Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 11 (2019) 797-808

Cognitive & Behavioral Assessment

Nonlinear Z-score modeling for improved detection of cognitive abnormality

John Kornak[a,*], Julie Fields[b], Walter Kremers[b], Sara Farmer[b], Hilary W. Heuer[a], Leah Forsberg[b], Danielle Brushaber[b], Amy Rindels[b], Hiroko Dodge[c,d], Sandra Weintraub[e], Lilah Besser[f], Brian Appleby[g], Yvette Bordelon[h], Jessica Bove[i], Patrick Brannelly[j], Christina Caso[k], Giovanni Coppola[h], Reilly Dever[a], Christina Dheel[b], Bradford Dickerson[l], Susan Dickinson[m], Sophia Dominguez[i], Kimiko Domoto-Reilly[k], Kelley Faber[n], Jessica Ferrall[o], Ann Fishman[p], Jamie Fong[a], Tatiana Foroud[n], Ralitza Gavrilova[b], Deb Gearhart[b], Behnaz Ghazanfari[q], Nupur Ghoshal[r], Jill Goldman[s], Jonathan Graff-Radford[b], Neill Graff-Radford[t], Ian M. Grant[e], Murray Grossman[i], Dana Haley[t], John Hsiao[u], Robin Hsiung[v], Edward D. Huey[s], David Irwin[i], David Jones[b], Lynne Jones[r], Kejal Kantarci[b], Anna Karydas[a], Daniel Kaufer[o], Diana Kerwin[w], David Knopman[b], Ruth Kraft[b], Joel Kramer[a], Walter Kukull[x], Maria Lapid[b], Irene Litvan[y], Peter Ljubenkov[a], Diane Lucente[l], Codrin Lungu[z], Ian Mackenzie[v], Miranda Maldonado[h], Masood Manoochehri[s], Scott McGinnis[l], Emily McKinley[aa], Mario Mendez[h], Bruce Miller[a], Namita Multani[q], Chiadi Onyike[p], Jaya Padmanabhan[l], Alexander Pantelyat[p], Rodney Pearlman[bb], Len Petrucelli[t], Madeline Potter[n], Rosa Rademakers[t], Eliana Marisa Ramos[h], Katherine Rankin[a], Katya Rascovsky[i], Erik D. Roberson[aa], Emily Rogalski-Miller[e], Pheth Sengdy[v], Les Shaw[i], Adam M. Staffaroni[a], Margaret Sutherland[z], Jeremy Syrjanen[b], Carmela Tartaglia[q], Nadine Tatton[m], Joanne Taylor[a], Arthur Toga[cc], John Trojanowski[i], Ping Wang[a], Bonnie Wong[l], Zbigniew Wszolek[t], Brad Boeve[b], Adam Boxer[a], Howard Rosen[a], on Behalf of the ARTFL/LEFFTDS Consortium

[a]*UCSF, San Francisco, CA, USA*
[b]*Mayo Clinic Rochester, Rochester, MN, USA*
[c]*Oregon Health and Science University, Portland, OR, USA*
[d]*University of Michigan, Ann Arbor, MI, USA*

Kornak J. has provided expert witness testimony for Teva Pharmaceuticals in *Forest Laboratories Inc. et al. v. 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 al. v. Apotex Inc.*, No. 1:15-cv-975 (D. Del. filed Oct. 26, 2015, regarding the drug Fingolimod; and for Puma Biotechnology in *Hsingching Hsu et al. vs. Puma Biotechnology, INC., et al.* 2018 regarding the drug Neratinib, Case No. 8:15-cv-00865. He has also provided testimony on behalf of Hikma Pharmaceuticals, USA, INC., et al. in *Amarin Pharma, Inc. and Amarin Pharmaceuticals Ireland Limited vs. Hikma Pharmaceuticals USA Inc., et al.* 2019 regarding the drug Vascepa. Case No. 2:16-cv-02562-MMD-NJK (Consolidated with 2:16-cv-02562-MMD-NJK). He receives research support from the NIH (R01EB022055, U01AG045390, U54NS092089 R01CA132870). Fields J receives research support from NIH. Kremers W receives research funding from AstraZeneca, Biogen, Roche, DOD, and NIH. Heuer H.W. receives research support from NIH. Forsberg L receives research support from NIH. Weintraub S. receives research support from the NIH. Appleby B. receives research funding from CDC (NU38CK000480). Brannelly P. employed by the Rainwater Charitable Foundation. Coppola G. receives research support from NIH. Dickerson B. receives research support from NIH. Dickinson S. is an staff at the Association for Frontotemporal Degeneration and a member of the National Institute for Neurological Disorders and Stroke Advisory Council. Faber K. receives research support from NIH. Foroud T. receives research support from NIH. Gavrilova R. receives research support from NIH. Ghoshal N. has participated or is currently participating in clinical trials of antidementia drugs sponsored by the following companies: Bristol Myers Squibb, Eli Lilly/Avid Radiopharmaceuticals, Janssen Immunotherapy, Novartis, Pfizer, Wyeth, SNIFF (The Study of Nasal Insulin to Fight Forgetfulness) study, and A4 (The Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease) trial. She receives research support from Tau Consortium and Association for Frontotemporal Dementia and is funded by the NIH.

https://doi.org/10.1016/j.dadm.2019.08.003

[e]Northwestern University, Chicago, IL, USA
[f]Florida Atlantic University, Boca Raton, FL, USA
[g]Case Western Reserve University, Cleveland, OH, USA
[h]Department of Psychiatry, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[i]University of Pennsylvania, Philadelphia, PA, USA
[j]Tau Consortium, Rainwater Charitable Foundation, Fort Worth, TX, USA
[k]University of Washington, Seattle, WA, USA
[l]Harvard University/MGH, Boston, MA, USA
[m]Association for Frontotemporal Degeneration, Radnor, PA, USA
[n]National Cell Repository for Alzheimer's Disease (NCRAD), Indiana University, Indianapolis, IN, USA
[o]University of North Carolina, Chapel Hill, NC, USA
[p]Johns Hopkins University, Baltimore, MD, USA
[q]University of Toronto, Toronto, Ontario, Canada
[r]Washington University, St. Louis, MO, USA
[s]Columbia University, New York, NY, USA
[t]Mayo Clinic, Jacksonville, FL, USA
[u]National Institute on Aging (NIA), Bethesda, MD, USA
[v]University of British Columbia, Vancouver, British Columbia, Canada
[w]UTSW, Dallas, TX, USA
[x]National Alzheimer Coordinating Center (NACC), University of Washington, Seattle, WA, USA
[y]UCSD, San Diego, CA, USA
[z]National Institute of Neurological Disorders and Stroke (NINDS), Bethesda, MD, USA
[aa]University of Alabama at Birmingham, Birmingham, AL, USA
[bb]Bluefield Project to Cure FTD, San Francisco, CA, USA
[cc]Laboratory of Neuroimaging (LONI), USC, Los Angeles, CA, USA

Abstract

Introduction: Conventional Z-scores are generated by subtracting the mean and dividing by the standard deviation. More recent methods linearly correct for age, sex, and education, so that these "adjusted" Z-scores better represent whether an individual's cognitive performance is abnormal. Extreme negative Z-scores for individuals relative to this normative distribution are considered indicative of cognitive deficiency.

