## The impact of performance models on MCI diagnosis

**BACKGROUND.** Cognitive-test scores in the bottom 7% of a normative distribution (z-score < -1.50) are an essential component of the diagnosis of Mild Cognitive Impairment (MCI). However, different patients show MCI-compatible performance when evaluated using different models that account for the influences of different demographic variables (e.g., Age only, Age and Gender, Age, Education, and Gender). Most conventional models exclude many factors (e.g., vocabulary, comorbidities, race, functional impairment, and emotional status) that correlate significantly with performance scores. As a result, they reduce performance score precision, inflate racial disparities in MCI classification, and compromise the detection of MCI in high-functioning individuals.

METHOD. 415 older participants (mean age 70.1 years) completed three 90-minute enrollment test sessions using the California Cognitive Assessment Battery (CCAB). Omnibus z-scores were obtained by averaging z-scores from 70 performance measures. Stepwise linear regression was used to identify factors that contributed significantly ($p<0.01$) to the solution, after outliers (typically ~3% of scores) were eliminated. We compared MCI classification using unadjusted performance scores and models with different predictors including (1) Age, (2) Age and Gender; (3) Age, Education, and Gender (AEG), and (4) a Comprehensive model that included 15 demographic factors (see Table 1). All models were tested for heteroskedasticity, multicollinearity, and normality.

RESULTS. Table 1 shows the Pearson correlations with omnibus z-scores of the 15 demographic factors included in the Comprehensive model. The five factors that contributed significantly ($p< 0.01$) and independently to the model were included in the final model solution. Figure 1 shows omnibus z-scores calculated with AEG and Comprehensive models. Table 2 shows the model root mean squared error and the racial bias of MCI classification for each of four models, along with the percentage of patients with identical MCI classifications on the different models. Less than one-third of participants with MCI-level performance on the Comprehensive model showed MCI-level performance on other models.

CONCLUSIONS. MCI classification is strongly influenced by the model used to score performance. Compared to other models, the Comprehensive model accounted for more variance, reduced racial bias in MCI classification, and identified different participants with MCI-level performance.

