Comprehensive Demographic Scoring Models Reduce Bias And Improve Sensitivity Of Cognitive Assessment

Background: Scores on cognitive tests are typically corrected for age, education, and gender (AEG). However, other demographic factors that influence performance are omitted including race, socioeconomic status, comorbidities, vocabulary, and computer familiarity. These omissions reduce scoring precision, disadvantage minorities, and potentially mask cognitive deficits in individuals with high levels of cognitive reserve.

Methods: A diverse normative San Francisco Bay Area sample (N = 1,913; mean age = 53.1, SD = 17.3 years; 56.6% female; 37.2% White, 22.6% Black, 17.9% Asian, 22.3% Other; 26.4% college-educated) completed the California Cognitive Assessment Battery (CCAB), a 2.5-hour telemedically-proctored battery. From 22 CCAB tests we extracted 146 unique performance measures including non-verbal (NV), simple verbal (Speech) , and more complex linguistic (Language) metrics, as well as an omnibus composite measure (OMNI). Vocabulary was measured with a 4-minute adaptive test. The Comprehensive (C-) model scores added vocabulary, age², race, socioeconomic status, computer use, daily medications, and Hispanic background to AEG regressors. For each C-model scoring equation, LASSO regression (500 random 80/20 samples) eliminated insignificant predictors to avoid model overfitting, followed by stable coefficient estimation across 1,000 bootstrap samples.

Results: As shown in Figure 1, the C-model accounted for substantially more variance than the AEG model across all task types. The corresponding reductions in root mean squared error (RMSE) were sufficient to enable sample size reductions in clinical trials of up to 49.4% for the OMNI) metric. As shown in Figure 2, White/Non-White disparities seen in raw and AEG-corrected scores were markedly reduced, with racial bias largely eliminated. Figure 3 shows that large performance disparities (up to 1.5 z-scores) between individuals in the highest and lowest vocabulary quintiles were largely eliminated by the C-model, enhancing sensitivity to early cognitive decline in this population.

Conclusion: Comprehensive scoring models improve test sensitivity, reduce racial bias, and correct for varying levels of cognitive reserve. Their increased statistical power enables sample size reductions in clinical trials targeting diverse aging populations. Evaluating unregressed, AEG, and C-model scores provides a comprehensive perspective on test performance facilitating the early detection of cognitive decline and MCI.

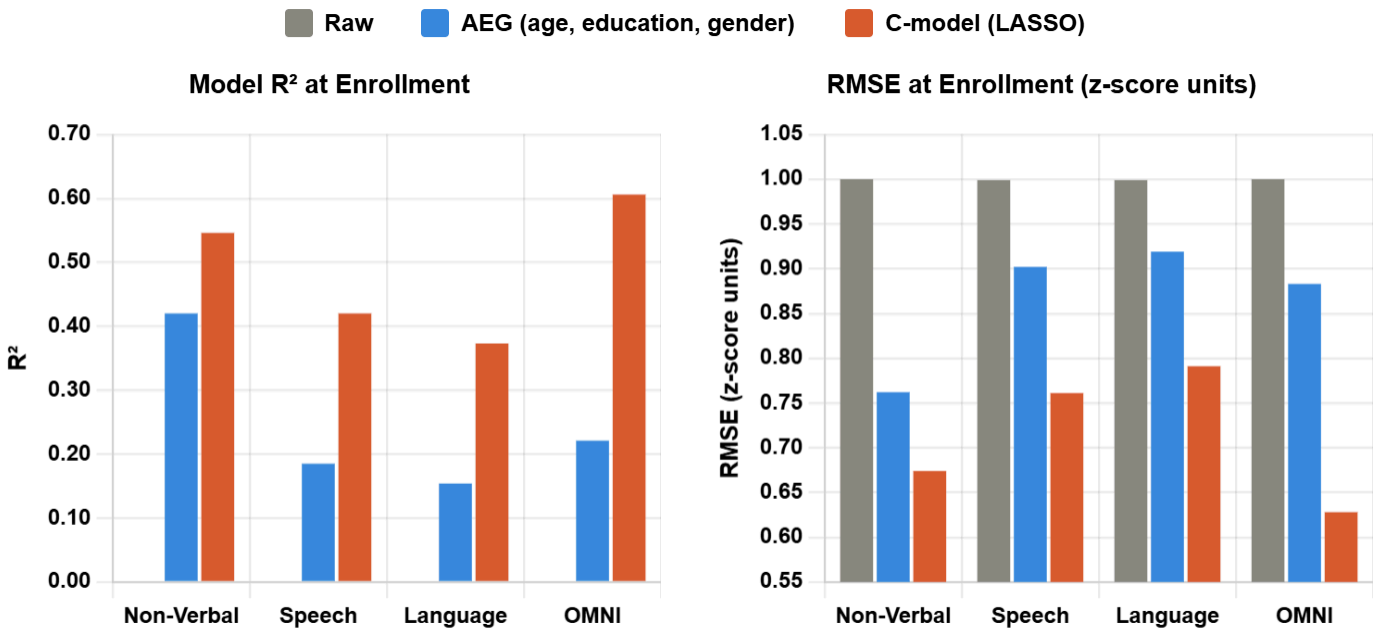

Figure 1. LEFT. Adjusted r^2 values for AEG and C-models for scores averaged over different test types. Right. RMSE values for AEG and C-models.