Methods: In this article, we consider nonlinear shape constrained additive models accounting for age, sex, and education (correcting for nonlinearity). Additional shape constrained additive models account for varying standard deviation of the cognitive scores with age (correcting for heterogeneity of variance).

Results: Corrected Z-scores based on nonlinear shape constrained additive models provide improved adjustment for age, sex, and education, as indicated by higher adjusted-R^2.

Discussion: Nonlinearly corrected Z-scores with respect to age, sex, and education with age-varying residual standard deviation allow for improved detection of non-normative extreme cognitive scores.

Keywords: Generalized additive models; Heterogenous variance modeling; Neuropsychological testing scores; Nonlinear Z-score correction; Shape constrained additive models

Goldman J. is serving as a consultant to the Novartis Alzheimer's Prevention Advisory Board. She receives research support from NIH, HDSA, and New York State Department of Health (RFA # 1510130358). Graff-Radford N.R. receives royalties from UpToDate and has participated in multicenter therapy studies sponsored by Biogen, TauRx, AbbVie, Novartis, and Lilly. He receives research support from NIH. Grossman M. receives grant support from NIH, Avid, and Piramal; participates in clinical trials sponsored by Biogen, TauRx, and Alector; serves as a consultant to Bracco and UCB; and serves on the Editorial Board of *Neurology*. Hsiung R. has served as an investigator for clinical trials sponsored by AstraZeneca, Eli Lilly, and Roche/Genentech. He receives research support from Canadian Institutes of Health Research and the Alzheimer Society of British Columbia. Huey E. receives research support from NIH. Irwin D. receives support from NIH, BrightFocus Foundation, and Penn Institute on Aging. Jones D. receives research support from NIH and the Minnesota Partnership for Biotechnology and Medical Genomics. Kantarci K. served on the Data Safety Monitoring Board for Takeda Global Research & Development Center, Inc.; data monitoring boards of Pfizer and Janssen Alzheimer Immunotherapy; and research support from the Avid Radiopharmaceuticals, Eli Lilly, the Alzheimer's Drug Discovery Foundation, and NIH. Knopman D. serves on the DSMB of the DIAN-TU study, is a site PI for clinical trials sponsored by Biogen, Lilly and the University of Southern California, and is funded by NIH. Kramer J. receives research support from NIH. Kukull W. receives research support from NIH. Litvan I. was a member of the AbbVie advisory board, consultant for Toyama Pharmaceuticals, and member of the Biotie/Parkinson Study Group Medical Advisory Board. Her research is supported by the National Institutes of Health grants: 5P50 AG005131-31, 5T35HL007491, 1U01NS086659, and 1U54NS092089-01; Parkinson Study Group; Michael J Fox Foundation; International Parkinson and Movement Disorder Society; Parkinson Foundation; AVID Pharmaceuticals; C2N Diagnostics/AbbVie; and Bristol-Myers Squibb/Biogen. She receives her salary from the University of California San Diego and as Chief Editor of Frontiers in Neurology. Lucente D. receives research support from NIH. Lungu C. receives honoraria for editorial work from Elsevier, Inc.

1. Introduction and background

The Advancing Research and Treatment for Frontotemporal Lobar Degeneration consortium and the Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects are both generating data from families with either a strong family history of frontotemporal dementia (FTD) or mutations in genes known to be associated with FTD: microtubule associated protein tau (*MAPT*), progranulin (*GRN*), or chromosome 9 open reading frame 72 (*C9orf72*). Both of these studies have the common goal of identifying the most robust and reliable methods to track disease progression in familial dementia, so that clinical trials of disease-modifying therapies can be designed appropriately. To this end, it is important to identify indications of a clinical change as early and accurately as possible.

A common approach to identifying cognitive deficits based on neuropsychological test scores is to compare the scores to those of a large set of cognitively normal individuals. The scores are flagged as potentially cognitively deficient if they are significantly worse than the scores represented in the cognitively normal data set.

1.1. Naïve (uncorrected) Z-score approach

The simplest approach to determine whether an individual's score is outside of the normal range is to calculate Z-scores. The method is to find the difference between the individual's obtained test score and the expected score (which is the mean of the score for the normative sample) and then divide that difference by the standard deviation (SD) of the normative sample's scores for that particular measure. When calculating Z-scores, one should be mindful of the fact that for some tests, a high score indicates good performance, but for other tests (e.g., tests where time to complete a task is the outcome being measured), a high score indicates poor performance. Common practice is to reverse the +/− Z-score sign in these cases such that a positive Z-score indicates better-than-expected performance and a negative score indicates poorer-than-expected performance. If the individual's Z-score is significantly below expectation compared with the normative sample, then the individual is flagged as being potentially cognitively deficient.

A major limitation of using a simple Z-score approach, however, is that it does not account for an individual's age, sex, or education level. These factors are known to influence what a "normal" score would be for a particular individual. For example, in some tests, higher education gives the false appearance of potentially protecting against dementia (a.k.a. the "cognitive reserve hypothesis" [1]). That is, individuals with higher education are less likely to show their early indications of dementia because cognitively normal individuals with higher education levels generally do better on the tests and so they need to decline further to be detected

Mackenzie I. receives research funding from Canadian Institutes of Health Research. McGinnis S. has served as an investigator for clinical trials sponsored by AbbVie, Allon Therapeutics, Biogen, Bristol-Myers Squibb, C2N Diagnostics, Eisai Inc., Eli Lilly and Co., Genentech, Janssen Pharmaceuticals, Medivation, Merck, Navidea Biopharmaceuticals, Novartis, Pfizer, and TauRx Therapeutics. He receives research support from NIH. Miller B. receives research support from NIH. Onyike C. receives research support from the NIH, CIHR, the Jane Tanger Back Fund for Young-Onset Dementias, the Nancy H. Hall Memorial Fund for Geriatric Psychiatry Research, and the Joseph Trovato Fund. He is a site PI for a study funded by Biogen, Inc. Pearlman R. employed as President by the Bluefield Project to cure FTD, a nonprofit medical research foundation. Petrucelli L. receives research support from NIH. Potter M. receives research support from NIH. Rademakers R. receives research funding from NIH and the Bluefield Project to Cure Frontotemporal Dementia. Rankin K. receives research support from NIH.

Rascovsky K. receives research support from NIH. Shaw L. receives research support from NIH.

Staffaroni A.M. receives research funding from the Larry H. Hillblom Foundation (2018-A-025-FEL). Tatton N. was employed by the Association for Frontotemporal Degeneration. Toga A. receives research support from the NIH and the Alzheimer's Association. Trojanowski J. may accrue revenue in the future on patents submitted by the University of Pennsylvania wherein he is a coinventor and received revenue from the sale of Avid to Eli Lily as coinventor on Aβ amyloid imaging–related patents submitted by the University of Pennsylvania. He receives research support from the NIH and several nonprofits. Wong B. receives research support from the NIH.

