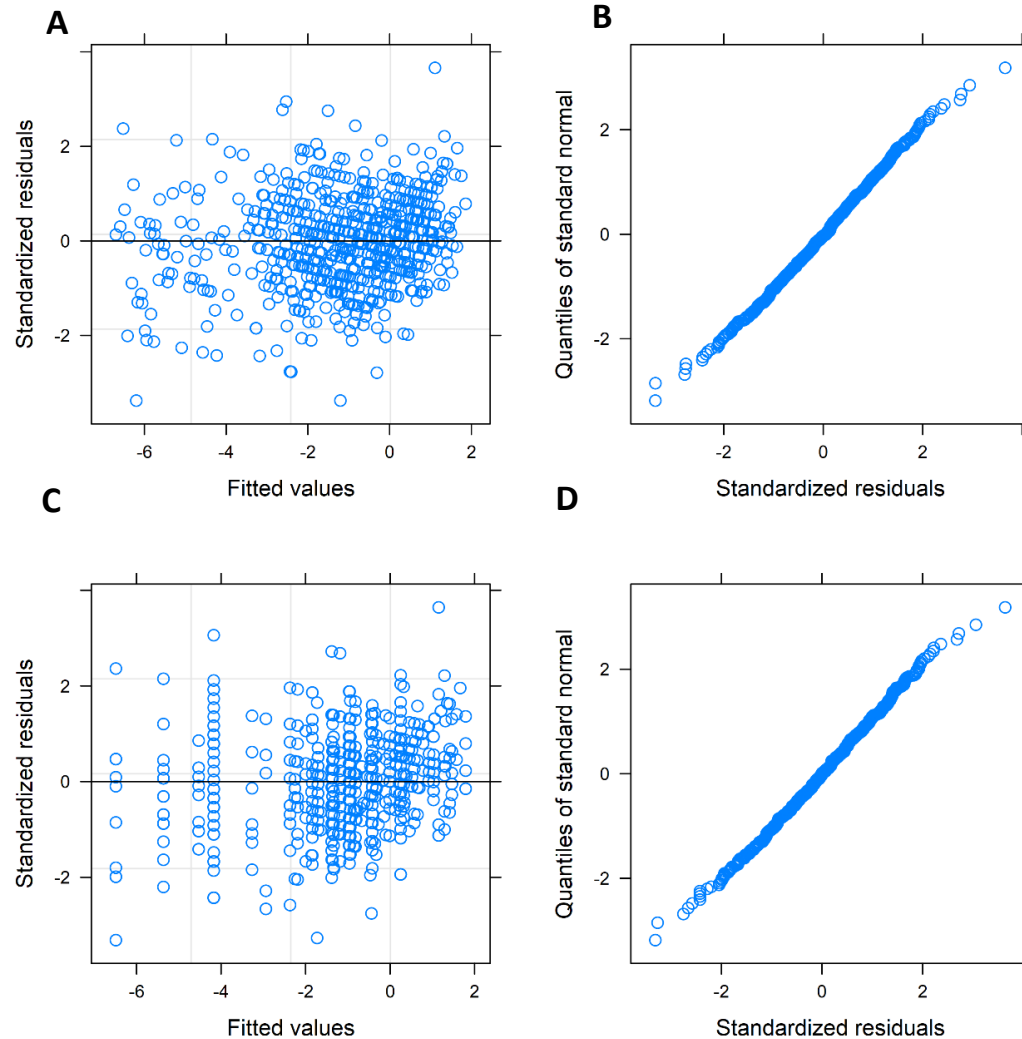


### Mixed-effects models

For calculating limits of agreement:  $(\text{trial 2} - \text{trial 1 score}) = \text{test sitting} + (1 \mid \text{PID})$

