Received: 13 April 2020 | Revised: 28 August 2020 | Accepted: 29 September 2020

DOI: 10.1002/alz.12214

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

FEATURED ARTICLE

Development and validation of the Uniform Data Set (v3.0) executive function composite score (UDS3-EF)

Adam M. Staffaroni[1] | Breton M. Asken[1] | Kaitlin B. Casaletto[1] | Corrina Fonseca[1] | Michelle You[1] | Howard J. Rosen[1] | Adam L. Boxer[1] | Fanny M. Elahi[1] | John Kornak[2] | Dan Mungas[3] | Joel H. Kramer[1]

[1] Department of Neurology, Memory and Aging Center, Weill Institute for Neurosciences, University of California at San Francisco (UCSF), San Francisco, California, USA

[2] Department of Epidemiology and Biostatistics, Memory and Aging Center, University of California at San Francisco (UCSF), San Francisco, California, USA

[3] Department of Neurology, University of California, Davis, Davis, California, USA

Correspondence
Adam M. Staffaroni, University of California, San Francisco, Weill Institute for Neurosciences, Department of Neurology, Memory and Aging Center, 675 Nelson Rising Lane, Suite 190, San Francisco, CA 94158, USA.
Email: adam.staffaroni@ucsf.edu

Adam M. Staffaroni and Breton M. Asken contributed equally to this study.

Funding information
National Institutes of Health; Larry L. Hillblom Foundation, Grant/Award Numbers: 2014-A-004-NET, 2018-A-006-NET, 2017-A-004-FEL, 2018-A-025-FEL; NIA, Grant/Award Numbers: K23AG061253, AG045390, AG032306, AG021886, AG016976, L30AG057123, K23AG058752; NIH/NINDS, Grant/Award Numbers: NS092089, UH3NS100608; National Institute of Neurological Disorders and Stroke; NIH/NIBIB, Grant/Award Number: R01EB022055; National Institute of Biomedical Imaging and Bioengineering

Abstract

Introduction: Cognitive composite scores offer a means of precisely measuring executive functioning (EF).

Methods: We developed the Uniform Data Set v3.0 EF composite score (UDS3-EF) in 3507 controls from the National Alzheimer's Coordinating Center dataset using item-response theory and applied nonlinear and linear demographic adjustments. The UDS3-EF was validated with other neuropsychological tests and brain magnetic resonance imaging from independent research cohorts using linear models.

Results: Final model fit was good-to-excellent: comparative fit index = 0.99; root mean squared error of approximation = 0.057. UDS3-EF scores differed across validation cohorts (controls > mild cognitive impairment > Alzheimer's disease-dementia ≈ behavioral variant frontotemporal dementia; $P < 0.001$). The UDS3-EF correlated most strongly with other EF tests (βs = 0.50 to 0.85, Ps < 0.001) and more with frontal, parietal, and temporal lobe gray matter volumes (βs = 0.18 to 0.33, Ps ≤ 0.004) than occipital gray matter ($\beta = 0.12$, $P = 0.04$). The total sample needed to detect a 40% reduction in UDS3-EF change (n = 286) was ≈40% of the next best measure (F-words; n = 714).

Conclusions: The UDS3-EF is well suited to quantify EF in research and clinical trials and offers psychometric and practical advantages over its component tests.

KEYWORDS
Alzheimer's disease, cognition, composite score, executive function, item response theory, mild cognitive impairment, National Alzheimer's Coordinating Center, uniform data set

1 | INTRODUCTION

Executive functioning (EF) is a multifaceted cognitive domain comprising several component processes including set-shifting, inhibition, planning, organization, and working memory.[1,2] Intact EF is critical for completing daily activities and mediates functional decline in many neurologic conditions.[3,4] The neuroanatomical substrate of EF spans fronto-parietal networks, subcortical-cortical circuits, and

HIGHLIGHTS

- The Uniform Data Set v3.0 executive function (UDS3-EF) is an item response theory-based composite score.
- The UDS3-EF demonstrates convergent validity with other EF tests and EF brain regions.
- Using the UDS3-EF reduces sample size estimates for powering clinical trials.
- The UDS3-EF is well suited as a cognitive endpoint for clinical trials.

RESEARCH IN CONTEXT

1. Systematic review: Executive functioning (EF) is a multifaceted cognitive domain affected by common age-related pathologies. Cognitive composite scores offer psychometric and practical advantages over individual tests as clinical trial endpoints. Modern nonlinear adjustments for demographic factors may further improve the precision of cognitive composite scores.
2. Interpretation: We developed an EF composite score using the National Alzheimer's Coordinating Center Uniform Data Set (version 3.0) Neuropsychological Battery and developed norms using nonlinear adjustments. This composite was then validated in independent healthy control, mild cognitive impairment, and dementia cohorts, and shown to provide lower sample size estimates for clinical trials than the individual tests.
3. Future directions: The UDS3-EF composite score evidenced strong utility as a cognitive endpoint for measuring executive function in research and clinical trials. Further work may explore correlates with advanced neuroimaging and fluid biomarkers. Future composites may benefit from incorporating tablet-based EF measures with better psychometric properties.

interhemispheric connections.[5–9] It therefore is not surprising that EF deficits are frequently observed in aging populations due to common pathological changes like Alzheimer's disease (AD) and cerebrovascular disease.[10–13] There is an emerging need for sophisticated and psychometrically robust quantification of EF in older adults at greatest risk for such diseases.

Neuropsychological test batteries administered through large-scale, longitudinal aging studies typically include several EF measures.[14] This ensures measurement of multiple EF components but can increase "false positive" errors if interpreting individual low scores or declines as evidence of true impairment or cognitive worsening.[15] On the other hand, EF composite scores offer advantages such as better reliability, fewer statistical comparisons (ie, lower false positive risk),[15–17] and improved power to detect longitudinal change with smaller sample sizes.[17–21]

Prior work leveraging the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort demonstrated the benefits of an EF composite score. Gibbons et al. showed that, compared to individual EF test scores, the ADNI-EF composite score was associated with greater ability to detect change over time, better prediction of conversion to dementia, and stronger associations with AD biomarkers.[17] Similar findings have been shown for the National Institutes of Health Executive Abilities: Measure and Instruments for Neurobehavioral Evaluation and Research (NIH-EXAMINER),[7] a computerized EF battery that uses item response theory (IRT) to derive a cognitive composite.[21] Data collected and stored by the National Institutes on Aging Alzheimer's Disease Centers (ADC) would benefit from an EF composite score that maximizes measurement precision using novel psychometric approaches.

