery, and, in some cases, these differences are altered dynamically with the progression of AD. To investigate the potential implications of this for clinical trials of the earliest stages of AD, we derived sex-specific optimally weighted composite scales informed by the ADNI sample and compared their performance to the PACC, a previously established composite score currently being used as the primary endpoint for phase 3 clinical trials in early-stage AD [6].

2. Methods

2.1. Participants

We used data from the ADNI. Informed consent was conducted locally at each site under the regulation of the local internal review boards. To represent the population of people recruited to preclinical and prodromal AD trials, we restricted our participant sample to cognitively normal and mild cognitively impaired subjects from the ADNI cohort who were biomarker positive for amyloid β (Aβ) plaque deposition as indicated by cerebrospinal fluid biomarker positivity using the newly released Elecsys cerebrospinal fluid Aβ assay data and previously established cutpoints [26].

2.2. Measures

We considered composites constructed from all possible combinations of 9 longitudinal measures, the Clinical Dementia Rating scale sum of boxes (CDR SOB) and eight cognitive measures (see Table 2). We used a calculated variable, Trail Making Test part B minus part A (Trails B − A), to extract the executive component from processing speed.

2.3. Statistical Methods

Descriptive data for men and women were compared using t-tests or χ^2 tests as appropriate. Optimal sex-specific composites were constructed as previously described [4,8], assuming a mixed-model repeated-measure analysis plan. Mixed-model repeated measure is preferred for FDA-registered phase 3 trials because the primary estimand is the mean difference between arms in change from first visit to last visit at the end of the trial without being unduly influenced by intermediate patterns of progression of disease. The optimal composite is the weighted sum of the component measures that maximizes the ratio of mean change to the standard deviation of change and, therefore, is the most statistically powered endpoint for detecting differences in change between treatment groups in a clinical trial [4]. Weights for the optimal composite are a function of the covariance of change scores (last observation minus first observation) of the component measures and the vector of mean change scores of the components. These values were estimated from the ADNI data, and the resulting weight applied to the component scores to calculate composite values [7].

To explore the relative utility of sex-specific composite measures, we compared performance of our identified composites with performance of the ADNI-PACC [27]. The ADNI-PACC is an approximation of the PACC, which is the primary endpoint for ongoing phase 3 treatment trials of prodromal AD [9,28]. The ADNI-PACC substitutes the Alzheimer's disease assessment scale - cognitive subscale Delayed Word Recall score for the Free and Cued Selective Reminding Test and switches the Trail Making Part B (log-transformed) for Digit Symbol, but otherwise contains the other two standard components of the PACC: the Mini–Mental Status Examination and the Logical Memory Delayed Recall test [6]. The components of the ADNI-PACC were standardized to the pooled male and female baseline data consistent with how this instrument is implemented in practice [21]. Standard power formulas were used to calculate statistical power as a function of sample size for the composite scales as well as the component measures assuming a two-arm trial with equal allocation to arms and type I error rate of 5% designed to detect a 25% slowing of progression of disease [4]. To

Table 1
Cognitive tests in the ADNI battery: Brief descriptions of each task and synopsis of sex differences reported to date in performance