Wszolek Z. supported by the NIH, Mayo Clinic Center for Regenerative Medicine, the gift from Carl Edward Bolch, Jr., and Susan Bass Bolch, The Sol Goldman Charitable Trust, and Donald G. and Jodi P. Heeringa. He has also received grant funding support from Allergan, Inc. (educational grant) and AbbVie (medication trials). Boeve B. has served as an investigator for clinical trials sponsored by GE Healthcare and Axovant. He receives royalties from the publication of a book entitled Behavioral Neurology of Dementia (Cambridge Medicine, 2009, 2017). He serves on the Scientific Advisory Board of the Tau Consortium. He receives research support from the NIH (U01 AG045390, U54 NS092089, R01 AG041797, P50 AG016574, U01 AG006786), the Mayo Clinic Dorothy and Harry T. Mangurian Jr. Lewy Body Dementia Program and the Little Family Foundation. Boxer A. receives research support from NIH, the Tau Research Consortium, the Association for Frontotemporal Degeneration, Bluefield Project to Cure Frontotemporal Dementia, Corticobasal Degeneration Solutions, the Alzheimer's Drug Discovery Foundation and the Alzheimer's Association. He has served as a consultant for Aeton, AbbVie, Alector, Amgen, Arkuda, Ionis, Iperian, Janssen, Merck, Novartis, Samumed, Toyama, and UCB and received research support from Avid, Biogen, BMS, C2N, Cortice, Eli Lilly, Forum, Genentech, Janssen, Novartis, Pfizer, Roche, and TauRx. Rosen H. has received research support from Biogen Pharmaceuticals, has consulting agreements with Wave Neuroscience and Ionis Pharmaceuticals, and receives research support from NIH.

Farmer S., Brushaber D., Rindels A., Dodge H., Besser L., Bove J., Bordelon Y., Caso C., Dever R., Dheel C., Dominguez S., Domoto-Reilly K., Ferrall J., Fishman A., Fong J., Gearhart D., Ghazanfari B., Graff-Radford J., Haley D., Hsiao J., Jones L., Karydas A., Kaufer D., Kerwin D., Kraft R., Lapid M., Ljubenkov P., Maldonado M., Manoochehri M., McKinley E., Mendez M., Multani N., Padmanabhan J., Pantelyat A., Ramos E.M., Roberson E., Rogalski-Miller E., Sengdy P., Sutherland M., Syrjanen J., Tartaglia C., Taylor J., and Wang P. have nothing to disclose. *Shared senior authors. Portions of this study were presented at the 2011 Alzheimer's Association International Conference, Paris, France, and the 2011 Clinical Trials on Alzheimer's Disease, San Diego, CA.*Corresponding author. Tel.: 415-514-8028; Fax: 415-514-8150.

E-mail address: john.kornak@ucsf.edu

relative to an uncorrected distribution of scores. Therefore, longitudinal studies using norms that do not account for education may fail to appropriately detect decline in those that are more highly educated or sometimes detect decline inappropriately among individuals who are less educated. (Note that although these norms are applied to longitudinal data, they are applied independently at each time point and not to measures of change.) By generating Z-score distributions specific to education levels (via modeling adjustments), we can determine whether an individual has an extreme Z-score relative to their own specific education level.

1.2. Z-score adjustment via linear regression

The Z-score approach to neurocognitive assessment has been extended by multiple groups to consider linear correction for age, sex, and education [2,3]. However, the linear assumption represents a strong constraint on the correction approach because it imposes the assumption of a linear relationship between each of age and education with the neuropsychological outcome scores in the regression. Furthermore, the standard linear model takes the homogeneous variance estimated for the residuals of the fitted regression and uses this constant value in the denominator of the Z-score estimate for all individuals. This use of constant SD in the Z-score estimates leads to severe underestimation or overestimation of the magnitude of the Z-score when the SD associated with a particular age, sex, and education score differs drastically from that of the mean SD of the linear regression residual errors. Edland et al. [4] had previously examined nonlinear correction of cognitive scores for the Cognitive Abilities Screening Instrument with respect to age, sex, and education. The authors dealt with the nonlinearity through Box-Cox transformation of the outcomes.

1.3. Z-score adjustment via nonlinear regression

In this article, we propose a direct nonlinear modeling approach for the Z-scores with respect to age, sex, and education to allow for nonlinear relationships with the neuropsychological outcomes examined. In addition, our approach explicitly incorporates a secondary nonlinear modeling of residual variance such that it can vary according to the covariates. This leads to improved determination of whether scores are extreme relative to the individual's age, sex, and education level.

2. Methods

2.1. Normative data

We used a slightly extended version of the National Alzheimer's Coordinating Center (NACC) database Uniform Data Set (UDS) of normal controls that was used by Weintraub et al. [3] (extended by additional data acquired by the NACC after December 2016). Informed consent was obtained by the NACC for all subjects in the study. Permission from the NACC has been obtained by this study to perform and publish this research on this data. These cross-sectional data were acquired from 29 Alzheimer's Disease Centers and considered visits between March 2015 and May 2017. The data set contained a total of 4287 individuals who were each tested one time. We removed individuals with missing education data and those who were aged 92 or over (numbers were small and outcomes erratic beyond this age), leaving 4193 individuals. We further restricted our normative data set to the 3461 Caucasians because there were only relatively small numbers for other racial groups, which overly limits size for building normative tables for those racial groups. Finally, we removed subjects that did not perform testing in English leaving 3430 participants.

2.2. Preprocessing steps

The data were further restricted to an age range of 40 to 91 by considering all individuals less than age 40 as being equivalent to individuals who are 40 years of age. The reason for this age shift (translation) was that the number of individuals was very small below age 40 and therefore sometimes led to poor model fits below age 40. In addition, most neuropsychological scores showed an approximately constant trend across the younger ages of 40 to 50, and thus, a constant effect for before 40 was thought to be a reasonable assumption on our part. Similarly, education levels of less than 10 years were mapped to 11 years of education and education levels above 21 years were counted as equivalent to 21 years due to the small number of subjects in those ranges.

2.3. Shape constrained additive model

Our approach of extending to nonlinear models with heterogeneous residual variance modeling is based on additive models [5,6], and more specifically, shape constrained additive models (SCAMs) [7]. Additive models enable the fitting of smoothly varying functions that relate the predictors to the outcome. SCAMs extend on additive models by constraining the shape of the fitted functions to be monotonically increasing or decreasing (they can also be used to enforce convexity/concavity, but we do not use that here). The monotonicity constraints are used to reflect our scientific knowledge that the neuropsychological scores that measure dementia generally decrease with age and increase with education level (assuming higher scores are good; the reverse is true for scores where higher scores imply poorer performance).