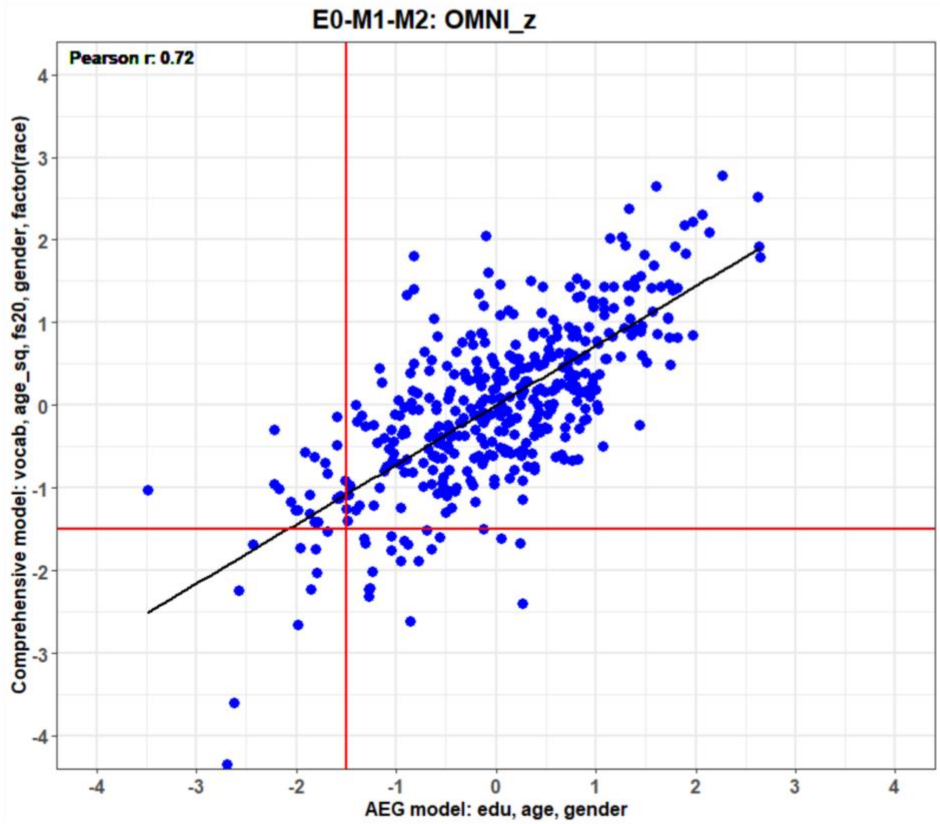


**Figure 1 Omnibus z-scores from enrollment tests calculated with comprehensive and AEG regressors.** Axis labels show significant predictors in order of significance (e.g., vocabulary was the most significant predictor in the Comprehensive model). Vertical and horizontal lines show MCI thresholds for the AEG and comprehensive models respectively: only 22% of participants in the MCI range on one model or the other (bottom left corner) were similarly classified on both models.

| Order | FACTOR | r |
|---|---|---|
| 1 | **Vocabulary*** | **0.58** |
| 2 | **Reading** | **0.36** |
| 3 | **Race*** | **0.35** |
| 4 | **Daily Meds** | **-0.33** |
| 5 | **Computer use** | **0.33** |
| 6 | **Education** | **0.31** |
| 7 | **$Age^2$*** | **-0.29** |
| 8 | **Age** | **-0.28** |
| 9 | **Gender*** | **0.24** |
| 10 | **SES** | **0.19** |
| 11 | **GDS** | **-0.19** |
| 12 | **FS20*** | **-0.17** |
| 13 | **Comorbidities** | **-0.14** |
| 14 | **GAD-7** | **-0.13** |
| 15 | **Hispanic** | **-0.03** |

**Table 1.** Pearson correlations of factors with omnibus performance. Correlations > |0.10| were significant at $p < 0.05$. Gender: Females vs. males, Race: Caucasian vs. non-Caucasian. Reading: hours reading per week. SES: socioeconomic status, GDS: Geriatric Depression Scale, FS20 : functional status, GAD-7: General Anxiety Disorders, Daily Meds: prescription medications taken daily. Asterisks show the five factors that made significant ($p< 0.01$) independent contributions to the Comprehensive model solution.

| Factors | RMSE | Racial bias | None | Age | AG | AEG | Comp |
|---|---|---|---|---|---|---|---|
| **None** | **1.00** | **1.56** | *** | **88%** | **84%** | **84%** | **28%** |
| **Age** | **0.96** | **1.70** | **66%** | *** | **90%** | **88%** | **31%** |
| **Age, Gender** | **0.85** | **1.95** | **67%** | **85%** | *** | **76%** | **27%** |
| **Age, Gender, Education** | **0.79** | **1.64** | **65%** | **77%** | **81%** | *** | **32%** |
| **Comprehensive** | **0.58** | **1.04** | **33%** | **27%** | **27%** | **33%** | *** |
| **Percent MCI** | | | **6.02%** | **7.71%** | **7.95%** | **7.47%** | **7.23%** |

**Table 2.** Root Mean Square Error (RMSE), racial bias (ratio of non-white to white MCI incidence), and the percentage of participants showing concordant MCI classification for different models. Percent MCI: the percentage of z-scores < -1.50 is for each model. AG = Age, Gender. AEG = Age, Education, Gender. Rows above the diagonal show the percentage of participants with MCI-level performance on the simpler model that were similarly classified with the more complex model, while columns show the percent of participants with MCI-level performance on the more complex model that were similarly classified. For example, 84% of participants with uncorrected ("None") z-scores in the MCI range showed AEG z-scores in the MCI range, while 65% of participants with abnormal AEG z-scores showed MCI-level performance in uncorrected scores. Patients with MCI-level performance on the Comprehensive model were largely distinct from those identified with simpler models.