Test Name	Description	Sex-specific findings in older adults
The Rey Auditory Verbal Learning Test [10]	15 Word List Learning task (5 trials) and 30-minute Delay Recall	Females perform better on this task [11]; but advantage is lost with cognitive decline in Aβ-positive women [12] and when the disease becomes more severe [2,13]
Alzheimer's disease assessment scale - cognitive subscale Word List [14]	10 Word List Learning task (3 trials) and 5-minute Delay Recall	Females perform better on this task [11]
Wechsler Memory Scale - Revised Logical Memory Story A [15]	Immediate recall of a story containing 25 items and 20-30 minute Delay Recall	Females have been shown to perform better on this task and to have enhanced blood flow to the left temporal lobe compared with men [16]
Digit symbol substitution	Processing speed: Key at the top of the page comprises 9 digit-symbol pairs; task is to complete several lines where the digit is presented and for each digit the subject should write down the corresponding symbol as fast as possible	Young women have been shown to have a slight advantage on this task [17], but there is less information for older adults
Boston Naming Test second edition [18]	Confrontation Naming Task consisting of 30 black and white drawings	Sex differences are not apparent on this task [19]
Category Fluency (Animals) [20]	Executive Functioning Task: naming as many different animals as fast as possible in one minute	Female advantage is commonly reported for this task and appears stable in old age [21]
Trail Making Test [22]	Part A (psychomotor processing): draw lines to connect numbers in ascending order Part B (executive): draw lines alternating between numbers and letters in ascending order	Large studies of normative data do not demonstrate sex differences [23]
Clock Drawing Test [24]	Visuospatial task: Draw the face of a clock, including the numbers, and place the hands at 10 after 11	Males have been shown to perform better on this task [25]

simplify presentation, we further assumed no dropout in power calculations.

3. Results

Twenty-four–month follow-up data were available for 86 men and 56 women. See Table 2 for details of demographic characteristics. Men had significantly more years of education than women. Neuropsychological test data were mostly similar between groups, although significant differences were noted with men doing better on the Boston Naming Test and women doing better on the learning trials of the Rey Auditory Verbal Learning Test. With optimally weighted composites, the mean-to-standard-deviation ratio (MSDR) and statistical power increase as the number of components included in the composite increase. We found that improvement in MSDR was negligible after a total of 5 components were included in the composites. That is, five-component composites were both parsimonious and near maximal in terms of MSDR and statistical power. For women, CDR SOB, Logical Memory Immediate Recall, Boston Naming Test, Trails B − A, and Category Fluency (animals) were included. In men, CDR SOB, clock drawing, learning trials of Rey Auditory Verbal Learning Test and delayed recall, and Logical Memory Delayed Recall were included (Table 3).

Power curves estimating statistical power to detect a 25% slowing of decline as a function of sample size [29] are illustrated in Fig. 1 (equal allocation to arms, mixed-model repeated-measure analysis, type I error rate α = 0.05). For example, given parameters estimated from ADNI pilot data, a single sex trial with approximately 250 subjects per arm (women) or 270 subjects per arm (men) would have 80 percent power to detect a treatment effect equivalent to a 25% slowing of rate of decline using the sex-specific composites as a primary outcome (Fig. 1). In comparison, assuming the parameters estimated from ADNI pilot data, a comparably powered trial using the ADNI-PACC would require 979 subjects per arm. Several factors explain this discrepancy in power. We note that one component of the ADNI-PACC, Logical Memory Delayed Recall, on average, did not decline in two years in the ADNI sample (Table 3), meaning this component contributed noise to the ADNI-PACC but no meaningful signal relevant to measuring treatment effects in a clinical trial. In total, the ADNI-PACC performed poorer than its single most sensitive component measure, which was the Mini–Mental Status Examination (Fig. 1, Table 3). Composite samples performing poorer than their component instruments have been demonstrated previously using computer simulations [4], highlighting the importance of carefully considering the weighting of components when constructing a composite scale. For comparison, we calculated optimal weights for components of the PACC as informed by available pilot data. The composite composed of the components of the PACC using optimal weighting had substantially greater statistical power as an endpoint for a clinical trial compared with the standard PACC (Fig. 1).