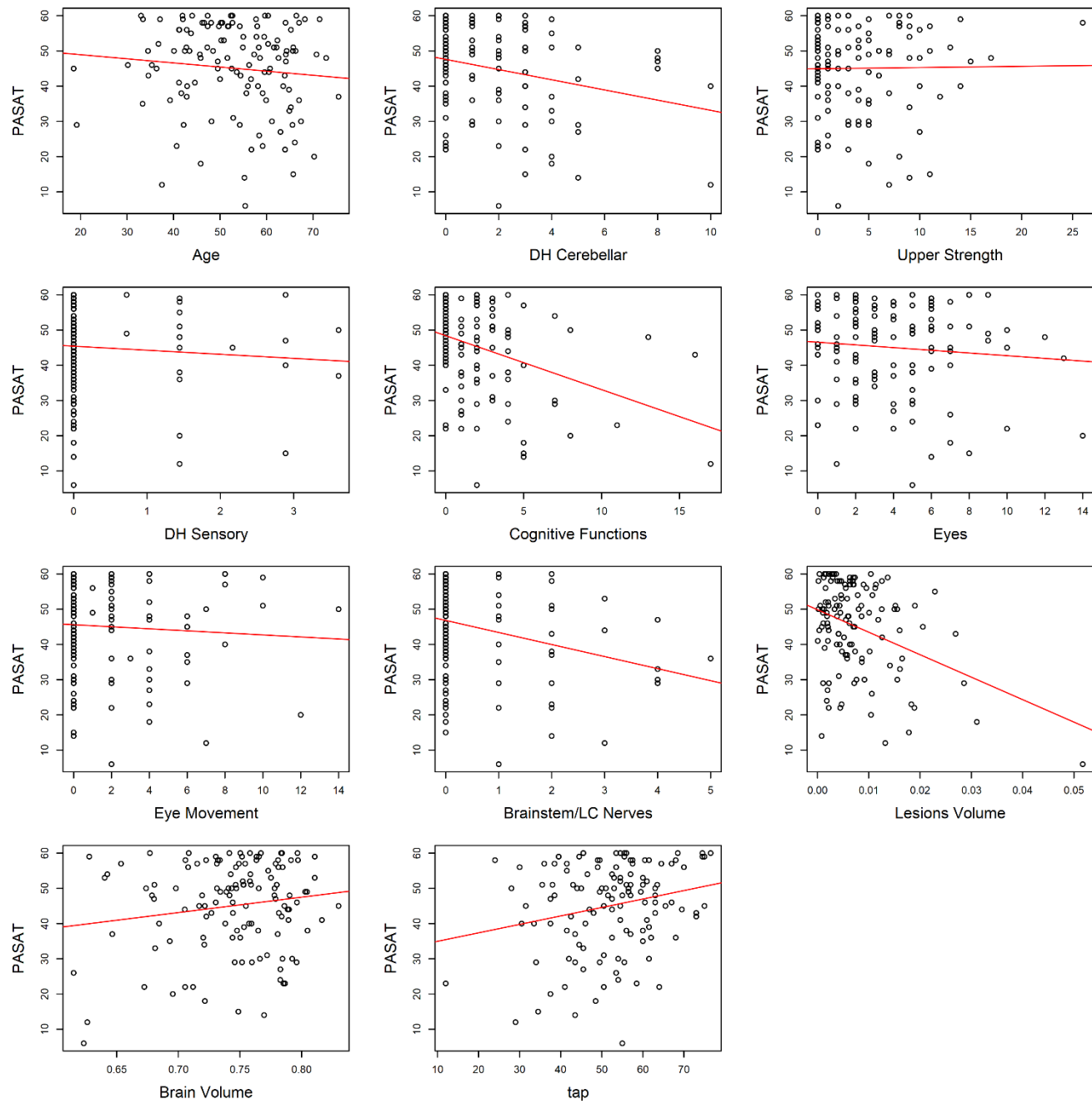
For calculating the cohort mean difference:  $(\text{trial 2} - \text{trial 1 score}) = 1 + (1 \mid \text{PID})$