Clinicians and researchers often interpret cognitive test performance using a demographically adjusted standardized score derived from a normative reference group. *z*-Scores, for example, represent the difference between an individual's score and the normative group mean, divided by the normative group's standard deviation. Adjustments commonly are made by linearly correcting for age, sex, and education,[15] which assumes their effect on any given cognitive skill is constant across the spectrum of that variable (eg, the effect of age on EF is the same between the ages of 40 to 50 as it is between 70 and 80). Relying on linear models when evaluating nonlinear relationships can over- or underestimate the magnitude of a *z*-score, and nonlinear regression approaches can improve the precision of normative comparisons, particularly for age effects.[22] Taken together, validating an EF composite score derived from several EF tests and then standardized using nonlinear adjustments for key demographic factors could optimize EF measurement in aging research.

We used the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS v3.0)[23] to develop an EF composite score (UDS3-EF). We then validated the UDS3-EF composite in independent research cohorts. We hypothesized the UDS3-EF would correlate more strongly with independent EF measures and frontal, parietal, and temporal brain volumes than with non-EF tasks and brain regions that do not directly support executive functions (eg, occipital lobes). We also hypothesized that older adults diagnosed with mild cognitive impairment (MCI), dementia due to suspected AD, or behavioral variant frontotemporal dementia (bvFTD) would have significantly lower UDS3-EF scores, and that using the UDS3-EF as an outcome would improve longitudinal change detection compared to its component tests.

2 | METHODS

2.1 | UDS3-EF composite development and norming

The normative database was an extension of the NACC-UDS database of normal controls from 29 ADCs used by Weintraub et al.,[23] with additional data collected through May 2017. This same sample was used by Kornak et al. to create nonlinear *z*-scores of the UDS measures.[22] Informed consent was obtained by all participants, and permission was obtained from the NACC to perform this study. The downloaded dataset contained baseline data for 4287 control participants. We then restricted the dataset to those whose primary language was English (excluded n = 111; final n = 3507). Non-English speakers were excluded to reduce test variance attributable to language rather than executive functions and to allow future research that directly studies these models in non-English speakers.

2.1.1 | Scale construction

Several of the authors with expertise in neuropsychological assessment (AMS, BMA, KBC, DM, JHK) reviewed the battery to make an initial selection of items that could be considered indicators of EF. EF is a multifaceted domain, and most measures of this construct also rely heavily on other cognitive processes, particularly processing speed.[24] Given the limited availability of measures, and the goal of creating a composite that is sensitive to the changes associated with aging and its associated pathologies, we favored inclusiveness when selecting tests similar to previous efforts using ADNI data.[17] The tests that we chose were Digit Span Backwards (total correct), Trail Making Test (TMT) parts A and B (correct lines per minute), lexical fluency (F and L words—total correct), and semantic fluency (animal and vegetable fluency—total correct). Model building steps are described in detail in the supporting information. IRT was used to calculate factor scores. IRT-derived scores have the important property of being invariant to the specific items used. Therefore, these scores should provide unbiased estimates of the latent trait regardless of which subtests are included.

2.1.2 | Shape constrained additive model (SCAM)

Additive models relate the predictors to the dependent variable by estimating smoothly varying functions. Shape constrained additive models (SCAMs) can incorporate constraints over and above smoothness on the form of the fitted functions.[25] In particular, the constraint used in this paper is such that the functions increase or decrease monotonically. Relevant to this study, SCAMs allow incorporation of scientific knowledge about the behavior of neuropsychological scores with respect to particular predictors; specifically, performance on measures of executive function typically decreases with age and increases with education. The application of SCAM models to neuropsychological data have been detailed elsewhere[22] and are described in the supporting information.

2.2 | UDS3-EF validation cohort participants

We assessed UDS3-EF validity in older adult participants from the UCSF Hillblom Aging Network (controls), ADRC (MCI and AD-dementia), and/or the Advancing Research and Treatment for Frontotemporal Lobar Degeneration/Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects (ARTFL/LEFFTDS; healthy controls and bvFTD) projects. ARTFL/LEFFTDS controls were excluded if they had a genetic mutation known to cause FTD. All participants underwent comprehensive annual assessments including the UDS. A large subset also completed structural neuroimaging. We limited analyses involving magnetic resonance imaging (MRI) and cognitive testing to individuals completing both within 90 days.

The UDS3-EF was calculated for each participant using data from the first visit at which the UDS v3.0[23] was completed. Therefore, this represents the first exposure to UDS3-EF tests added to UDS v3.0 (F-words and L-words) but not necessarily other components included in prior UDS versions (eg, animal fluency). All participants spoke English as their primary language. All participants classified as MCI or dementia were suspected to have a primary AD etiology based on clinical history and available neuroimaging, biomarkers, and family history. Classification as cognitively normal, MCI, AD-dementia, or bvFTD was made through a multidisciplinary consensus conference. We excluded individuals diagnosed with language-predominant syndromes regardless of suspected pathology. Individual components of the UDS3-EF occasionally were available during consensus conference when determining functional status, but the UDS3-EF composite score was not.

2.3 | Other neuropsychological tests

EF tests not included in the UDS3-EF composite were used in validation analyses: Modified Trail Making Test,[26] Letter Fluency (D-words), Design Fluency, and the NIH-EXAMINER Executive Composite score. Non-EF tests performed were the Craft Story, Benson Figure, Number Location subtest of the Visual Object and Space Perception battery (VOSP), 15-item Boston Naming Test, Mini-Mental State Examination (MMSE),[27] and Montreal Cognitive Assessment (MoCA)[28] (see supporting information).

2.4 | Structural neuroimaging

T1-weighted structural MRI scans were obtained on a 3.0 Tesla Siemens TIM Trio scanner and a 3.0 Tesla Siemens Prisma Fit scanner at the University of California at San Francisco (UCSF) Neuroscience Imaging Center. Scanner parameters and processing steps are included in the supporting information.