Table 2
Demographic, biomarker, and clinical characteristics of the sample

Variable	Men (n = 86), mean (SD)	Women (n = 56), mean (SD)	P value
Age	75.28 (6.39)	73.47 (7.06)	.12475
Education (years)	16.47 (2.72)	15.08 (3.12)	.00977
ADNI diagnosis normal: n (%)	21 (24.42%)	20 (35.71%)	.20690
ADNI diagnosis mild cognitively impairment: n (%)	65 (75.58%)	36 (64.29%)	
E4 negative: n (%)	32 (37.21%)	26 (46.43%)	.35883
E4 positive: n (%)	54 (62.79%)	30 (53.57%)	
CSF amyloid β level	598.57 (171.15)	637.70 (176.54)	.19393
CSF phospho-tau level	29.62 (12.63)	32.36 (14.38)	.25220
CSF total tau level	293.55 (113.89)	321.41 (124.69)	.18636
CSF phospho-tau/amyloid β ratio	0.05 (0.03)	0.05 (0.03)	.81476
CSF total tau/amyloid β ratio	0.53 (0.24)	0.54 (0.28)	.81896
Clock Draw: Copy	**4.71 (0.61)**	**4.64 (0.82)**	**.60406**
Clock Draw: Command	**4.13 (1.13)**	**4.39 (0.97)**	**.13725**
Rey AVLT: Trial I-V Sum	**32.49 (9.87)**	**36.04 (10.55)**	**.04703**
Rey AVLT: 30-minute Recall	**3.5 (3.76)**	**3.86 (4.17)**	**.60528**
Logical Memory: Immediate Recall	**8.58 (3.91)**	**9.23 (5.25)**	**.42880**
Logical Memory: Delayed Recall	**5.77 (4.43)**	**6.80 (6.32)**	**.28832**
Boston Naming Test	**27.01 (2.58)**	**25.59 (4.91)**	**.04974**
Trails B − Trails A	**79.78 (59.42)**	**72.05 (53.28)**	**.42146**
Category Fluency: Animals	**17.34 (5.34)**	**16.70 (4.80)**	**.45903**
CDR Sum of Boxes	**1.17 (1.05)**	**1.07 (1.06)**	**.57209**
Mini–Mental Status Examination	27.44 (1.86)	27.62 (1.98)	.58184
ADNI-PACC standard	0.19 (2.55)	0.89 (3.48)	.19814

NOTE. Measures in bold were used in developing the composites.

Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; CDR, Clinical Dementia Rating; CSF, cerebrospinal fluid; PACC, Prodromal Alzheimer's Cognitive Composite; SD, standard deviation.

4. Discussion

The optimal sex-specific composites outperformed the ADNI-PACC in terms of statistical power to detect percentage difference in rate of decline between treatment arms in clinical trials. Similar improvements in power would be expected for detecting predictors of decline in cohort studies of comparable populations. The improved performance follows both from using sex-specific, optimal weighting when constructing the composite, and from the fact that the optimal composite includes a cognitive-functional measure, the CDR SOB, which is sensitive to disease progression in the early stages of AD. The FDA both has approved and supports composite cognitive functional scales as primary endpoints to clinical trials for this very reason [30].

These data suggest that sex-specific composites may provide a more sensitive predictor of decline in the population represented by this sample compared with currently

Table 3
Baseline to two-year mean change of components and composite scales

Men				Women				ADNI-PACC (all subjects)			
Item	Mean	SD	MSDR	Item	Mean	SD	MSDR	Item	Mean	SD	MSDR
Men Optimal	0.807	0.832	0.970	Women Optimal	1.050	1.051	0.999	ADNI-PACC Optimal	0.404	0.626	0.646
Clock Draw: Command	0.093	1.164	0.080	Category Fluency: Animals	1.071	5.023	0.213	ADNI-PACC Standard	1.361	2.684	0.507
RAVLT: Trial I-V Sum	1.547	6.799	0.227	Boston Naming Test	0.429	3.138	0.137	Trails B Logarithm	0.049	0.203	0.241
RAVLT: 30 Minute Recall	0.174	2.680	0.065	Trails B − Trails A	33.714	60.417	0.558	Logical Memory: Delayed	0.000	3.435	0.000
Logical Memory: Delayed Recall	0.267	3.654	0.073	Logical Memory: Immediate Recall	0.125	2.848	0.044	Alzheimer's disease assessment scale: Delayed Word Recall	0.725	1.830	0.396
CDR Sum of Boxes	1.280	1.533	0.835	CDR Sum of Boxes	1.065	1.783	0.597	Mini–Mental Status Examination	1.690	3.205	0.527

Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; CDR, Clinical Dementia Rating; PACC, Prodromal Alzheimer's Cognitive Composite; MSDR, mean-to-standard-deviation ratio; RAVLT, Rey Auditory Verbal Learning Test; SD, standard deviation.
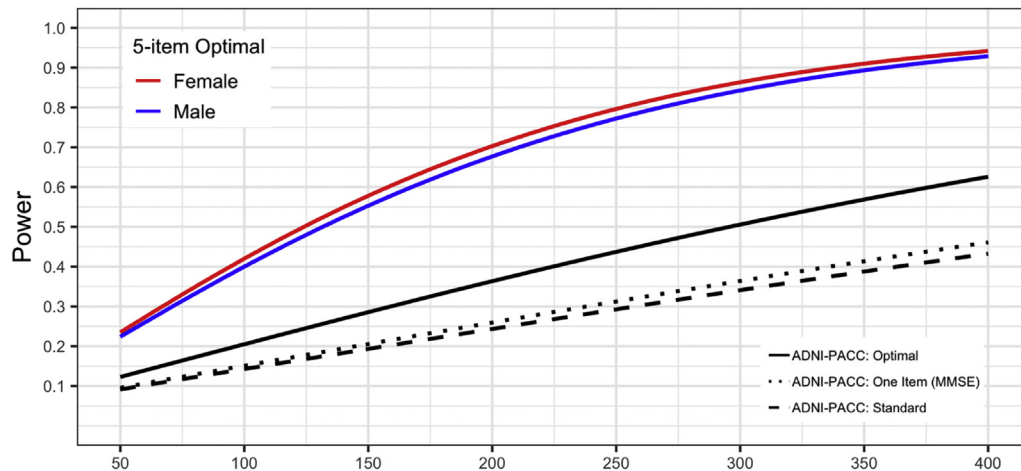
Fig. 1. Power to detect 25% reduction in rate of decline (2-year clinical trial, equal allocation to arms, type I error rate 5%). Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; PACC, Prodromal Alzheimer's Cognitive Composite; MMSE, Mini–Mental Status Examination.

available composites such as the PACC. Among men, instruments dominated by the single visuospatial task (clock drawing) and memory tasks were most sensitive collectively to cognitive change, whereas among women, instruments sensitive to language and executive functioning were predominant. These findings suggest that studies targeting these earliest stages of disease would benefit from targeted sex-specific assessment instruments. However, we caution that the available sample size is not sufficient to definitively answer this question. We have previously shown that sample sizes 100 to 300 are required to meaningfully determine optimal weighting of component measures when constructing a composite [8]. Future investigation with larger samples than available in this pilot study will be required to better determine the relative efficiency of sex-specific composites as outcome measures for clinical trials.

We also investigated the performance of the standard ADNI-PACC vis-à-vis the ADNI-PACC constructed using the weighted sum of the ADNI-PACC components that maximized signal to noise of the resulting composite (Fig. 1). The resulting "optimal" ADNI-PACC would require approximately 40% less subjects than the standard ADNI-PACC to obtain comparable power to detect a given percent slowing of rate of decline in a clinical trial (Fig. 1). Notable, one component of the ADNI-PACC, the Mini–Mental Status Examination, had a better signal-to-noise ratio in the ADNI sample and therefore based on these data would be a more powerful endpoint for a clinical trial. These observations underscore the influence and importance of the weighting of components when constructing a composite outcome measure. It should also be noted that the original PACC features the Free and Cued Selective Reminding Test, which is a sensitive measure to early dementia [31].

