**Supplemental Material**

**Demographic Influences**

Although statistically significant due to large sample size, the largest correlation present was *r* = .22 between education and Osterrieth CF Copy and *r* = -.25 between race and Emory CF Recognition.

**Table 1 – Supplement**

**Pearson Correlations Reflecting Bivariate Demographic Effects on CF scores**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Age* | *Gender* | *Race (B/W)* | *Education (years)* |
| Osterrieth CF Copy | -.02*p*=.667 | .08*p=.*153 | -.20*p*<.001 | .22*p*<.001 |
| Osterrieth CF Immediate | -.20*p*<.001 | -.02*p=.*789 | -.20*p*<.001 | .12*p=.*039 |
| Osterrieth CF Delay | -.22*p*<.001 | .02*p=.*762 | -.20*p*<.001 | .13*p=.*025 |
| Meyers & Meyers CF Recognition | -.05*p=.*355 | .07*p=*.236 | -.19*p*<.001 | .02*p=*.669 |
| Emory CF Copy | .01*p=.*840 | .03*p=.*577 | -.09*p=.*116 | .14*p=.*015 |
| Emory CF Immediate | .-24*p*<.001 | -.01*p=.*804 | -16*p=*.006 | .05 *p=.*377 |
| Emory CF Delay | -.19*p=.*001 | .01*p=.*909 | -.16*p=.*005 | .04*p=.*310 |
| Emory CF Recognition | -.20*p*<.001 | .06*p=.*287 | -.25*p*<.001 | .11*p=.*051 |

**Prediction of Osterrieth Scores from Emory Scores using Regression**

EHBS volunteer data were partitioned into three segments using SAS PROC SURVEYSELECT for simple random sampling with a split of 60% for training, 30% for validation, and 10% for testing. Participants in racial groups other than Black or White were excluded from the regression analyses due to their low representation. Multiple linear regression models were built, and the distribution of the residuals was examined. Predictive models were selected using backward selection with corrected Akaike Criterion (AICC) as a criterion for model selection in the training data, and model performance was validated in the validation data. All the quadratic and pairwise interaction terms were examined. R-squared (*R2*) was used to measure how well the model fits the data. Average squared error (ASE) represents the average of the squared difference between the predicted and observed values. When ASEs for the three subsets are similar, the selected predictive model is reasonable. Root Mean Square Error (RMSE) is the square root of ASE and was used to estimate prediction error. RMSE is reported using the same units as the dependent variable, with RMSE less than 10% considered good. The resultant multiple linear regression equation allows the prediction of traditional Osterrieth values from Emory scores. In all analyses, a *p*-value less than or equal to 0.05 was used to indicate statistical significance. Age was a significant predictor in Pearson correlation with immediate recall but not in the multiple linear regression.

 Only Emory Immediate and Education remained in the selected predictive model for Osterrieth Immediate recall, with the ASEs being similar across the three datasets (11.0, 11.5, and 11.2 for training, validation, and testing, respectively). *R2*was 0.74 and RMSE was 3.3 for the Osterrieth immediate recall which ranged from 1 to 33. For consistency and simplicity, quadratic and interaction terms were removed from the final model for the prediction of Osterrieth delayed recall, and the exclusion resulted in little change in *R2*. The selected effects included Emory Delay, age, race, and education, with ASEs being a little bit higher in the validation (13.1) and testing (13.3) data compared to that in the training set (10.5). *R2*was 0.76 and RMSE was 3.3 for the Osterrieth delayed recall ranging from 0 to 32. About 95% of the observed values would fall within +/- 6.6 points of the predicted values based on RMSE. Clinical application of these formulae should be made with great caution given effect size differences in CF scoring approaches between EHBS and DBS groups. Further caution is needed for individual patients or respondents with less than 11 years of education, individuals with a race other than White or Black, and individuals with a gender other than male or female.

Specific regression values are presented in Tables 2 and 3. Figures 1 and 2 present confidence intervals for the estimated mean (in tan) and individual predicted values (in blue).

**Table 2 - Supplement**

 **Regression Coefficients for Emory Immediate Predicting Osterrieth Immediate Recall**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Independent Variable | Parameter Estimate | Standard Error | *p* value | Lower CI | Upper CI |
| Intercept | -3.56 | 2.07 | 0.09 | -7.64 | 0.52 |
| Emory Immediate Recall | 1.36 | 0.06 | <0.0001 | 1.25 | 1.47 |
| Education, years | 0.26 | 0.12 | 0.0280 | 0.03 | 0.49 |

**Table 3 - Supplement**

**Regression Coefficients for Emory Delay Predicting Osterrieth Delay Recall**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Independent Variable | Parameter Estimate | Standard Error | *p* value | Lower CI | Upper CI |
| Intercept | 3.55 | 3.11 | 0.26 | -2.59 | 9.69 |
| Emory Delayed Recall | 1.285 | 0.06 | <0.0001 | 1.17 | 1.39 |
| Age, years | -0.09 | 0.04 | 0.0201 | -0.16 | -0.01 |
| African American vs Caucasian | -1.16 | 0.61 | 0.06 | -2.37 | 0.05 |
| Education, years | 0.21 | 0.12 | 0.07 | -0.02 | 0.44 |

**Figure 1 - Supplement**

**Predicted Osterrieth Immediate CF Scores from Emory CF Immediate Scores with Confidence Intervals**



**Figure 2 - Supplement**

**Predicted Osterrieth Delay CF Scores from Emory CF Delay Scores with Confidence Intervals**



**Logistic Regression Group Classification**

 All four CF scores from each scoring approach were entered into separate logistic regressions to compare their ability to correctly classify participants into EHBS and DBS groups. The aim of these analyses was to evaluate the different scoring methods as a whole, rather than to create an efficient classification model with the fewest variables. These analyses are not meant to determine specific diagnostic sensitivity since DBS is not a homogeneous diagnostic group, but rather to demonstrate how each CF method distinguishes a patient group with a high prevalence of spatial, executive function, and memory deficits from cognitively healthy controls. The classification differences between the groups in addition to the parametric differences reported in the main report indicate a lack of strict equivalence among the scoring methods and suggest varying neuropsychological sensitivity despite significant correlations between methods.

**Table 4**

**Logistic Regression Classification Tables for Traditional CF Scoring and Emory CF Scoring**

|  |
| --- |
| *Traditional Scoring* |
|  | Predicted EHBS | Predicted DBS | Percentage Correct |
| Observed EHBS | 307 | 8 | 97.5 |
| Observed DBS | 70 | 14 | 16.5 |
| Overall Percentage |  | 80.5 |
|  |
| *Emory Scoring* |
|  | Predicted EHBS | Predicted DBS | Percentage Correct |
| Observed EHBS | 302 | 13 | 95.9 |
| Observed DBS | 57 | 27 | 32.1 |
| Overall Percentage |  | 82.5 |