The general form for an additive model is

$$\begin{aligned} y_i &= \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \ldots + f_p(x_{pi}) + \varepsilon_i \\ &= \beta_0 + \sum_{k=1}^{p} f_k(x_{ki}) + \varepsilon_i \end{aligned}$$

where the term y_i represents the outcome value for individual i. The term β_0 is the intercept term for the model (i.e., the

fitted value of the outcome when all predictors are set to zero). The functions, $f_k(x_{ki})$, are smooth functions (i.e., not containing sharp fluctuations) of the covariates, x_{ki}, to be determined from the data. The smoothness is commonly imposed on the functions by defining them to be spline functions, that is, suitably constrained combinations of polynomial functions, and this is the approach taken here. The ε_i terms are assumed to be independently and identically distributed as normal with mean 0 and SD of σ.

2.4. Model fitting

SCAMs were fit separately to each of the following outcomes (variable names with capital letters are those used in National Alzheimer's Coordinating Center Uniform Data Set Data Dictionary found at https://www.alz.washington.edu/WEB/rdd_uds.pdf): Trail Making Test A (TRAILA), Trail Making Test B (TRAILB), Letter fluency F (UDSVERFC), Letter fluency L (UDSVERLC), Category fluency–animals (ANIMALS), Category fluency–vegetables (VEG), Multilingual Naming test–Test (or MINT) total (MINTTOTS), Number Span longest digit forward (DIGFORSL), Number Span longest digit backward (DIGBACLS), Craft memory–immediate (CRAFTVRS), Craft memory–delay (CRAFTDVR), Benson figure–copy (UDSBENTC), Benson figure–recall (UDSBENTD), Montreal Cognitive Assessment (or MoCA) total (MOCATOTS), Number Span forward total correct trials (DIGFORCT), and Number Span backward total correct trials (DIGBACCT) [3].

Models were originally fit allowing nonlinear corrections for age and education (each constrained to be monotonic), along with an additive term for sex; we do not have a nonlinear term for sex because it is a binary predictor (male/female). For all models, the nonlinear education effect was close to linear, so we refitted education in the SCAM as a simple linear term.

For the neuropsychological outcomes, the form of the SCAM we used was

$$y_i = \beta_0 + f_{\text{age}}(\text{age}_i) + \beta_{\text{sex}}\text{sex}_i + \beta_{\text{educ}}\text{educ}_i + \varepsilon_i$$

The only term that has a nonlinear relationship with the outcome in this model is age; the sex and education variables both have linear multipliers β_{sex} and β_{educ}, respectively.

The residuals for each outcome were then extracted from the corresponding fitted model, and the SD of the residuals at each age was estimated using a sliding window of width 11 years centered at the age value being considered; the SD at an age was estimated by the sample SD of all residuals within the corresponding window. A second SCAM was then fitted to the estimated SD as a function of age. (The choice of an 11-year window was based on a compromise between having enough data in each window vs. the risk of smoothing out the change with age too much.) Note that the original nonlinear model for the outcome was fitted using either a homogeneous SD assumption or occasionally modeling a linear relationship of age with SD whenever it appeared to lead to a better visual fit. The point at which we really capture the nonlinear component of the SD is in the second nonlinear model for SD described below. (Note that we provide explicit details on whether the homogeneous variance assumption was used along with other model settings for each outcome in Supplementary Materials).

We observed changing variability of residuals with increasing age. However, for education versus SD, we did not see a clear relationship that warranted adjustment, and so we did not consider the potential for varying SD with education any further. For a few outcomes (Category fluency–animals and Number Span longest digit forward), the SD appeared to be clearly nonmonotonic with respect to age, in which case we fit unconstrained additive models rather than SCAMs.

The additive model for the SD therefore took the following form:

$$SD_i = \beta_0 + f_i(\text{age}_i) + \varepsilon_i$$

where, $i=1, \ldots, M$, indexes the set of windowed SD estimates (one estimate for each age).

A lookup table for each outcome was generated based on these two SCAM/additive model fits. For each value of age, sex, and education level, an adjusted Z-score was generated, using the fitted mean and SD for that age, sex, and education level from the fitted SCAM/additive models.

The fitting of the nonlinear models often required some additional processing to the models or the data. For example, the direction of monotonic relationships (i.e., whether increasing or decreasing) or lack thereof needed to be specified; outliers were sometimes removed when they had a large effect on the model fit; SD additive model fits were sometimes weighted by the sample size (number of individuals) in each age window, and sometimes were not. The decisions as to the preferred modeling approach were made in discussions with clinical investigators. Full details of additional steps taken for each variable are given in Supplementary Material with this article.

For comparison purposes, we also fitted linear models corresponding to the models described in Weintraub et al. [3]. That is, we fitted linear models with age, sex, and education as additive predictors and with a constant residual variance.

$$y_i = \beta_0 + \beta_{\text{age}}\text{age}_i + \beta_{\text{sex}}\text{sex}_i + \beta_{\text{educ}}\text{educ}_i + \varepsilon_i$$

Note that the results from these fitted linear models differ from those for the UDS, version 3, calculator in the study by Weintraub et al. because our snapshot of the data was taken later than theirs and therefore includes more data.

Models were fitted using the R statistical programming environment [8] along with the mgcv package for additive models [6] and the scam package for SCAMs [7]. All models

Table 1
Comparisons of adjusted-R^2 for linear versus additive models

Neuropsych measure	Adjusted-R^2 linear model	Adjusted-R^2 additive model	Adjusted-R^2 SD additive model
Trail Making Test A	0.152	0.164	0.983
Trail Making Test B	0.161	0.183	0.989
Letter fluency F	0.050	0.055	0.302
Letter fluency L	0.061	0.066	0.000
Category fluency–animals	0.114	0.120	0.863
Category fluency–vegetables	0.137	0.145	0.000
Multilingual Naming Test (MINT) total	0.062	0.069	0.716
Number Span longest digit forward	0.021	0.021	0.911
Number Span longest digit backward	0.035	0.035	0.935
Craft Story memory–immediate	0.055	0.062	0.022
Craft Story memory–delay	0.062	0.068	0.177
Benson figure–copy	0.016	0.017	0.984
Benson figure–recall	0.084	0.086	0.959
MoCA total	0.140	0.149	0.996
Number Span forward total correct trials	0.027	0.027	0.732
Number Span backward total correct trials	0.040	0.040	0.360

NOTE. The adjusted-R^2 linear and adjusted-R^2 additive model columns indicate the adjusted-R^2 for the respective models. Note that the results from the linear models differ from those for the UDS, version 3, calculator in Weintraub et al. because it was based on fitting to a later (larger) snapshot of the data. The final adjusted-R^2 SD additive model column represents the adjusted-R^2 for the second SCAM model fitting the SD across ages (i.e., comparing a model that allows the SD to vary with a constant estimated SD across all the ages). Note that there is no comparison for the SD model because we are really comparing against a constant SD model which would have an R^2 of zero.

Abbreviations: MoCA, Montreal Cognitive Assessment; UDS, Uniform Data Set; SD, standard deviation.

were fitted with the default parameters, except that the SCAM package requires manual definition of the direction of monotonicity.