Supplementary Figure 1



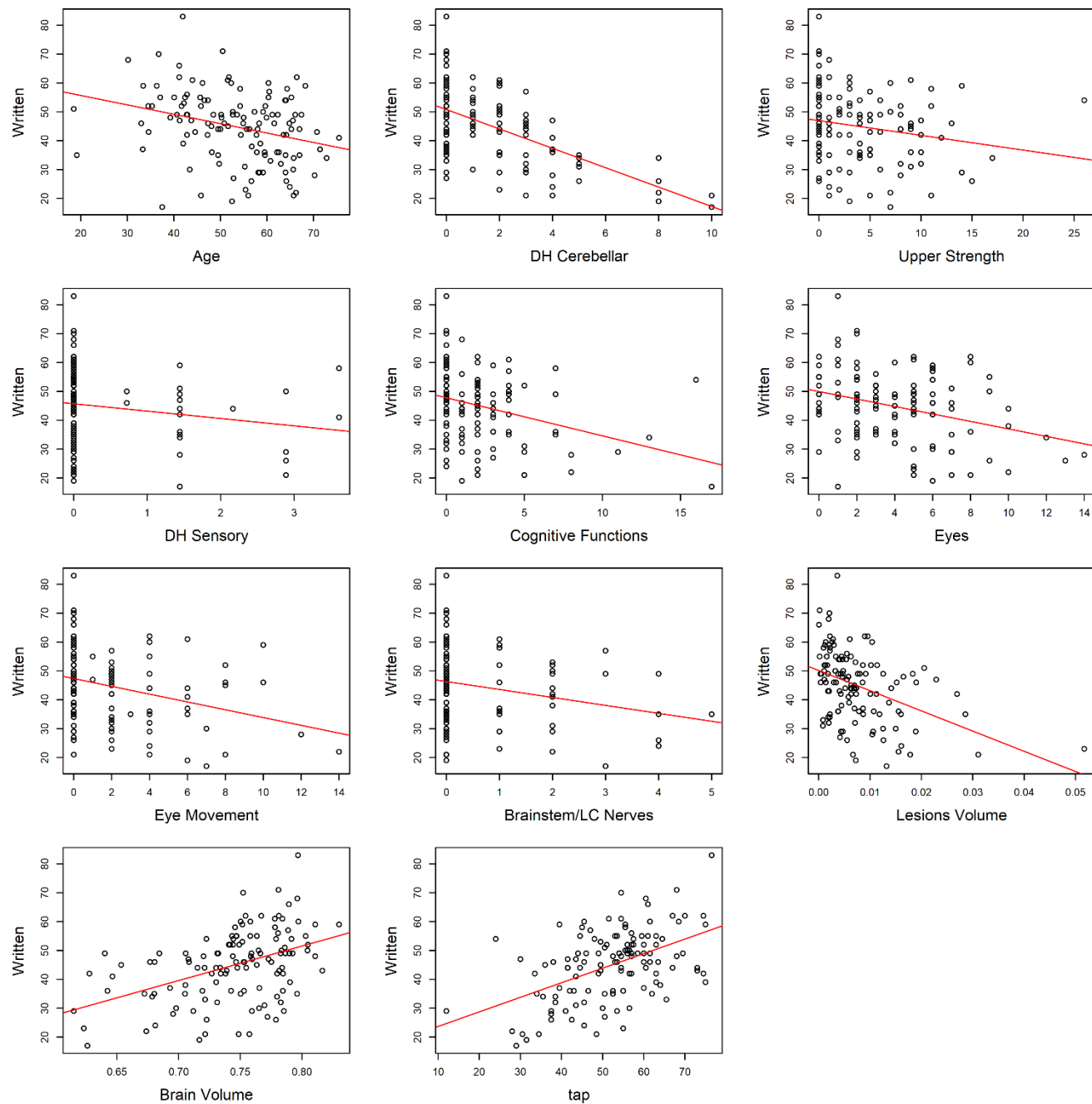
Diagnostic plots for the models that make up the mixed-effects Bland-Altman plot. **(A-B)** Diagnostic plots for the limits of agreement model. **(C-D)** Diagnostic plots for the mean difference model.

