Comprehensive Demographic Correction Improves Sensitivity and Reduces Bias in Cognitive Assessment

David L. Woods[1], Kathleen Hall[1], Isabella Jaramillo[1], Mike Blank[1], Kristin Geraci[1], Peter Pebler[1], and David K. Johnson[2]

[1]Neurobehavioral Systems, Inc., Berkeley, CA, [2]UC Davis Alzheimer's Disease Research Center, Walnut Creek, CA

Introduction

Standard cognitive test scoring corrects for the effects of age, education, and gender (AEG models). However, the standard scoring model ignores other demographic factors that may significantly affect test performance. We developed a Comprehensive (C) model that expands the standard model and includes vocabulary, race/ethnicity, and other demographic factors that significantly influence performance.

Methods

Overview. We developed a Comprehensive (C-) scoring model for the California Cognitive Assessment Battery (CCAB) to expand on the standard age + education + gender (AEG) scoring model by correcting for additional predictors, including vocabulary, age², race/ethnicity, Hispanic ethnicity, socioeconomic status (SES), computer use, and daily medications (a proxy for health and comorbidities)..

Participants. A normative sample of 1,914 community-dwelling adults ages 18–90 (mean age 53.9 years) with substantial racial and ethnic diversity (38% White, 23% Black, 18% Asian, and 22% mixed-race participants. All participants were telemedically tested with the CCAB in their homes (Woods et al., 2024).

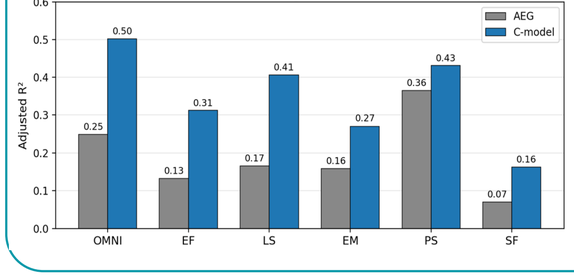
Tests, scores, and cognitive domains. Each participant completed 24 tests, generating 118 scores including Speech and Language Biomarkers (SLBs). Scores were analyzed individually and, based on correlation analysis, grouped into five cognitive domains: Executive Function (EF), Processing Speed (PS), Memory (EM), Language (LS), and Speech Fluency (SF).

C-model fitting. To avoid model overfitting, the C-model included only robust predictors for each test and domain score, identified using stability-selection LASSO with a strict selection criterion (λ = 1 SE) in 300 random 80/20 splits, with the further requirement that predictors were retained in at least 80% of samples. C-model coefficients were then estimated from 1,000 bootstrap 80/20 samples drawn from the 1,914 participants, and predictors with coefficients <0.10 were eliminated.

C-model generalizability. Cross-sample frozen coefficient validation was used to evaluate model generalizability. Models were developed in one cohort (1,083 older subjects or 881 younger subjects) and evaluated using frozen-coefficient predictions in the complementary cohort and the full dataset.

C-model parsimony. LASSO produced parsimonious models: the mean number of retained predictors was 2.81 (range 1–6) across 118 measures—only marginally less than the standard model's three mandatory predictors Three predictors dominated C-model retention: vocabulary (72.9%), age (66.9%), and race (61.0%) dominated model selection, whereas sex and education were retained in only 31.4% and 11.9% of measures, respectively.