3. Results

We focus the results to illustrate the model fits for two outcomes: Trail Making Test B (TRAILS B), which is measured in seconds taken to complete (max. 300), and total number of animals (ANIMALS) named in 60 seconds (max. 77). Complete results for all remaining outcomes are provided between Supplementary Plots and Table 1.

3.1. TRAILS B

Fig. 1 displays plots of (A) TRAILS B versus age (years) and (B) TRAILS B versus education (years). SCAM-based fitted lines are added to Fig. 1A based on males with 10, 15, and 20 years of education and to Fig. 1B based on males of age 50, 60, 70, and 80. The (additive) sex effect was very small by comparison to the (nonlinear) age and (linear) education effects (i.e., the predicted TRAILS B scores were very close for the two sexes for a specific age and education level). Plots showing sex differences are given in Supplementary Materials for all outcomes and were actually

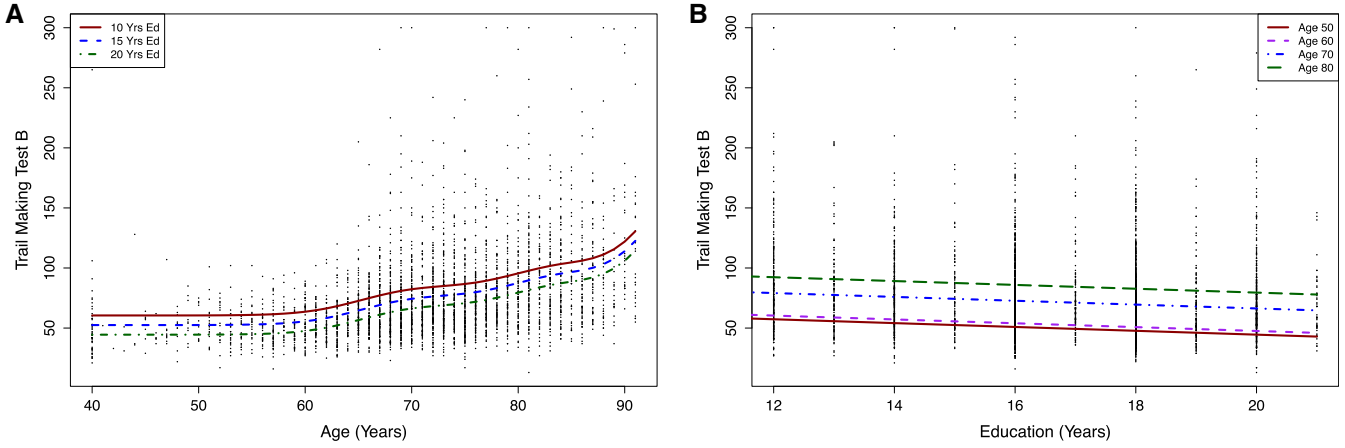

Fig. 1. (A) shows a plot of Trail Making Test B scores versus age in years (based on males with 10, 15, and 20 years of education), and (B) shows Trail Making Test B scores versus education level measured in years (based on males of age 50, 60, 70, and 80). The small difference between ages 50 and 60 in the education plot lines reflects the nonlinearity in the age plot where the lines are relatively flat at younger ages. The sex effects were very small in comparison with age and education. Plots showing sex differences are given in Supplementary Materials.

relatively small (compared with age and education effects) for all outcomes. Note that, the small difference observed between age 50 and 60 in the education plot lines of Fig. 1B reflects the nonlinearity of the age effect where the lines can be seen in Fig. 1A to be relatively flat at younger ages.

There is clear nonlinearity in the relationship between age and TRAILS B, with an increasing trend that accelerates with increasing age. The SCAM model provided for an adjusted-R^2 of 0.183 for the nonlinear model versus 0.161 for the corresponding linear correction model. (The adjusted-R^2 is a version of R^2 that is adjusted to account for the number of parameters in the model, the difference being to determine genuinely better fitting models from those that simply explain more variance by overfitting to noise). Note that although this observed increase in adjusted-R^2 does not appear to be large, it does represent a 13% increase, and this improvement is going to have its largest effect at the age extremes where the linear fit of the mean function would depart the most from the true mean. In particular, notice that between age 40 and 55 (critical for FTLD assessment), the mean score is very flat. A linear fit would severely underestimate the mean score in individuals under age 45 which would lead to overcalling extreme Z-scores (based on high test scores) in younger subjects. Note also that in these data, the between-subject variability is actually very high and that places an upper ceiling on what (adjusted) R^2 can be attained even with a perfect mean function. Indeed, it is this high variation in the scores of cognitively normal subjects that we are trying to capture in the normative scores distributions, but with removal of bias due to age, sex, and education.

Fig. 2 displays the window estimated SD versus age (blue line) and the corresponding fitted SCAM model (red line). Increasing SD was seen with age, and the SD increased by a factor of approximately 3 between age 40 and age 90. The nonlinear approach explicitly considers the age-specific value of the SD, in contrast to the linear correction approach, which assumes a constant SD. The large difference in variability across ages has major implications because the SD forms the denominator of the Z-score, thereby affecting up to a factor of 3 difference in the Z-score. The adjusted-R^2 of the nonlinear fit to SD when compared with assuming a constant SD is 0.989. Note that the R^2 for a constant fit to a varying SD is effectively 0 because none of the variation in the SD is explained by a constant value.

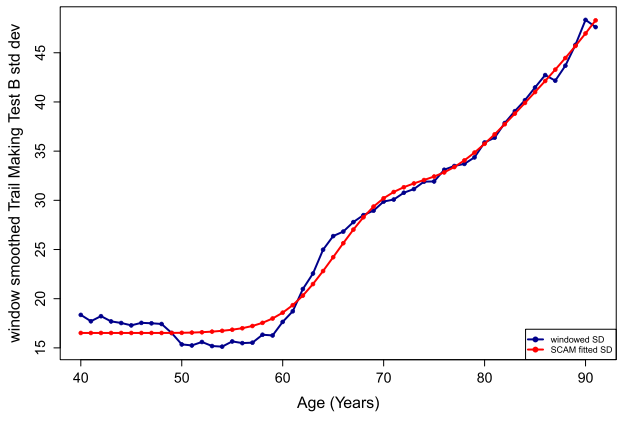

Fig. 2. Plot of SD of residuals for Trail Making Test B model versus age in years. Blue line shows raw SD curve based on sample SD estimates within the 11-year window centered on each point. The red line shows the corresponding SCAM model fit. Abbreviations: SCAM, shape constrained additive model; SD, standard deviation.

3.2. ANIMALS

Fig. 3 displays plots of (A) ANIMALS versus age (years) and (B) ANIMALS versus education (years). SCAM-based fitted lines are added to Fig. 3A based on males with 10, 15, and 20 years of education and to Fig. 3B based on males of age 50, 60, 70, and 80. The (additive) sex effect was small by comparison to the (nonlinear) age and (linear) education effects, and the interested reader can examine the (small) differences in the curves between the sexes via the plots in Supplementary Materials. No outliers were removed when fitting the SCAM models for ANIMALS.