**Supplementary Figure 2**



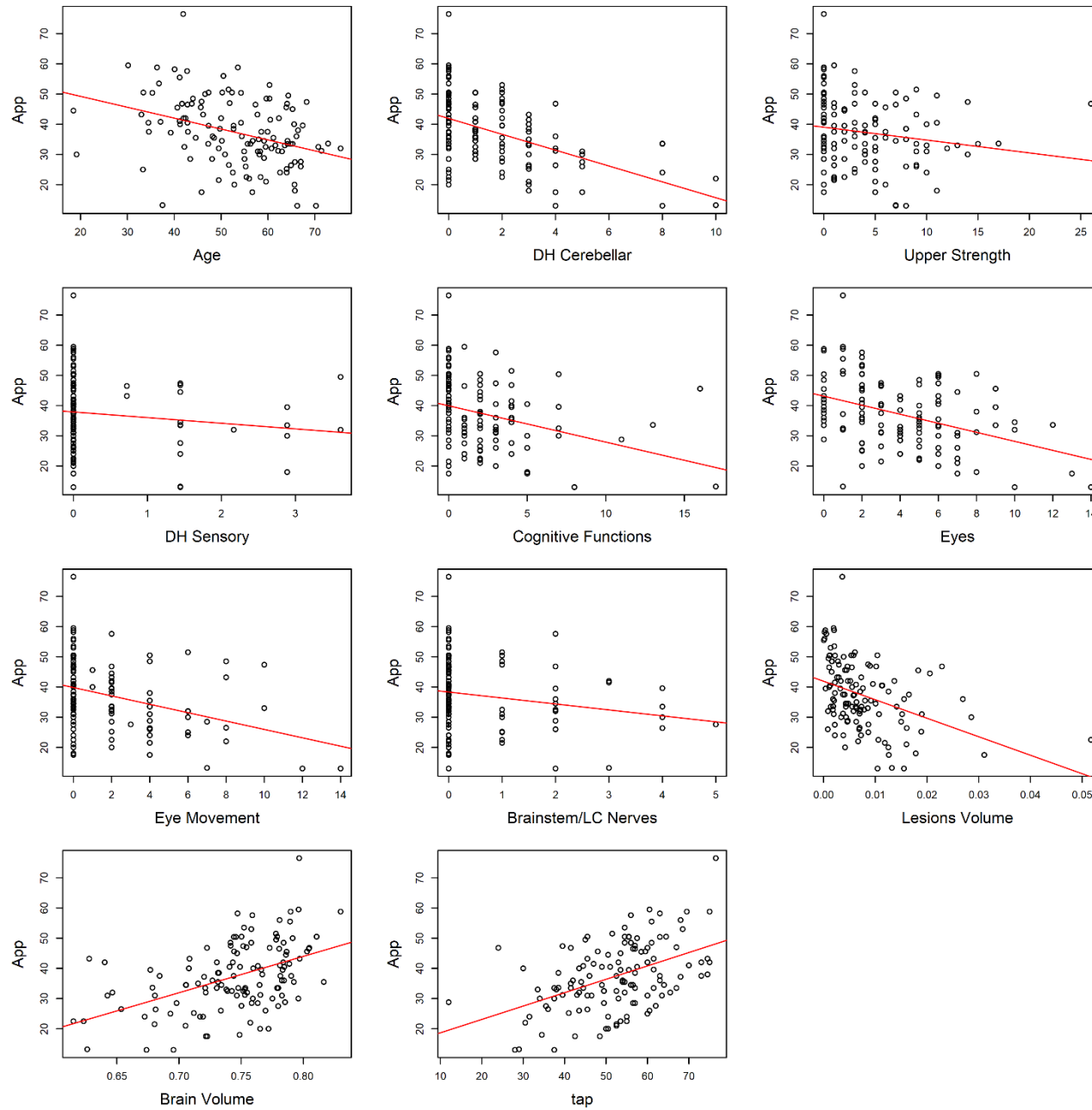
Plots of PASAT scores against clinical features, volumetric MRI, and smartphone dominant hand tapping.