2.5 | UDS3-EF validation analyses

All model building, IRT, and SCAM analyses were performed in R (version 3.6.1), while the remaining validation steps were performed in SPSS version 25 (IBM, Armonk, New York, USA). Raw test scores from other neuropsychological measures and region of interest (ROI) volumes were converted to z-scores based on the mean and standard deviation of a non-overlapping sample of cognitively healthy participants from the Hillblom Aging Network (n = 201 to 718 across tests). NIH-EXAMINER z-scores were based on the normative sample. We excluded participants with UDS3-EF standard error >0.75[7] (N = 8; 0 controls, two MCI, four AD-dementia, two bvFTD).

We performed four sets of validation analyses. First, we compared the UDS3-EF among diagnostic groups (controls, MCI, AD-dementia, bvFTD) cross-sectionally using analysis of variance (ANOVA).

Second, we assessed associations between the UDS3-EF and other test scores using linear regression covarying for age, sex, education, and Clinical Dementia Rating Scale Sum of Boxes (CDR-SB). Divergent validity was assessed through associations between the UDS3-EF and tests of memory, language, and spatial abilities.

Third, we assessed associations between the UDS3-EF and ROI gray matter volumes using linear regression covarying for age, sex, total intracranial volume, and CDR-SB. Putative "executive" regions included frontal gray matter, parietal gray matter, temporal gray matter, dorsolateral prefrontal cortex (DLPFC; caudal and rostral middle frontal gyrus), orbitofrontal cortex (OFC; medial and lateral orbital frontal gyrus), and anterior cingulate cortex (ACC; caudal and rostral anterior cingulate gyrus). We assessed potential divergent validity from total occipital and pericalcarine gray matter volume.

For all regression models, statistical significance was defined as $P < 0.005$ to account for multiple comparisons.

Last, we assessed the ability to detect changes in EF over time (UDS3-EF score without norming vs individual component tests) using longitudinal data. Consistent with published methodology,[29] we estimated annualized changed scores for the executive composite and for each of the component subtests. We included those with a second assessment that occurred within 2 years of baseline. We compared annualized change scores between groups using linear regression. We also used the annualized change score to estimate the sample sizes needed to detect a small (25%) and moderate (40%) reduction in the mean rate of decline in 12 months, with 80% power and alpha = .05 (two-sided). To improve comparability, we restricted the sample to the MCI and dementia patients that were not missing any data for this analysis; no bvFTD cases met these criteria.

3 | RESULTS

3.1 | Model fit

First, an exploratory factor analysis was conducted to inform the number of latent factors. Eigenvalues and the resulting scree plot strongly suggested a one-factor model: eigenvalue 1 = 3.25, 2 = 1.02, 3 = 0.89. We fit a confirmatory factor model without any residual covariances and extracted modification indices. Fit was poor for this initial model (C_2[df = 14] = 1779.3, $P < 0.001$; comparative fit index [CFI] = 0.82; root mean squared error of approximation [RMSEA] = 0.194). Consistent with expectations, modification indices suggested residual covariances between several measures. We observed greatly improved fit ($P < 0.001$ based on χ^2 difference test) when allowing residual covariances among semantic fluency measures, lexical fluency measures, and TMT A and B. Final model fit was excellent for most statistics (C_2[df = 11] = 130.72, $P < 0.001$; CFI = 0.99; Tucker-Lewis index [TLI] = 0.98). RMSEA (0.057) suggested good fit.[30] Standardized loadings are presented in Figure 1. Factor scores were calculated for the NACC sample; the resulting score distribution was symmetrical and bell shaped, and there was no obvious departure from normality based on a Q-Q diagnostic plot (Figures SA–B in supporting information).

3.2 | SCAM models

Figure 2 display plots of the UDS3-EF score against age and education. There was a clear nonlinear effect of age. This nonlinear adjustment was most noticeable at younger ages, where a linear fit would have likely overestimated the mean score, leading to overcalling impairments in this group. The additive sex effect was very small compared to the nonlinear age and linear education effects (Figure SC in supporting information).

3.3 | Validation sample description

We calculated UDS3-EF composite scores for 305 participants from UCSF research cohorts (96 controls, 84 MCI, 87 AD-dementia, 38 bvFTD) with varying availability across test- and MRI-specific analyses (Table SA in supporting information). Diagnostic groups did not differ significantly in sex distribution, years of education, or race (% White), but controls (mean ± standard deviation [SD] age = 65.2 ± 14.0 years) were on average younger than both MCI (70.5 ± 10.1) and AD-dementia (69.2 ± 9.4) groups ($P = 0.008$). The MCI, AD-dementia, and bvFTD groups had lower MMSE, lower MoCA, higher Geriatric Depression Scale, higher global CDR, and higher CDR-SB than controls (Table 1).

There were statistically significant differences in UDS3-EF scores between diagnostic groups ($P < 0.001$, partial eta squared = .53) in the hypothesized direction (Figure 3): controls (mean z = −0.08, 95% confidence interval [CI −0.10, 0.26]) > MCI (mean z = −0.92, 95% CI [−1.13, −0.71]) > AD-dementia (mean z = −2.53, 95% CI [−2.82, −2.25]) ≈ bvFTD (mean z = −2.72, 95% CI [−3.10, −2.34]).