Nonlinearity in the relationship between age and ANIMALS is again clear, but this time with a decreasing trend that accelerates with increasing age. The SCAM model provided for an adjusted-R^2 of 0.120 for the nonlinear model versus 0.114 for the corresponding linear correction model.

Fig. 4 displays the window estimated SD versus age (blue line) and the corresponding fitted SCAM model (red line). The SD was seen to vary with age, but for this outcome, the relationship did not appear to be monotonic, and so we modeled with an unconstrained additive model. The adjusted-R^2 of the fit when compared with assuming a constant SD is 0.863.

Plots of nonlinear fits to the complete set of neuropsychological outcomes are available in Supplementary Materials.

3.3. Overall results

Table 1 shows the improvements in adjusted-R^2 obtained by using nonlinear fitted models for all of the outcomes considered. The improvements in fit for the nonlinear models can be seen by noting that the adjusted-R^2 for the additive models are always equal to or higher than the corresponding adjusted-R^2 for the linear models. The final column shows the adjusted-R^2 in the second SCAM model for the outcome of SD compared with a constant SD model (i.e., by allowing it to vary with age rather than fixing to the estimated mean SD across all the data). Note that there is no comparison model for the SD model because we are really comparing against a constant SD model which would have an R^2 of zero. It can be seen that in many cases, the gains obtained by nonlinearly fitting the SD are very high, indicating a clear pattern of variation in the SD with age across many outcomes. For some outcomes, the gains to fitting variable SD are not as strong; in cases where the adjusted-R^2 for

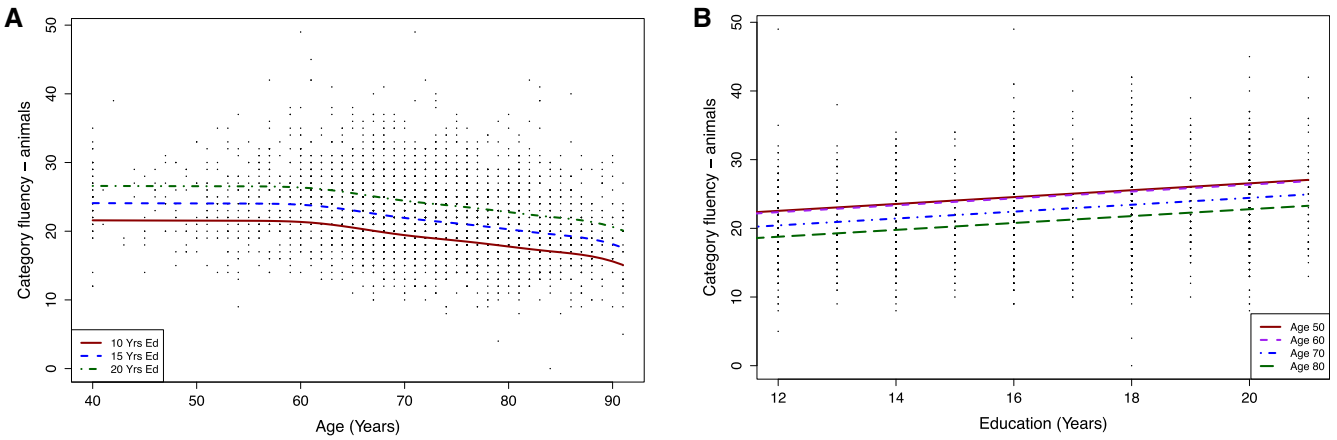

Fig. 3. (A) shows a plot of Category fluency–animals versus age in years (based on males with 10, 15, and 20 years of education), and (B) shows Category fluency–animals versus education level measured in years (based on males of age 50, 60, 70, and 80). The sex effects were small in comparison with age and education. Plots showing sex differences are given in Supplementary Materials.

SD is essentially zero, the result is indicating that the data are providing no indication of a consistent pattern of variation in the SD with age and that fitting a constant SD across all ages is adequate.

3.4. Comparison study

For illustrative purposes, a comparison was undertaken of Animals and Trails B Z-scores based on the UDS calculator [3], published normative data of Heaton et al. and Tombaugh et al. [9,10], and the adjusted calculator utilizing the nonlinear SCAM corrections. Z-scores were calculated for young, middle-aged, and older (ages 40, 55, and 75) males and females with varying degrees of education (10, 13, 16, and 20 years). Total words correct for Animals was randomly predetermined to be 18, and the time to complete Trails B was set at 65 seconds to directly compare methodologies. Consistent improvements (i.e., relative to published norms) in the interpretations of Z-scores were seen across these neurocognitive measures when allowing for nonlinear adjustment via SCAM fitting as compared with the NACC linear model for calculating Z-scores [3]. Table 2 depicts the Z-score comparisons between methods.

3.5. Case example

A 40-year-old male with 13 years of education presents for evaluation because he has a family history of FTD and has found out that he is a gene mutation carrier. He has no complaints, and his informant expresses no concern. He would like to establish a cognitive baseline. On testing, he generates 18 words on the animal fluency task and completes Trails B in 65 seconds. Utilizing the UDS norms, these scores are found to be markedly abnormal (see Table 2), $Z = -2.64$ and $Z = -3.27$, respectively. Similar scores are obtained on other measures of attention, language, and working memory, whereas the remainder of his cognitive profile is relatively normal. In the context of his family history and gene status, he is told that he has cognitive impairment consistent with FTD, that is, impairments in language and executive functioning. He is referred for further diagnostic testing, which is time-consuming, logistically and emotionally burdensome, and expensive. Ultimately, these diagnostics are negative, but he is told that an evolving process cannot be ruled out given his cognitive testing and gene status, although having the mutation does not determine with certainty that he will even develop the disease. He becomes anxious and despondent. Alternatively, with nonlinear adjustments from the proposed SCAM models, which are generally in agreement with well-established published norms, this individual would be found to be functioning in the low average range for animal fluency and within the average range for Trails B and other measures. He would

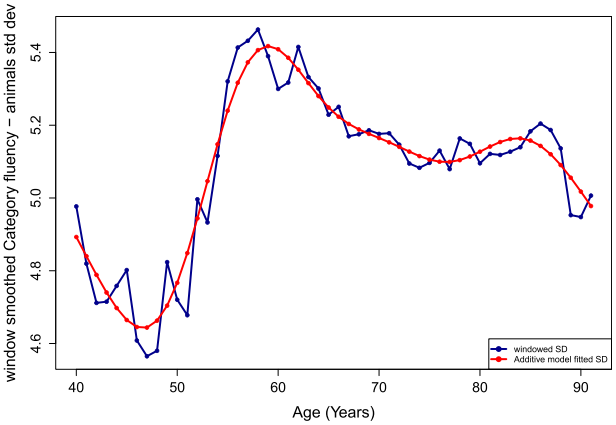

Fig. 4. Plot of SD of residuals for Category fluency–animals versus age in years. Blue line shows raw SD curve based on sample SD estimates within the 11-year window centered on each point. The red line shows the corresponding SCAM model fit. Abbreviations: SCAM, shape constrained additive model; SD, standard deviation.