### Supplementary Figure 3



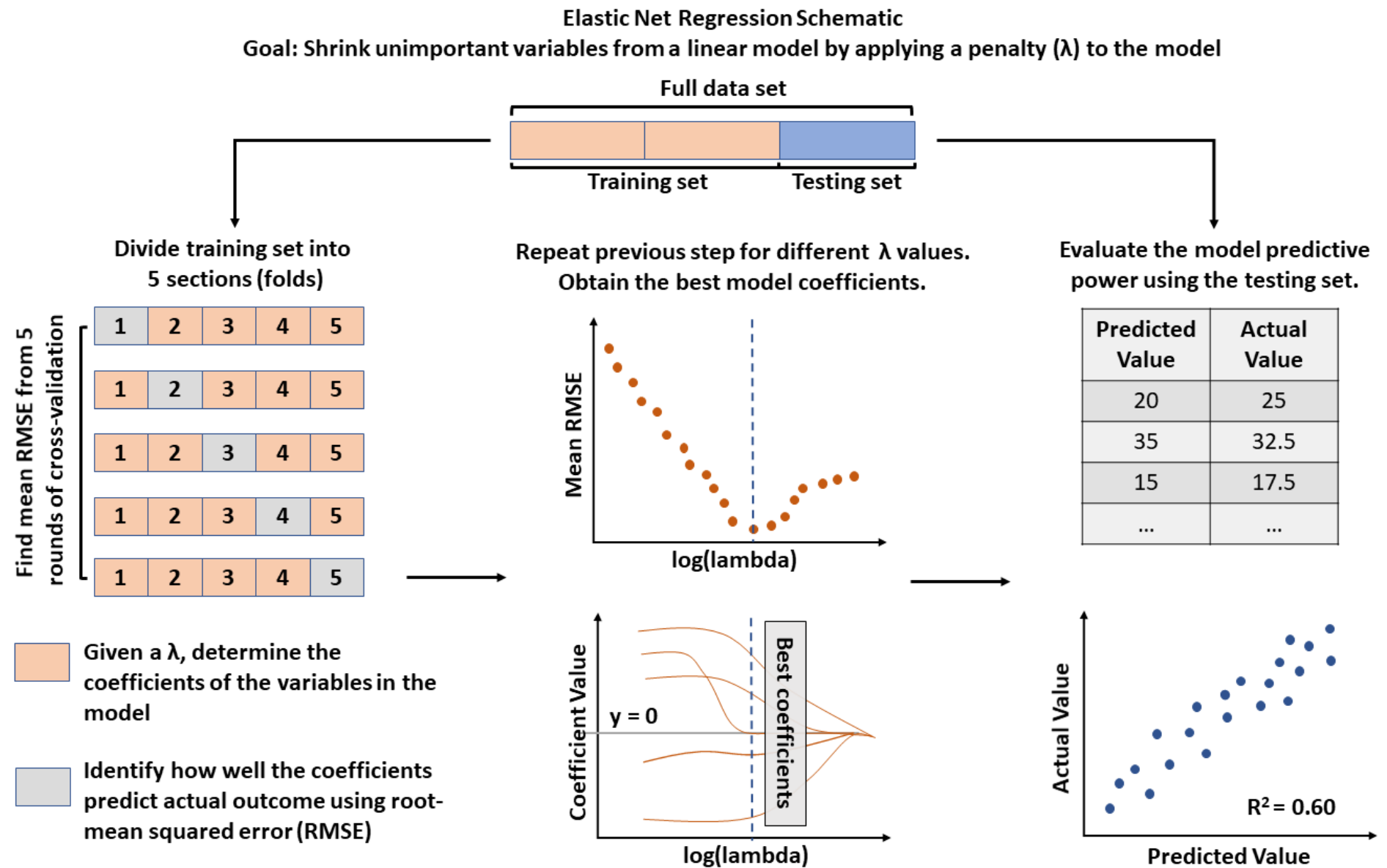
Plots of written SDMT scores against clinical features, volumetric MRI, and smartphone dominant hand tapping.

**Supplementary Figure 4**



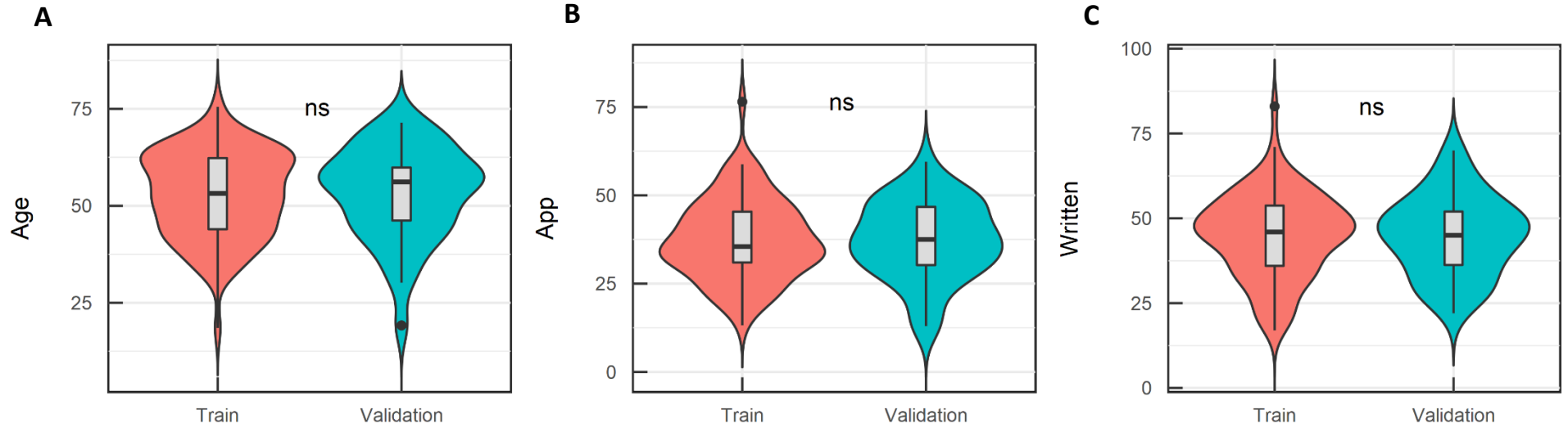
Plots of app SDMT scores against clinical features, volumetric MRI, and smartphone dominant hand tapping.