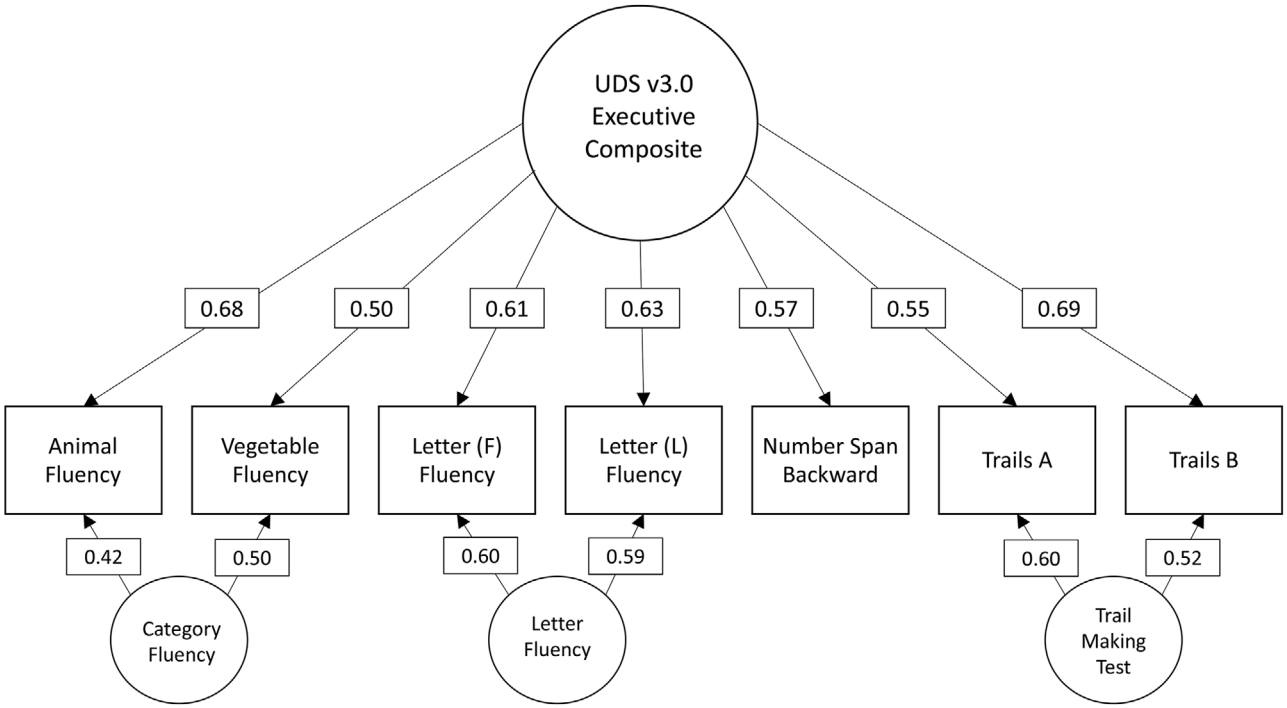

FIGURE 1 Standardized factor loadings for the final confirmatory factor analysis model. Trails, Trail Making Test; UDS3, Uniform Data Set version 3.0

3.4 | Cognitive and structural neuroimaging correlates of the UDS3-EF

Standardized (z) UDS3-EF scores, other cognitive test scores, and brain structure volumes are provided by diagnostic group (Table SB in supporting information). As shown in Figure 4A, the UDS3-EF correlated more strongly with other EF scores (lowest $\beta = 0.50$ [95% CI 0.39, 0.61]; highest $\beta = 0.85$ [95% CI 0.78, 0.92], $Ps < 0.001$) than with spatial (VOSP Number Location: $\beta = 0.42$ [95% CI 0.28, 0.56], $P < 0.001$), language (BNT-15: $\beta = 0.38$ [95% CI 0.24, 0.52], $P < 0.001$) and memory tests (Craft Story %Retention: $\beta = 0.21$ [95% CI 0.05, 0.36], $P = 0.008$; Benson %Retention: $\beta = 0.23$ [95% CI 0.10, 0.35], $P = 0.001$).

Figure 4B shows that the UDS3-EF generally correlated most strongly with frontal ($\beta = 0.18$ [95% CI 0.06, 0.31], $P = 0.002$), temporal ($\beta = 0.33$ [95% CI 0.22, 0.44], $P < 0.001$), and parietal lobes ($\beta = 0.26$ [95% CI 0.15, 0.38], $P < 0.001$), as well as frontal subregions (DLPFC: $\beta = 0.20$ [95% CI 0.07 0.33], $P = 0.003$; ACC: $\beta = 0.15$ [95% CI 0.02, 0.28], $P = 0.025$; OFC: $\beta = 0.14$ [95% CI 0.02, 0.26], $P = 0.027$), compared to the occipital lobe ($\beta = 0.12$ [95% CI 0.01, 0.23], $P = 0.039$) and pericalcarine cortex ($\beta = 0.06$ [95% CI −0.08, 0.21], $P = 0.372$).

3.5 | Longitudinal change

Mean annualized change for healthy controls (n = 32) was an increase of 0.11 units/year (SD = 0.36). Statistically significantly greater decline compared to controls was observed for the MCI group (n = 22; difference = −0.36 units/year, $P = 0.001$, [95% CI −0.57, −0.15), the AD-dementia group (n = 27; difference = −0.39 units/year, $P = 0.002$, [95% CI −0.63, −0.15]), and the bvFTD group (n = 20, difference = −0.42 units/year, $P = 0.001$, [95% CI −0.66, −0.18]). In a combined group of 26 MCI and AD-dementia participants with data for all measures, the UDS3-EF measure showed the greatest estimated decline (Table 2). The sample size needed to detect a 40% reduction in change (n = 286) was 40% of the next best measure (F-words: n = 714), and ≈16% of the sample required compared to CDR-SB (n = 1814).

4 | DISCUSSION

Our study describes the development, norming, and initial validation of an EF composite score derived from the NACC UDS3-EF. The UDS3-EF was developed using modern psychometric methods in a large sample of healthy controls from the NACC and then validated in participants from independent research cohorts classified as controls, MCI or dementia with expected AD pathology, or bvFTD. UDS3-EF scores were significantly worse in MCI than control participants, and significantly worse in AD-dementia and bvFTD than MCI participants. The UDS3-EF score correlated well with other EF test scores and was associated with frontal, parietal, and temporal lobe gray matter volumes, along with several frontal subregions (predominantly DLPFC). Longitudinal analysis showed greater declines in MCI, AD-dementia, and bvFTD compared to controls, and we estimated that a clinical trial using the UDS3-EF composite as a cognitive outcome would require less than