Table 2
Normative data for Category fluency–animals (ANIMALS) and Trail Making Test B (TRAILB) based on different normative calculators

	Age	Education	Sex	ANIMALS UDS Raw 18 Words/Z	ANIMALS Tombaugh Raw 18 Words/Z	ANIMALS Nonlinear Raw 18 Words/Z	TRAILB UDS Raw 65 Seconds/Z	TRAILB Heaton Raw 65 Seconds/Z	TRAILB Nonlinear Raw 65 Seconds/Z
1	40	10	M	**−1.90**	−0.43	−0.73	−1.10	+0.10	−0.27
2		10	F	**−2.05**	−0.43	−0.60	−0.85	0.00	−0.18
3		13	M	**−2.64**	−0.72	−1.04	**−3.27**	−0.50	−0.56
4		13	F	**−2.79**	−0.72	−0.90	**−3.03**	−0.60	−0.48
5		16	M	**−3.39**	−0.72	−1.35	**−5.45**	−0.80	−0.85
6		16	F	**−3.54**	−0.72	−1.21	**−5.21**	−0.90	−0.77
7		20	M	**−4.38**	−0.72	**−1.76**	**−8.35**	−1.00	−1.24
8		20	F	**−4.53**	−0.72	**−1.62**	**−8.11**	−1.10	−1.16
9	55	10	M	−0.93	−0.43	−0.67	*+2.73*	+0.60	−0.24
10		10	F	−1.08	−0.43	−0.55	*+2.97*	+0.50	−0.16
11		13	M	**−1.67**	−0.72	−0.96	+0.55	0.00	−0.52
12		13	F	**−1.82**	−0.72	−0.83	+0.80	−0.10	−0.44
13		16	M	**−2.42**	−0.72	−1.25	**−1.63**	−0.30	−0.81
14		16	F	**−2.57**	−0.72	−1.12	−1.38	−0.40	−0.73
15		20	M	**−3.41**	−0.72	**−1.63**	**−4.53**	−0.50	−1.19
16		20	F	**−3.56**	−0.72	**−1.50**	**−4.28**	−0.60	−1.11
17	75	10	M	+0.36	+0.37	−0.12	*+7.83*	+1.60	+0.67
18		10	F	+0.21	+0.37	+0.00	*+8.08*	+1.50	+0.71
19		13	M	−0.38	−0.05	−0.42	*+5.65*	+0.90	+0.52
20		13	F	−0.53	−0.05	−0.29	*+5.90*	+0.80	+0.56
21		16	M	−1.12	−0.05	−0.71	*+3.48*	+0.80	+0.37
22		16	F	−1.27	−0.05	−0.58	*+3.72*	+0.60	+0.42
23		20	M	**−2.11**	−0.05	−1.11	+0.57	+0.60	+0.18
24		20	F	**−2.27**	−0.05	−0.98	+0.82	+0.40	+0.22

NOTE. Z-scores in bold are impaired ($\geq$ 1.5 SD below the mean); Z-scores in italics are $\geq$ 1.5 SD above the mean.
Abbreviations: UDS, UDS Calculator; Tombaugh, Published Norms of Tombaugh et al. (stratified by age/education); Nonlinear, Adjusted Calculator (Kornak et al.); Heaton, Published Norms of Heaton et al. (stratified by age/education/sex/Caucasian).

be counseled on the fact that just because he has the mutation, it does not mean he will develop FTD, and his scores at present are not concerning. It would be recommended that he return for reevaluation in about a year, or sooner if he started to show symptoms.

4. Discussion

In this article, we have described a nonlinear extension to generating age-, sex-, and education-corrected cognitive Z-scores. This nonlinear approach does a good job of modeling potential nonlinearities between the predictors (age, sex, and education) and the cognitive outcomes. Specifically, age was seen to display clear nonlinearity with respect to multiple cognitive outcomes. Furthermore, the additional modeling of the nonlinear relationships between age and SD has a high impact on determining extreme scores; some individuals would be given Z-scores that are too extreme (i.e., individuals in age ranges that have high variance) and others would be given Z-scores that are too conservative (i.e., individuals in age ranges that have low variance).

As shown in Table 2, extreme Z-scores were much more likely to be obtained using the NACC calculator without nonlinear adjustments. Specifically, according to the published norms for Animals, none of the Z-scores calculated, regardless of age, sex, or education, were abnormally low or abnormally high (i.e., $\geq \pm 1.5$ SD from the mean) for a set point of 18 words generated. The same was true for completing Trails B in 65 seconds. Nearly, all of the Z-scores obtained from the NACC calculator for both of these measures fell to one extreme or the other, most frequently toward impairment (-1.67 to -4.53 for animals and -1.63 to -8.35 for Trails B). In contrast, no Z-scores for Trails B using the nonlinear (SCAM) adjusted model were abnormally high or low. Only 4 of the 24 Z-scores calculated for Animals using the nonlinear (SCAM) adjusted model deviated from the norm, and for those that did, it was only to a mild degree (range of -1.50 to -1.76).

4.1. Limitations

There are some limitations with our nonlinear modeling approach.

First, we needed to occasionally manipulate data and variables, such as removing outliers or changing the form of the smoothing function (monotonic vs. nonmonotonic, etc.). Invariably, nonlinear modeling has increased difficulties over linear approaches with respect to linear approach and often require user manipulation. Clearly, a fully automated process for model estimation that provides completely accurate results every time is the ideal. However, although we

acknowledge the need for manipulation in model estimation is as an issue, we argue that it is less problematic than when performing statistical inference. The goal here is optimal estimation of Z-scores, with optimal adjustment for variability, and not testing for statistically significant effects. The concern about potential manipulation (with implicit multiple comparison concerns) and possible overfitting are less central to estimation. Future research should carefully evaluate the generalizability of our models on Advancing Research and Treatment for Frontotemporal Lobar Degeneration/Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects patient data over time.

Second, the SD windowing estimation approach is admittedly *ad hoc*. To the best of our knowledge, existing methodology and software is not readily available for the simultaneous estimation of nonlinear trend for both outcome and residual SD with respect to a continuous predictor. Future research could extend the methodology to simultaneous (joint) modeling for both the outcome and nonlinear heterogeneous SD.

Third, there are clearly non-normally distributed errors for many outcomes. Again, because statistical inference is not the primary concern, the effects of this limitation are mitigated. However, in our joint modeling of outcome and SD methodology development, we plan to incorporate the possibility of non-normal errors. Furthermore, there were a couple of outcomes (i.e., Number Span longest digit forward and longest digit backward) that were so heavily discretized (integer scale of 3 to 9, and 2 to 8, respectively) that we do not advocate performing any kind of corrected Z-score estimation for these outcomes.