Supplementary Figure 5



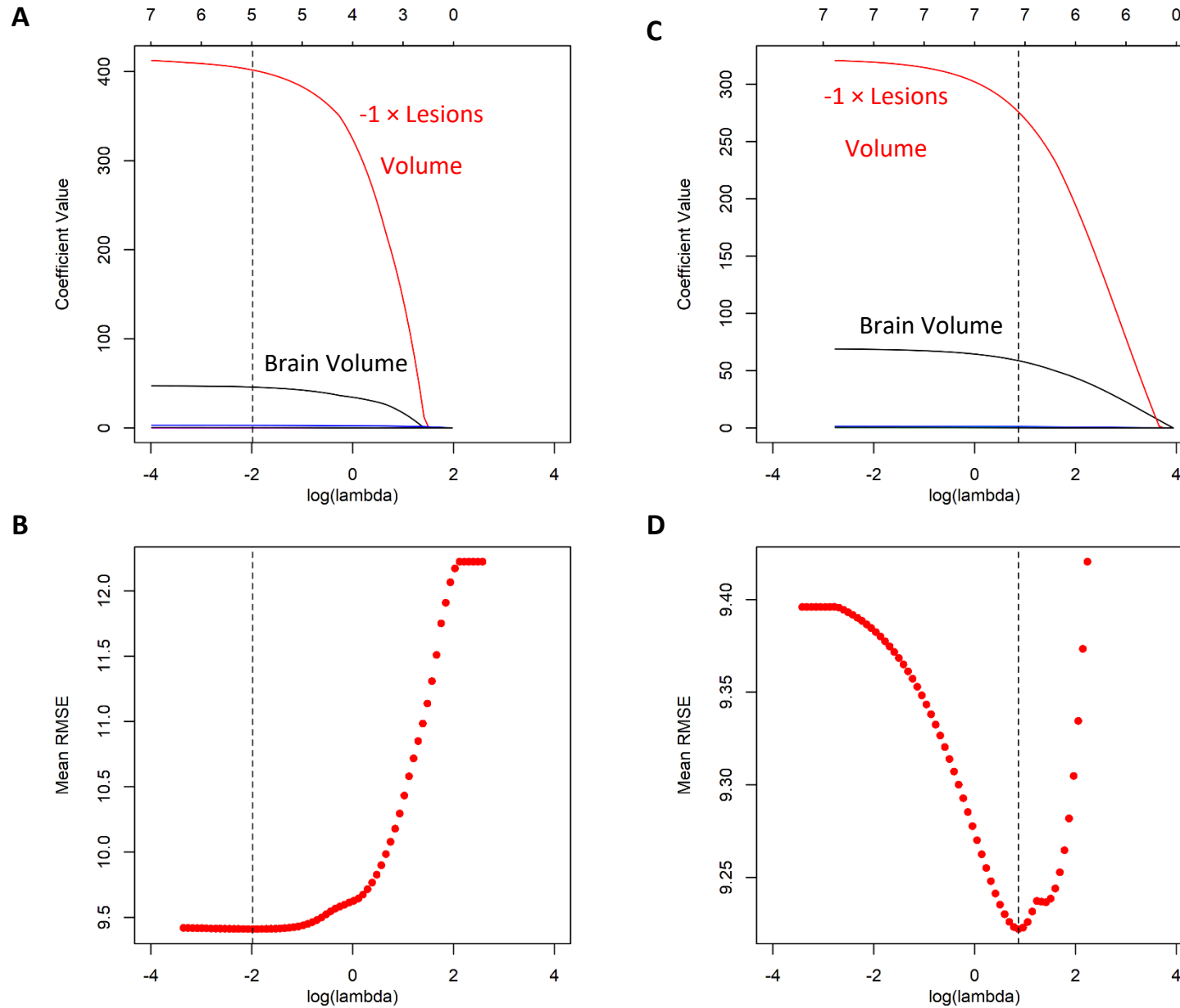
Explanation diagram for how elastic net works.