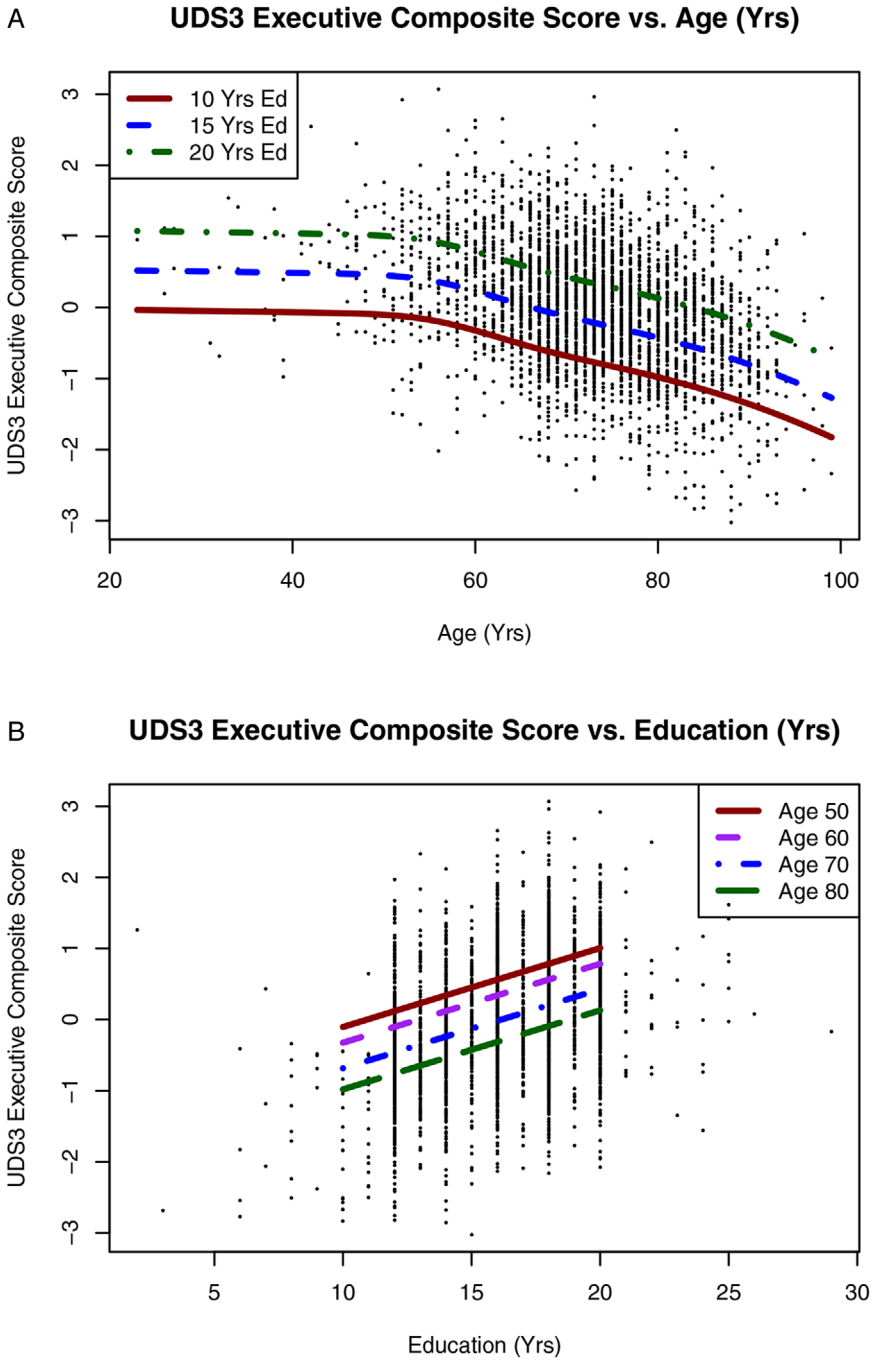

FIGURE 2 A–B, Scatterplots showing UDS3-EF scores in the NACC sample as a function of age with separate fit lines for years of education (A) and as a function of education with separate fit lines for age groups (B). NACC, National Alzheimer's Coordinating Center; UDS3, Uniform Data Set version 3.0

half the number of participants to detect a treatment effect than the best individual EF test and ≈80% fewer participants than the CDR-SB.

These results underscore the benefits of psychometrically robust composite score endpoints in research studies and clinical trials across the neurodegenerative disease spectrum.[17,18,21] Studies from ADNI have developed ADNI-specific composite scores for memory,[18] EF,[17] and global cognition.[31] An IRT-derived EF composite from the NIH-EXAMINER has been shown to detect longitudinal declines in asymptomatic carriers of mutations that cause FTD.[21] Notably, the NIH-EXAMINER composite score showed the strongest correlation with the UDS3-EF (>0.8) and participants with bvFTD obtained the lowest UDS3-EF scores, on average, of all diagnostic groups. Such work consistently demonstrates psychometric and practical advantages over using a single cognitive test score or multiple isolated test scores to quantify cognitive performance.

The UDS3-EF additionally improves precision of normative prediction models by accounting for nonlinear age effects. Despite the popularity of normative reference approaches linearly correcting for factors like age, sex, and education, concerns about underlying assumptions (eg, consistent test score variance across the demographic spectrum) have prompted alternative approaches including quantile[32,33] and nonlinear[22] regression. Our study extended recent work showing improved precision associated with nonlinear regression norms in the NACC dataset[22] to norm an EF composite score derived from UDSv3.0

TABLE 1 Descriptive statistics for the NACC development sample and independent validation cohorts

		NACC development sample	UCSF controls	UCSF MCI	UCSF dementia	UCSF bvFTD	Sig.[a]
N	–	3507	96	84	87	38	–
Age (y)	Mean (SD)	73.1 (10.1)	65.2 (14.0)	70.5 (10.1)	69.2 (9.4)	66.6 (7.4)	.008
	Mdn (IQR)	73 (67–80)	63.5 (50–79)	73 (64–78)	68 (62–76)	68 (60–72)	
Sex	% Female	64.2	60.4	42.9	50.6	50.0	.13
Education (y)	Mean (SD)	17.0 (7.8)	16.7 (2.3)	16.9 (2.6)	16.2 (2.4)	16.6 (2.7)	.37
	Mdn (IQR)	16 (14–18)	16 (16–18)	18 (16–20)	16 (16–18)	16 (14–18)	
Race/ethnicity	% White	82.1	71.9	67.9	72.4	86.8	.33
	% Black	13.9	0.0	2.4	1.1	2.6	
	% Asian	1.3	7.2	7.2	9.3	0.0	
	% Hispanic	2.2	0.0	0.0	0.0	0.0	
	% Native American	0.04	0.0	0.0	0.0	0.0	
	% Missing	0.1	20.8	22.6	17.2	10.5	
MMSE	Mean (SD)	–	28.7 (1.4)	26.9 (2.8)	21.0 (5.3)	22.4 (6.1)	<.001
	Mdn (IQR)	–	29 (28–30)	28 (25–29)	22 (18–25)	24 (20–27)	
MoCA	Mean (SD)	26.3 (2.8)	27.0 (2.1)	23.3 (4.0)	16.1 (5.6)	17.1 (6.2)	<.001
	Mdn (IQR)	27 (25–28)	27 (26–29)	24 (21–26)	17 (12–21)	17 (13–22)	
GDS	Mean (SD)	–	4.6 (4.9)	7.6 (5.7)	7.2 (5.2)	8.9 (7.3)	<.001
	Mdn (IQR)	–	3.5 (1–7)	6 (3–12)	7 (3–10)	8 (2–15)	
CDR Global	Mean (SD)	–	All = 0	0.5 (0.22)	0.9 (0.4)	1.1 (0.8)	<.001
	Mdn (IQR)	–	–	0.5 (0.5–0.5)	1.0 (0.5–1.0)	1.0 (0.5–2.0)	
CDR-SB	Mean (SD)	–	0.00 (0.09)	1.9 (1.2)	4.8 (2.1)	5.8 (3.9)	<.001
	Mdn (IQR)	–	0.0 (0.0–0.0)	2.0 (1.0–2.5)	4.5 (4.0-6.0)	5.0 (3.0–9.5)	