What are the potential implications for clinical trials of the finding that sex-specific composites perform well as clinical trial endpoints? Although this study was performed on a relatively small sample and requires replication in a larger sample, including more diverse participants, there are several implications that should be considered. Given recent findings of sex differences in AD tau pathophysiology [32,33] and rate of cognitive decline in AD [34] and evidence that the influence of genetic variants may differ by sex [35–37], there may be a time when single-sex clinical trials are considered. Sex-specific composites would be the obvious choice for primary endpoints for such trials. For clinical trials recruiting both men and women, sex-specific instruments constructed from the same components but weighted differently for men and women would be optimal and could possibly allow for pooled analyses if the instrument had similar longitudinal properties in men and women. Alternatively, sex-stratified meta-analyses with overall Fisher meta-analysis *P* values [38] that would preserve power *and* realize the advantages of sex-specific outcome assessment could be considered. Optimal endpoints for clinical trials decrease required sample size (minimize study subject burden and study cost) and increased statistical power (increasing the likelihood that effective treatments are identified) and therefore have vast clinical implications.

RESEARCH IN CONTEXT

1. Systematic review: Composites are important outcome measures in clinical trials. They are usually developed in a sex-agnostic fashion, despite known and important sex differences in cognition, both throughout the life span and in response to Alzheimer's disease. We used empirical methods to develop sex-specific composites using data from Alzheimer's Disease Neuroimaging Initiative.

2. Interpretation: Our men- and women-specific composites both included the Clinical Dementia Rating Sum of Boxes, but otherwise included distinct test selections. In comparison with other composites or single-outcome measures, use of these composites would allow for reduced numbers of participants in clinical trials using them as a primary outcome.

3. Future directions: These results will need to be replicated on other cohorts. If similar findings result, implementation of sex-specific composites might aid in improving clinical trial design.

References

[1] Pauls F, Petermann F, Lepach AC. Gender differences in episodic memory and visual working memory including the effects of age. Memory 2013;21:857–74.

[2] Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM, et al. Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men? J Alzheimers Dis 2017;56:947–57.

[3] 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. Alzheimer's Dement Transl Res Clin Interv 2016;2:177–81.

[4] Ard MC, Raghavan N, Edland SD. Optimal composite scores for longitudinal clinical trials under the linear mixed effects model. Pharm Stat 2015;14:418–26.

[5] Weintraub S, Carrillo MC, Farias ST, Goldberg TE, Hendrix JA, Jaeger J, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. Alzheimer's Dement Transl Res Clin Interv 2018;4:64–75.

[6] 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.

[7] Buckley RF, Mormino EC, Amariglio RE, Properzi MJ, Rabin JS, Lim YY, et al. Sex, amyloid, and APOE ε4 and risk of cognitive decline in preclinical Alzheimer's disease: findings from three well-characterized cohorts. Alzheimer's Dement 2018;14:1193–203.

[8] Edland SD, Ard MC, Li W, Jiang L. Design of pilot studies to inform the construction of composite outcome measures. Alzheimers Dement Transl Res Clin Interv 2017;3:213–8.

[9] Papp KV, Rentz DM, Orlovsky I, Sperling RA, Mormino EC. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5. Alzheimers Dement Transl Res Clin Interv 2017;3:668–77.

[10] Rey A. L'examen clinique en psychologie [The clinical examination in psychology]. Oxford, England: Presses Universitaries De France; 1958.

[11] Bleecker ML, Bolla-Wilson K, Agnew J, Meyers DA. Age-related sex differences in verbal memory. J Clin Psychol 1988;44:403–11.

[12] Caldwell JZK, Berg J-L, Cummings JL, Banks SJ. Moderating effects of sex on the impact of diagnosis and amyloid positivity on verbal memory and hippocampal volume. Alzheimers Res Ther 2017;9:72. https://doi.org/10.1186/s13195-017-0300-8.

[13] Sundermann EE, Maki PM, Rubin LH, Lipton RB, Landau S, Biegon A. Female advantage in verbal memory: evidence of sex-specific cognitive reserve. Neurology 2016;87:1916–24.

[14] Mohs RC. Administration and Scoring Manual for the Alzheimer's Disease Assessment Scale 1994 Revised Edition. The Mount Sinai School of Medicine; 1994.