Fourth, we have less data near the edges of our age and education ranges. Consequently, when considering individuals with age and/or education close to or beyond the limits of the data, it is important to be aware of the context of the model predictions. Indeed, when interpreting scores for individuals with covariates that are associated with few data points in the UDS, it would be important to report on the additional model uncertainty involved.

Fifth, our modeling approach treats the effects from each predictor as additive. Extension to nonadditive age and sex are available within the additive model framework; these extended models allow for arbitrary two-dimensional smooth functions of pairs of predictors (and can be extended to yet more dimensional predictors), allowing for complex interactions between these predictors. Although the extended modeling can provide improved fit in many cases (internal validity), such models are more difficult to interpret, and therefore there is a stronger need for external validation. This led to our decision to initially avoid such models for our Z-score correction process.

Sixth, a side effect of the nonlinear fitting approach with heterogeneous variance is that it is not always true that a Z-score will improve with age (i.e., become less extreme) for a constant value of a test score (which is what one might initially expect given that you expect decline with age). For example, if the SD is increasing with age at a faster rate than the mean score that is declining with age, then a nonchanging test score could become less extreme relative to the normative distribution as age increases. Whether this is viewed as a negative depends on your perspective, but when the objective is to determine whether or not a score is extreme given a specific age, it is the correct approach to consider the variance specific to that individual's age in the calculation. Note also that this effect of a constant score becoming less extreme with respect to the normative distribution with increasing age might be exacerbated for outcomes where the residuals are not normally distributed. In particular, scores which have a limiting value at certain ages (i.e., ceiling or floor effects) because the SD can change radically over a short age range. The fact that our approach accounts for this changing variance is positive relative to conventional approaches in that we are still able to detect which observations are extreme at a specific age. Further improvements might potentially be made on this front by explicitly modeling the non-normality in the data.

Seventh, our data was only based on Caucasians. Normative tables should be developed in the future for other racial groups as appropriate data become available.

Finally, we note that our nonlinear Z-score approach to flagging individuals with cognitive deficiency is based only on normative individuals, that is, we are examining what are unusual scores for normal controls, who are from different sites and in different settings. The conventional method for identifying abnormal cognitive performance is to compare an individual's obtained cognitive scores to scores from normative samples that are derived from studies examining large numbers (usually) of healthy individuals at one time point. These samples may or may not include individuals that at some point go on to demonstrate cognitive impairment. This potentially mixed sample actually *decreases* the ability to detect early cognitive changes, particularly if normative data are derived from small samples. Ideally, one would want a *robust* normative dataset that is derived from cognitively normal individuals who are subsequently tracked longitudinally and shown to be stable over time. Several studies have found superiority of robust norms over conventional norms in identifying cognitive impairment at baseline [11–13]. By utilizing robust norms, comparison of test scores in individuals with scores that decline over time to scores from those who remain cognitively stable has greater diagnostic utility both in terms of current status and predicting change over time. Unfortunately, the current normative sample has not been assessed longitudinally, so we do not have the capability to develop robust norms at this time.

The changing of environments is a concern when examining normal controls because the distribution of cognitive

scores will differ according to whether data are collected in rural versus urban, multisite versus single site, hospital versus specialty clinic settings, etc. Examining such differences between environments needs to be a focus of future research as more data become available.

5. Conclusions

Allowing nonlinear model fitting via SCAM models provides improvement in fit for the relationship between the predictors (age, sex, and education level) and neurocognitive outcomes in control normal relative to linear models. The nonlinear modeling thereby leads to adjusted Z-scores that are more representative of the departures from cognitively normal levels relative to their specific age, sex, and education level.

Hence, adjusting Z-scores using nonlinear SCAM models provides improved adjustment for age, sex, and education level compared with linear adjustment.

The ultimate danger in basing judgments on inaccurate and unreliable scores is misdiagnosing an individual, which results not only in inadequate and/or inappropriate care but also limits our ability to develop disease-modifying treatments.

Acknowledgments

This study was supported by grants U01 AG045390 (PIs Bradley F. Boeve and Howard J. Rosen), U54 NS092089 (PI Adam Boxer, MD), and R01 EB022055 (PI John Kornak, PhD).

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), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), 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), P50 AG005131 (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), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.dadm.2019.08.003.

RESEARCH IN CONTEXT

1. Systematic review: The Z-score approach to normalizing cognitive scores has had a long history using the simple approach of subtracting the mean and dividing by the standard deviation. More recently, linear methods corrected for age, sex, and education have been adopted, so that ensuing adjusted Z-scores better represent whether a score is outside the expected, or "normal," range.
2. Interpretation: Cognitive scores versus each of age and education show clear nonlinear relationships and varying residual standard deviation with age. Adjusted Z-scores based on nonlinear shape constrained additive models (SCAMs) improve estimation of the degree of departure from the norm more precisely than linear or unadjusted Z-scores.
3. Future directions: Improved Z-score estimation allows for improved assessment of cognitive decline and treatment effects, leading to improved clinical diagnosis and treatment planning. We will post the nonlinear calculator on the National Alzheimer's Association Uniform Data Set website: https://www.alz.washington.edu/WEB/data_descript.html.

References

[1] Tucker-Drob EM, Johnson KE, Jones RN. The cognitive reserve hypothesis: a longitudinal examination of age-associated declines in reasoning and processing speed. Dev Psychol 2009;45:431–46.
[2] Berres M, Monsch AU, Bernasconi F, Thalmann B, Stähelin HB. Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. Stud Health Technology Inform 2000:195–202.
[3] Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, et al. Version 3 of the Alzheimer disease centers' neuropsychological test battery in the Uniform Data Set (UDS). Alzheimer Dis Associated Disord 2018;32:10–7.
[4] McCurry SM, Edland SD, Teri L, Kukull WA, Bowen JD, McCormick WC, et al. The cognitive abilities screening instrument (CASI): data from a cohort of 2524 cognitively intact elderly. Int J Geriatr Psychiatry 1999;14:882–8.
[5] Hastie TJ, Tibshirani RJ. Generalized Additive Models. Boca Raton, Fla: Chapman and Hall/CRC; 1990.

[6] Wood Simon. Generalized Additive Models: An Introduction with R. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 2017.
[7] Pya N, Wood SN. Shape constrained additive models. Stat Comput 2015;25:543–59.
[8] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available at: https://www.R-project.org/. Accessed November 26, 2019.
[9] Heaton RK, Miller SW, Taylor MJ, Grant I. Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz, FL: Psychological Assessment Resources; 2004.
[10] Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Arch Clin Neuropsychol 1999;14:167–77.
[11] De Santi S, Pirraglia E, Barr W, Babb J, Williams S, Rogers K, et al. Robust and conventional neuropsychology norms: diagnosis and prediction of age-related cognitive decline. Neuropsychology 2008; 22:469–84.
[12] Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. Archives of Clinical Neuropsychology 2008;23:531–41.
[13] Pedraza O, Lucas JA, Smith GE, Petersen RC, Graff-Radford NR, Ivnik RJ. Robust and expanded norms for the Dementia Rating Scale. Archives of Clinical Neuropsychology 2010;25:347–58.