Abbreviations: bvFTD, behavioral variant frontotemporal dementia; CDR, Clinical Dementia Rating scale; CDR-SB, CDR Sum of Boxes; GDS, Geriatric Depression Scale; IQR, interquartile range; MCI, mild cognitive impairment; MMSE, Mini Mental State Exam; MoCA, Montreal Cognitive Assessment; NACC, National Alzheimer's Coordinating Center; SD, standard deviation; UCSF, University of California, San Francisco.
[a]Statistical significance of comparisons between UCSF cohorts using either analysis of variance or chi-square tests.

tests. We further observed the UDS3-EF to be approximately normally distributed even in our dementia cohorts. This has analytic and interpretive advantages over individual test scores that are typically skewed in populations with cognitive impairment.[17]

The UDS3-EF offers several practical advantages as a cognitive composite outcome score. We showed that longitudinal measurement of the UDS3-EF in a simulated clinical trial setting is more sensitive to detecting performance changes than its component scores. The UDS3-EF is estimated to achieve greater statistical power at smaller sample sizes, which has direct implications for clinical trial recruitment targets. Using the UDS3-EF as a cognitive score outcome would require ≈286 total participants to detect a 40% treatment effect, whereas ≈700 to 13,000 would be required if using any single component test, or 1814 if using CDR-SB. Gibbons et al.[17] report similar findings with the ADNI-EF composite score. These and other studies consistently show the benefit of composite score outcomes on measurement precision and sample recruitment goals in clinical trials.

Quantifying EF with a single score made up of numerous aspects of EF arguably simplifies the interpretive considerations inherent to obtaining multiple scores, such as inflated type I error and inter-test performance variability.[15] Recent work proposed a factor structure for the UDS v3.0 with separate factors for "speed/executive" (TMT A and B, lexical fluency) and "attention" (digit span) domains, along with category fluency scores within the "language" domain.[34] These are reasonable factor classifications but arguably spread several "executive" functions over separate, related domains.[1,2,7] The UDS3-EF distills EF measurement into a single score capturing several aspects of EF that otherwise might be untenable to interpret either in isolation or across different domain factors. A limitation of composite scores, however, is the masking of test-specific, within-domain scores that may inform diagnosis and recommendations on a case-by-case basis. The ability to interpret a single low test score might be preferred in settings in which potential "false positive" diagnoses are an acceptable trade-off for maximizing sensitivity to the earliest cognitive changes.

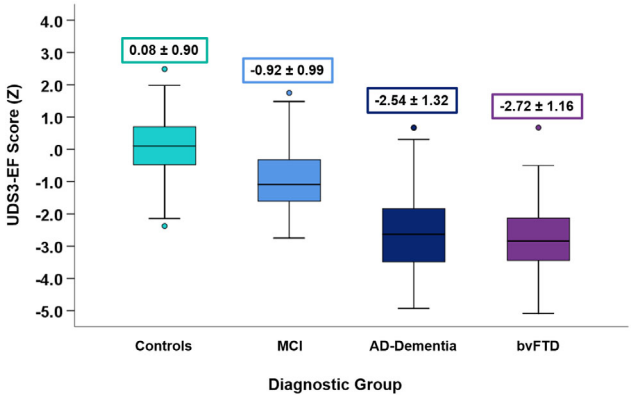

FIGURE 3 UDS3-EF score comparison between diagnostic groups. UDS3-EF score differences among controls, MCI, AD-dementia, and behavioral variant frontotemporal dementia diagnostic groups in the validation sample. The mean ± standard deviation of the UDS3-EF score is provided. Box plots represent the median (horizontal line) and interquartile range (top and bottom whiskers) of UDS3-EF scores. AD, Alzheimer's disease; MCI, mild cognitive impairment; UDS3, Uniform Data Set version 3.0

TABLE 2 Total sample sizes to detect treatment effects (25% and 40% reductions in score change) in sample of 26 MCI and AD-dementia participants with complete longitudinal cognitive data for UDS subtests

Measure	n	25%	40%
UDS measures			
UDS3-EF	26	728	286
TMT A	26	9016	3524
TMT B	26	1966	770
Digit span backward	26	1988	778
F-words	26	1822	714
L-words	26	4994	1952
Animal fluency	26	34680	13548
Vegetable fluency	26	3654	1428
Common measures			
CDR Sum of Boxes	24	4640	1814
MoCA	25	4222	1650
MMSE	20	13296	5196

Note: We also present commonly used outcome measures, when available (ns 20 to 25), to provide an independent reference against which to illustrate the power of the UDS3-EF for tracking longitudinal change.
Abbreviations: AD-dementia; Alzheimer's disease dementia; CDR, Clinical Dementia Rating Scale; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; TMT, Trail Making Test; UDS3-EF, Uniform Data Set v3.0 Executive Function composite score.

We limited our neuroimaging analyses to relatively crude associational investigations of discrete regional gray matter volumes. Data clearly support distributed brain regions and networks being responsible for EF tasks.[8] Converging evidence points toward frontal-parietal-subcortical networks as key EF regions, but additional areas including temporal and cerebellar regions also compellingly contribute to EF.[5,7–9,35–37] Consistent with prior work, the UDS3-EF correlated well with multiple regions of interest throughout the frontal lobe as well as the temporal and parietal lobes. These relationships held even when controlling for degree of functional impairment (CDR-SB), but were strongest in participants with the most functional impairment (ie, classified as having dementia; data not shown), which may reflect generalized brain changes associated with later-stage neurodegenerative disease and introduces specificity concerns.