[15] Wechsler D. Wechsler Memory Scale-Revised. San Antonio, Texas: Psychological Corporation; 1987.

[16] Ragland JD, Coleman AR, Gur RC, Glahn DC, Gur RE. Sex differences in brain-behavior relationships between verbal episodic memory and resting regional cerebral blood flow. Neuropsychologia 2000; 38:451–61.

[17] Roivainen E. Gender differences in processing speed: a review of recent research. Learn Individ Differ 2011;21:145–9.

[18] Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. 2nd. Philadelphia: Lea and Febiger; 1983.

[19] Saxton J, Ratcliff G, Munro CA, Coffey EC, Becker JT, Fried L, et al. Normative data on the Boston naming test and two equivalent 30-item short forms. Clin Neuropsychol 2000;14:526–34.

[20] Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. Neurology 1989;39:1159–65.

[21] De Frias C, Nilsson LG, Herlitz A. Sex differences in cognition are stable over a 10-year period in adulthood and old age. Aging Neuropsychol Cogn 2006;13:574–87.

[22] Partington J, Leiter R. Partington's pathways test. Psychol Serv Cent Bull 1949;1:1–11.

[23] Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004;19:203–14.

[24] Rouleau I, Salmon DP, Butters N, Kennedy C, McGuire K. Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. Brain Cogn 1992;18:70–87.

[25] Seigerschmidt E, Mösch E, Siemen M, Förstl H, Bickel H. The clock drawing test and questionable dementia: Reliability and validity. Int J Geriatr Psychiatry 2002;17:1048–54.

[26] Shaw LM, Waligorska T, Fields L, Korecka M, Figurski M, Trojanowski JQ, et al. Derivation of cutoffs for the Elecsys® amyloid β (1–42) assay in Alzheimer's disease. Alzheimers Dement 2018; 10:698–705.

[27] Donohue MC, Sperling RA, Petersen R, Sun CK, Weiner M, Aisen PS. Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. JAMA 2017;317:2305–16.

[28] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228fs13. https://doi.org/10.1126/scitranslmed.3007941.

[29] Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. Adv Alzheimer's Dis 2011;26:369–77.

[30] Sabbagh MN, Hendrix SB, Harrison JE. FDA position statement "Early Alzheimer's disease: Developing drugs for treatment, Guidance for Industry". Alzheimer's Dement Transl Res Clin Interv 2019;5:13–9.

[31] Grober E, Sanders AE, Hall C, Lipton RB. Free and cued selective reminding identifies very mild dementia in primary care. Alzheimer Dis Assoc Disord 2010;24:284–90.

[32] Oveisgharan S, Arvanitakis Z, Yu L, Farfel J, Schneider JA, Bennett DA. Sex differences in Alzheimer's disease and common neuropathologies of aging. Acta Neuropathol 2018;136:887–900.

[33] Liesinger AM, Graff-Radford NR, Duara R, Carter RE, Hanna Al-Shaikh FS, Koga S, et al. Sex and age interact to determine clinicopathologic differences in Alzheimer's disease. Acta Neuropathol 2018; 136:873–85.

[34] Lin KA, Choudhury KR, Rathakrishnan BG, Marks DM, Petrella JR, Doraiswamy PM. Marked gender differences in progression of mild cognitive impairment over 8 years. Alzheimer's Dement Transl Res Clin Interv 2015;1:103–10.

[35] Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. J Steroid Biochem Mol Biol 2016;160:134–47.

[36] Sundermann EE, Tran M, Maki PM, Bondi MW. Sex differences in the association between apolipoprotein E ε4 allele and Alzheimer's disease markers. Alzheimer's Dement 2018;10:438–47.

[37] Fan CC, Banks SJ, Thompson WK, Chen CH, McEvoy LK, Tan CH, et al. Sex-dependent polygenic effects on the clinical progressions of Alzheimer's disease. BioRxiv 2019; https://doi.org/10.1101/613893.

[38] Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J Evol Biol 2011; 24:1836–41.