Appendix 2 (as supplied by the authors): Derivation of the model

Initial analyses of the training sample suggested that a linear model taking into account age and education could predict Mini-Mental State Examination scores. Considering that subjects had up to three measures in the Canadian Study of Health and Aging[1] we investigated the model using repeated measures regression analyses. We explored three correlation structures (compound symmetry, order 1 autoregressive, and variance components) and chose compound symmetry as it presented the lowest Akaike Information Criterion. We attempted to extract a theoretical working model from the data-driven results. Supplementary Appendix 3 illustrates the relationship between age and Mini-Mental State Examination, and education and Mini-Mental State Examination. As suggested by cubic smoothing splines, we postulated that age was quadratically associated with Mini-Mental State Examination while a linear relationship appeared between education and Mini-Mental State Examination; the latter becoming null past 14 years of education.

Observations noted above lead us to propose a candidate regression model for Mini-Mental State Examination (M) that would include age (A) as a quadratic factor, also supported by comparing the Akaike Information Criterion for linear and cubic functional forms, education (E) which we recoded past the 14 years cut point as a linear factor, and an interaction between age and education.

$$\text{Model 1: } M = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 E + \beta_4 A \cdot E$$

Because of correlated errors due to repeated measures, we estimated the parameters of the regression model using maximum likelihood methods for repeated measures. Using the Akaike Information Criterion, we found that a compound symmetry correlation structure fit the data appropriately. Table 2 (see Results section) presents the parameter estimates, confidence interval bounds and significance levels of Model 1.

 Copyright © 2017 Joule Inc. or its licensors

We further simplified the model for practical use. We begin with the postulated linear model presented in the article:

$$\text{Model 1: } M = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 E + \beta_4 A \cdot E$$

Simple algebraic manipulation allowed us to reformulate this model:

$$\text{Model 1: } \frac{M - \beta_0}{A} = \beta_1 + \beta_2 A + \beta_3 \frac{E}{A} + \beta_4 E$$

From the parameter estimation, results presented in the text, we deducted that β_0 and β_3 are not significantly different from 0. Thus, a first simplification arose:

$$\text{Simplified Model 1: } \frac{M}{A} = \beta_1 + \beta_2 A + \beta_4 E$$

Conceptual simplification generated Model 2:

$$\text{Model 2: } \frac{M}{A} = \beta_1 + \beta_2 \left(A + \frac{\beta_4}{\beta_2} E \right)$$

Because this is a derivation of Model 1, there was no need for further parameter estimation. We therefore substituted the values from Table 1 (see Results section), and obtained the estimated model:

$$\text{Model 2: } \frac{M}{A} = 0.786 - 0.00577(A - 0.515E)$$

For further convenience, we multiplied both sides of the equation by 1,000:

$$\text{Model 2: } \frac{M}{A} \times 1000 = 786 - 5.77(A - 0.515E)$$

 Copyright © 2017 Joule Inc. or its licensors

And proposed to name the left-hand side of the equation the cognitive quotient:

$$Cognitive\ quotient = \frac{M}{A} \times 1000$$

And proposed to name the right-hand side of the equation standardized age:

$$Standardized\ age = A - 0.515E$$

While the above equation was used for all analyses reported in this manuscript, for ease of use, and with very minimal loss of precision, we propose that clinicians use $Standardized\ age = A - 0.5E$. Since standardized age's scale uses age and education, we termed this scale 'standardized years'. The estimated Cognitive Quotient Model can then be formulated as:

$$Cognitive\ quotient = 786 - 5.77S_A$$

Since every individual entered the study at different standardized age, we used the model to project cognitive quotient scores at 60 years of standardized age, also termed cognitive $quotient_{60}$. We then chose quantiles of this projected cognitive $quotient_{60}$ value so that spacing between two curves would represent an annual drop over standardized years which should maximize classification efficiency. We started from the 10^{th} percentile and built up using discrimination declines that optimized Youden index and clinical applicability as the spacing between percentiles.

Reference

1. The Canadian Study of Health and Aging: risk factors for Alzheimer's disease in Canada. *Neurology*. 1994;44:2073-80.