The UDS3-EF was developed using test measures from the UDS v3.0 and likely does not fully capture all aspects of EF. Given the limited number of subtests, we made the decision to include tests such as TMT A and semantic fluency that could be considered measures of processing speed and language, respectively. Processing speed is intimately related to EFs,[24] and ultimately, a single metric that captures both speed and EFs might be the most sensitive to pathological changes in aging. Despite this, our composite showed good convergent validity with other EF measures. Researchers interested in deriving a purer estimate of EFs can choose to remove these tests when creating the composite. Future studies might consider including additional, psychometrically robust EF measures such as those from ADNI[17] or UCSF Brain Health Assessment.[38] The lack of racial and ethnic diversity in the NACC and UCSF cohorts (≈80% White) and overall high education levels is another limitation precluding generalizability. Similarly, there were few individuals in the NACC dataset below age 50 and above age 90, and therefore caution should be exercised when applying the normative corrections to those outside this range. We identified participants from UCSF observational research cohorts for validating the UDS3-EF. Slightly different sample sizes were available for different aspects of the analyses, which may result in biases associated with missing data. Some test scores may have been missing as a function of impairment level (eg, higher proportion of missing data in the dementia groups).

In conclusion, the UDS3-EF appears to be a valid composite measure of EF in older adults. The UDS3-EF appears well suited to quantify EF in research and clinical trials and offers several psychometric and practical advantages over its component tests. R scripts to calculate factor scores and normative lookup tables are available upon request to the corresponding author.

ACKNOWLEDGMENTS

This work is supported by the National Institutes of Health (grants K23AG061253, AG045390, NS092089, AG032306, AG021886, AG016976, K23 AG058752, R01EB022055, and L30AG057123) and the Larry L. Hillblom Foundation (2014-A-004-NET, 2018-A-006-NET, 2017-A-004-FEL, and 2018-A-025-FEL).

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG062428-01 (PI James Leverenz, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266

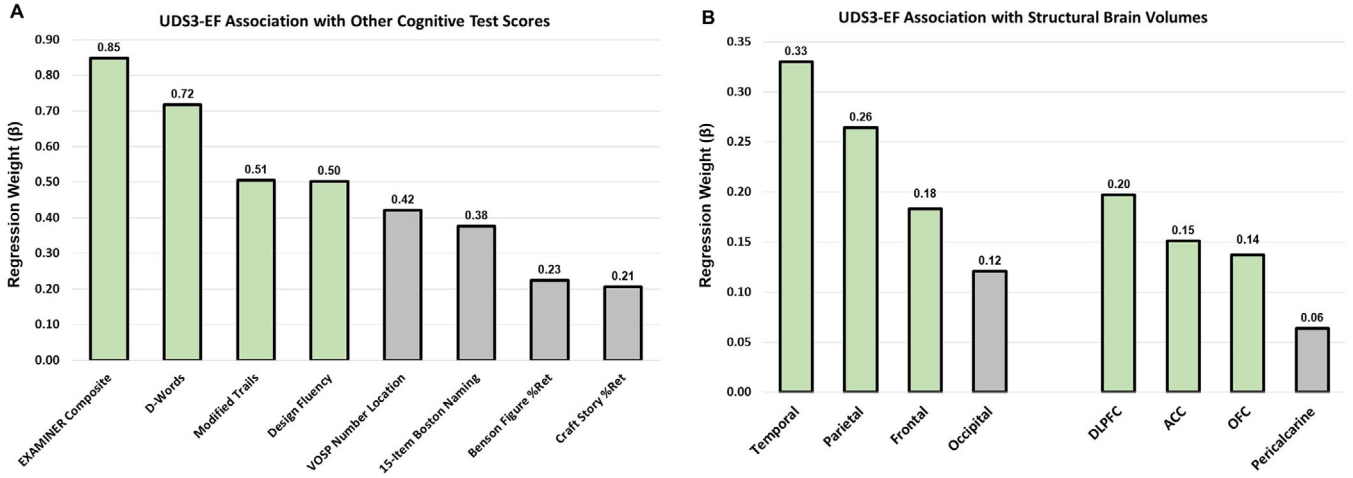

FIGURE 4 Association of the UDS3-EF score with other cognitive test scores and brain volumes. A, There were 305 participants with a UDS3-EF and a standard error <0.75 (age 68.0 ± 11.2 years old, 51% female, education 16.6 ± 2.5 years; AD-dementia, N = 87; mild cognitive impairment, N = 84; behavioral variant frontotemporal dementia, N = 38; controls, N = 96). Regression coefficients describe the strength of association between the UDS3-EF and other cognitive tests (N = 234 to 280 per test). Standardized beta-weights reflect associations after covarying for age, sex, education, and CDR Sum of Boxes ($P < 0.001$ for all except Craft Story and Benson Figure). B, Structural neuroimaging data were available for 210 participants (age 66.6 ± 11.4 years old, 52% female, education 16.6 ± 2.5 years; AD-dementia, N = 52; mild cognitive impairment, N = 46; behavioral variant frontotemporal dementia, N = 37; controls, N = 75). Regression coefficients describe the strength of association between the UDS3-EF and brain gray matter volumes. Standardized beta-weights reflect associations after covarying for age, sex, total intracranial volume, and CDR Sum of Boxes ($P < 0.005$ for all listed βs > .18). AD-dementia, Alzheimer's disease dementia; ACC, anterior cingulate cortex; CDR, Clinical Dementia Rating; DLPFC, dorsolateral prefrontal cortex; OFC, orbitofrontal cortex; UDS3, Uniform Data Set version 3.0

(PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P30 AG062429-01(PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

CONFLICTS OF INTEREST

Dr. Kornak reports providing expert witness testimony for Teva Pharmaceuticals in *Forest Laboratories Inc. et al.* versus *Teva Pharmaceuticals USA, Inc.*, Case Nos. 1:14-cv-00121 and 1:14-cv-00686 (D. Del. filed Jan. 31, 2014 and May 30, 2014) regarding the drug memantine; for Apotex/HEC/Ezra in *Novartis AG et a*l. versus *Apotex Inc.*, No. 1:15-cv-975 (D. Del. filed Oct. 26, 2015), regarding the drug fingolimod. He has also given testimony on behalf of Puma Biotechnology *in Hsingching Hsu et al.* versus *Puma Biotechnology, INC., et al.* 2018 regarding the drug neratinib. Dr. Kramer reports receiving research support from NIH, the Tau Research Consortium, and the Larry L. Hillblom Foundation, and he has provided consultation to Biogen. Dr. Rosen reports having a consulting agreement with Ionis pharmaceuticals. Sources of NIH funding support are listed above. All authors report no conflicts of interest directly related to this project.