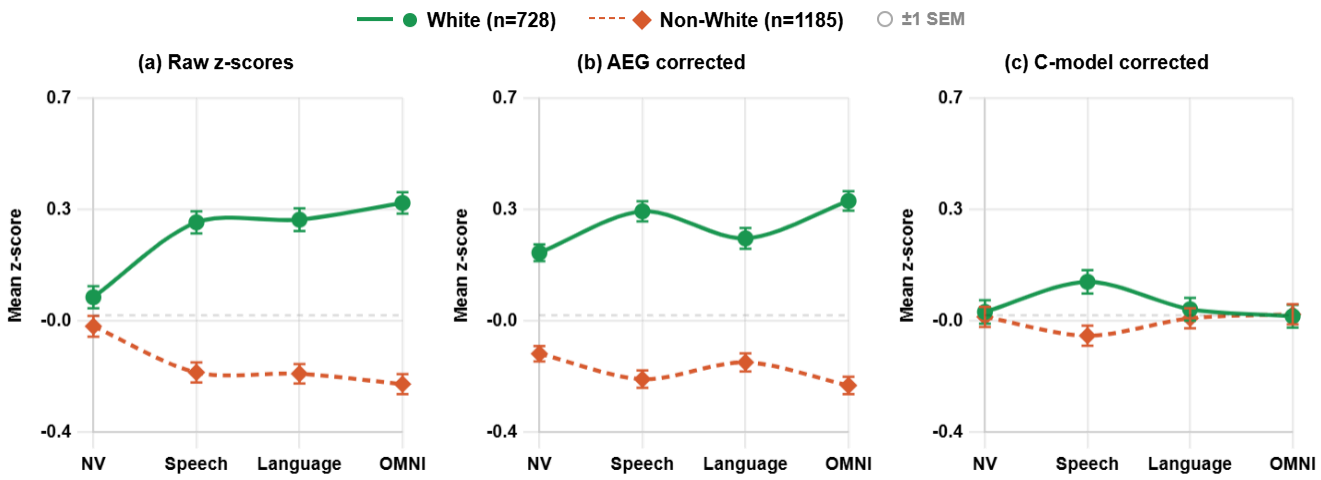

Figure 2. Racial bias in unadjusted (left) and AEG-correct z-scores (center) are largely eliminated with C-model scores (right).

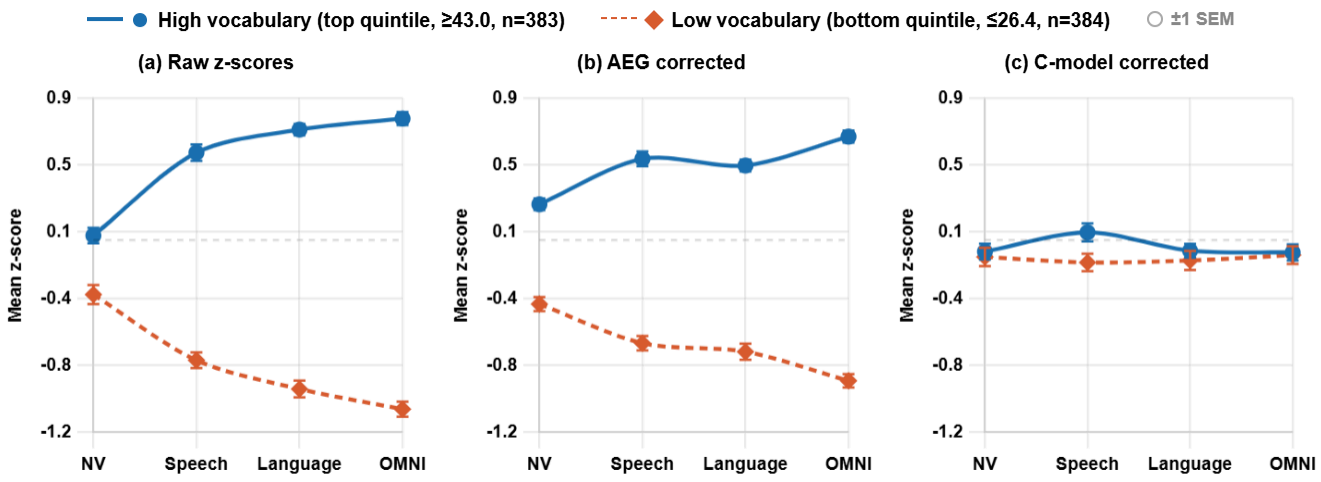

Figure 3. Individuals with high vocabulary scores have much higher raw z-scores (left), and AEG corrected z-scores (center) than individuals with small vocabularies for all test types. These discrepancies are largely eliminated with C-model scoring (right).