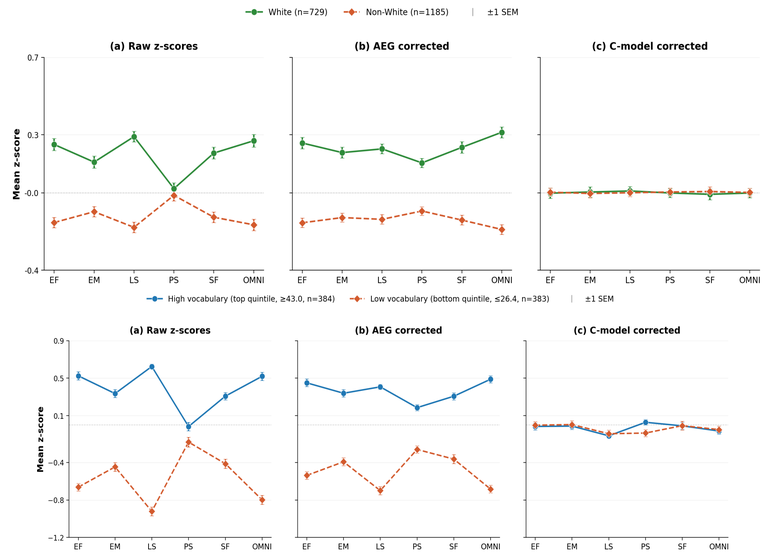
Generalizability. Both methods generalized well across samples: pooled r² values fell within 0.03 of the full-sample R² across all five cognitive domain averages and the OMNI composite for both methods.



Variance explained by the C-model and the standard model. r^2 comparisons showed that the C-model accounted for roughly twice the variance in omnibus and cognitive domain scores as the standard model. The C-model also showed better model fit (lower Akaike Information Criterion values) and improved residual normality..

Racial Differences. Mean domain and omnibus z-scores showed substantial White–non-White differences in raw scores and after standard-model correction. These differences were largely eliminated by the C-model. With the standard model, candidate MCI rates (bottom 7% of omnibus scores) were approximately fivefold higher in Black than in White participants; with the C-model they were ~2-fold higher.

Influence of Premorbid Ability. Mean domain and omnibus z-scores were compared between participants in the highest and lowest vocabulary quintiles. Unregressed and standard-model scores showed performance differences exceeding 1 SD that were largely eliminated by the C-model. With the standard model, candidate MCI rates were approximately elevenfold higher in low-vocabulary than in high-vocabulary participants.; with the C-model they were ~2-fold higher.



SUMMARY

- The C-model explains nearly twice as much score variance as the standard model.
- The C-model reduces racial disparities in the classification of mild cognitive impairment (MCI).
- By integrating corrections for premorbid function, the C-model triples MCI classification rates in participants with high levels of crystallized ability
- C-model solutions generalized across two large cohorts, recruited years apart.
- The C-model corrects scores for premorbid crystallized ability (estimated from vocabulary), obviating the need for post-hoc score corrections.
- Importantly, both standard and C-model scores have clinical utility, in that their agreement adds diagnostic confidence, and their disagreement carries clinically relevant information.

Discussion

..

The scoring pipeline of the CCAB reports standard and C-model scores because the pattern of agreement and disagreement adds diagnostic information and confidence. For example, when standard and C-model classifications agree—both flag abnormal performance, or neither does—diagnostic confidence is increased. When the standard model and C-model disagree, the direction and type of disagreement is diagnostically informative.

Among participants with MCI-level performance (bottom 7%) on either score:

- 47.5% (mean age 55.6) showed abnormalities in both omnibus standard scores (mean = -2.24) and C-model scores (mean = -2.28).
- 26.2% (mean age 51.4) had abnormal omnibus standard scores (mean = -1.81) but C-model scores in the low normal range (mean = -0.94). These participants had small vocabularies (83% in the lowest quartile) and were disproportionately Black (50% vs. 23% of the cohort).
- 26.2% (mean age 50.5) had standard scores in the low normal range (mean = -1.07) but abnormal C-model scores (mean = -1.76). These participants had large vocabularies (only 4% in the lowest vocabulary quartile) and were disproportionately White (52% vs. 38% of the cohort).

References

Woods, D., Pebler, P., Johnson, D. K., Herron, T., Hall, K., Blank, M., Geraci, K., Williams, G., Chok, J., Lwi, S., Curran, B., Schendel, K., Spinelli, M., & Baldo, J. (2024). The California Cognitive Assessment Battery (CCAB). *Frontiers in Human Neuroscience, 18*, 1305529. *https://doi.org/10.3389/fnhum.2023.1305529*

Contact us

drdlwoods@neurobs.com
ccabresearch.com
neurobs.com

Visit us at booth 1417!

Supported by NIA R44AG062076, NIA R44AG080951