REFERENCES

1. Packwood S, Hodgetts HM, Tremblay S. A multiperspective approach to the conceptualization of executive functions. *J Clin Exp Neuropsychol*. 2011;33(4):456-470.
2. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognit Psychol*. 2000;41(1):49-100.
3. Puente AN, Lindbergh CA, Miller LS. The relationship between cognitive reserve and functional ability is mediated by executive functioning in older adults. *Clin Neuropsychol*. 2015;29(1):67-81.
4. Cahn-Weiner DA, Boyle PA, Malloy PF. Tests of executive function predict instrumental activities of daily living in community-dwelling older individuals. *Appl Neuropsychol*. 2002;9(3):187-191.
5. Bettcher BM, Mungas D, Patel N, et al. Neuroanatomical substrates of executive functions: beyond prefrontal structures. *Neuropsychologia*. 2016;85:100-109.
6. Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol Rev*. 2006;16(1):17-42.
7. Kramer JH, Mungas D, Possin KL, et al. NIH EXAMINER: conceptualization and development of an executive function battery. *J Int Neuropsychol Soc*. 2014;20(1):11-19.
8. Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci*. 2012;12(2):241-268.

9. Reineberg AE, Banich MT. Functional connectivity at rest is sensitive to individual differences in executive function: a network analysis. *Hum Brain Mapp*. 2016;37(8):2959-2975.
10. Rabinovici GD, Stephens ML, Possin KL. Executive dysfunction. *CONTINUUM: Lifelong Learning in Neurology*. 2015;21. (3 Behavioral Neurology and Neuropsychiatry):646.
11. Swanberg MM, Tractenberg RE, Mohs R, Thal LJ, Cummings JL. Executive dysfunction in Alzheimer disease. *Arch Neurol*. 2004;61(4):556-560.
12. Chong JSX, Liu S, Loke YM, et al. Influence of cerebrovascular disease on brain networks in prodromal and clinical Alzheimer's disease. *Brain*. 2017;140(11):3012-3022.
13. Levit A, Hachinski V, Whitehead SN. Neurovascular unit dysregulation, white matter disease, and executive dysfunction: the shared triad of vascular cognitive impairment and Alzheimer disease. *GeroScience*. 2020;42(2):445-465.
14. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's disease centers' uniform data set (UDS): the neuropsychological test battery. *Alzheimer Dis Assoc Disord*. 2009;23(2):91.
15. Brooks BL, Sherman EMS, Iverson GL, Slick DJ, Strauss E. Psychometric foundations for the interpretation of neuropsychological test results. *The little black book of neuropsychology*. Berlin, Germany: Springer; 2011:893-922.
16. Iverson GL, Brooks BL. Improving accuracy for identifying cognitive impairment. *The little black book of neuropsychology*. Berlin, Germany: Springer; 2011:923-950.
17. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6(4):517-527.
18. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6(4):502-516.
19. Crane PK, Narasimhalu K, Gibbons LE, et al. Composite scores for executive function items: demographic heterogeneity and relationships with quantitative magnetic resonance imaging. *J Int Neuropsychol Soc*. 2008;14(5):746-759.
20. Mungas D, Beckett L, Harvey D, et al. Heterogeneity of cognitive trajectories in diverse older persons. *Psychol Aging*. 2010;25(3):606.
21. Staffaroni AM, Bajorek L, Casaletto KB, et al. Assessment of executive function declines in presymptomatic and mildly symptomatic familial frontotemporal dementia: nIH-EXAMINER as a potential clinical trial endpoint. *Alzheimers Dement*. 2020;16(1):11-21.
22. Kornak J, Fields J, Kremers W, et al. Nonlinear Z-score modeling for improved detection of cognitive abnormality. *Alzheimers Dement*. 2019;11(C):797-808.
23. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the Uniform Data Set (UDS). *Alzheimer Dis Assoc Disord*. 2018;32(1):10.
24. Salthouse TA. Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*. 2005;19(4):532.
25. Pya N, Wood SN. Shape constrained additive models. *Stat Comput*. 2015;25(3):543-559.
26. Kramer JH, Jurik J, Sha SJ, et al. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. *Cogn Behav Neurol*. 2003;16(4):211-218.
27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198.
28. Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. 1993;43(11):2412-2414.
29. Sakpal T. Sample size estimation in clinical trial. *Perspect Clin Res*. 2010;1(2):67-67.
30. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1-55.
31. Asken BM, Thomas KR, Lee A, et al. Discrepancy-Based evidence for loss of thinking abilities (DELTA): development and validation of a novel approach to identifying cognitive changes. *J Int Neuropsychol Soc*. 2019;26(5):464-479.
32. Koenker R. Regression quantiles. *Econometrica*. 1978:33-50.
33. Sherwood B, AX-H Z, Weintraub S, Wang L. Using quantile regression to create baseline norms for neuropsychological tests. *Alzheimers Dement*. 2016;2:12-18.
34. Kiselica AM, Webber TA, Benge JF. the uniform dataset 3.0 neuropsychological battery: factor structure, invariance testing, and demographically adjusted factor score calculation. *J Int Neuropsychol Soc*. 2020;26(6):576-586.
35. Koziol LF, et al. Consensus paper: the cerebellum's role in movement and cognition. *Cerebellum*. 2014;13(1):151-177.
36. Farrar DC, Mian AZ, Budson AE, et al. Retained executive abilities in mild cognitive impairment are associated with increased white matter network connectivity. *Eur Radiol*. 2018;28(1):340-347.
37. Nowrangi MA, Okonkwo O, Lyketsos C, et al. Atlas-based diffusion tensor imaging correlates of executive function. *J Alzheimers Dis*. 2015;44(2):585-598.
38. Possin KL, Moskowitz T, Erlhoff SJ, et al. The brain health assessment for detecting and diagnosing neurocognitive disorders. *J Am Geriatr Soc*. 2018;66(1):150-156.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Staffaroni AM, Asken BM, Casaletto KB, et al. Development and validation of the Uniform Data Set (v3.0) executive function composite score (UDS3-EF). *Alzheimer's Dement*. 2021;17:574–583.
https://doi.org/10.1002/alz.12214