**Supplementary Figure 6**



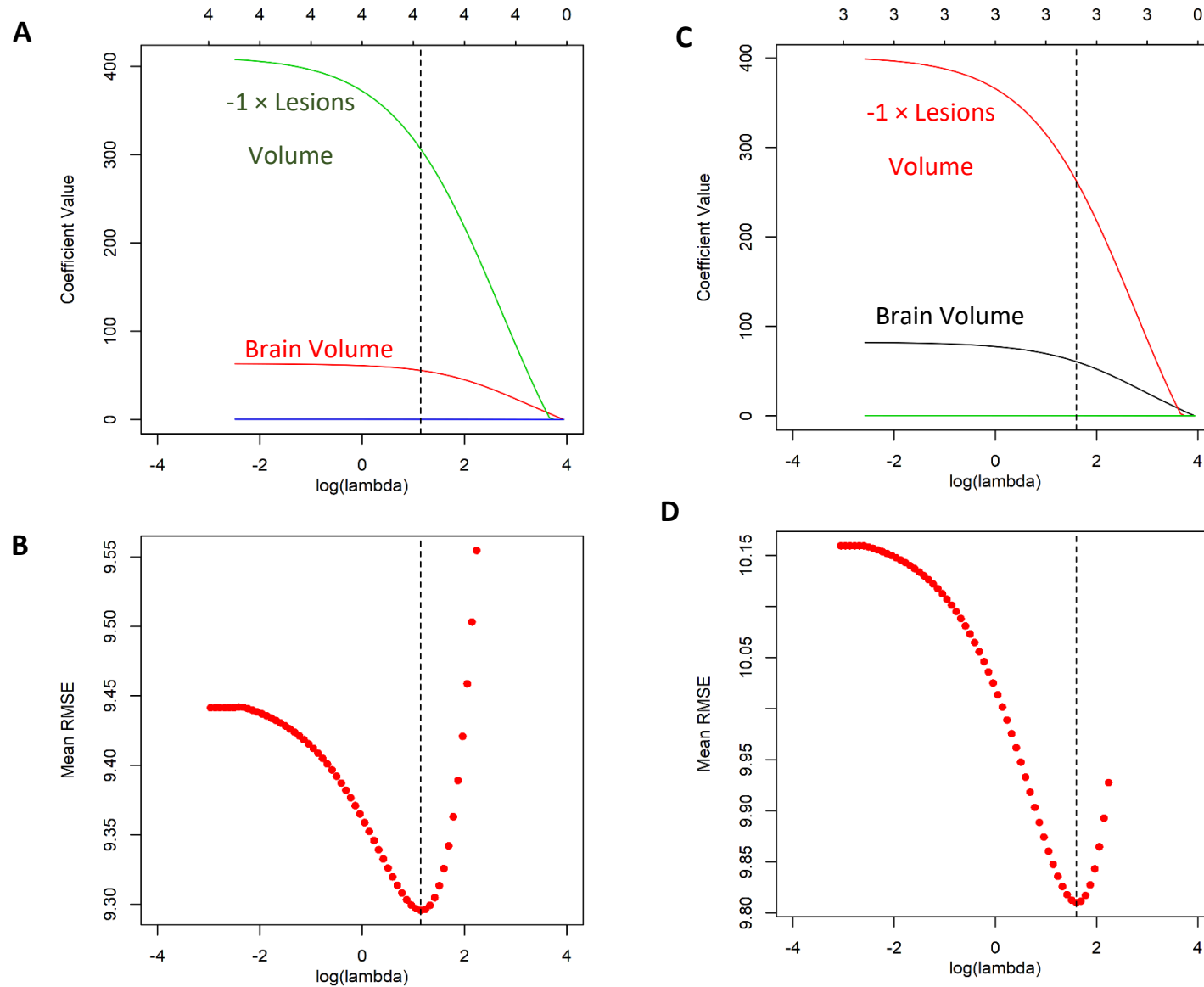
Age and SDMT distributions (app and written) between the cohorts in elastic net regression. The training cohort has 86 individuals and the validation set has 38 individuals. Wilcoxon rank-sum test determined that age and SDMT scores were not significantly different (ns) between these groups.

Supplementary Figure 7



Coefficient path and mean RMSE plots for written and app SDMT elastic net regressions. **(A-B)** Plots for written SDMT. **(C-D)** Plots for app SDMT.

Supplementary Figure 8



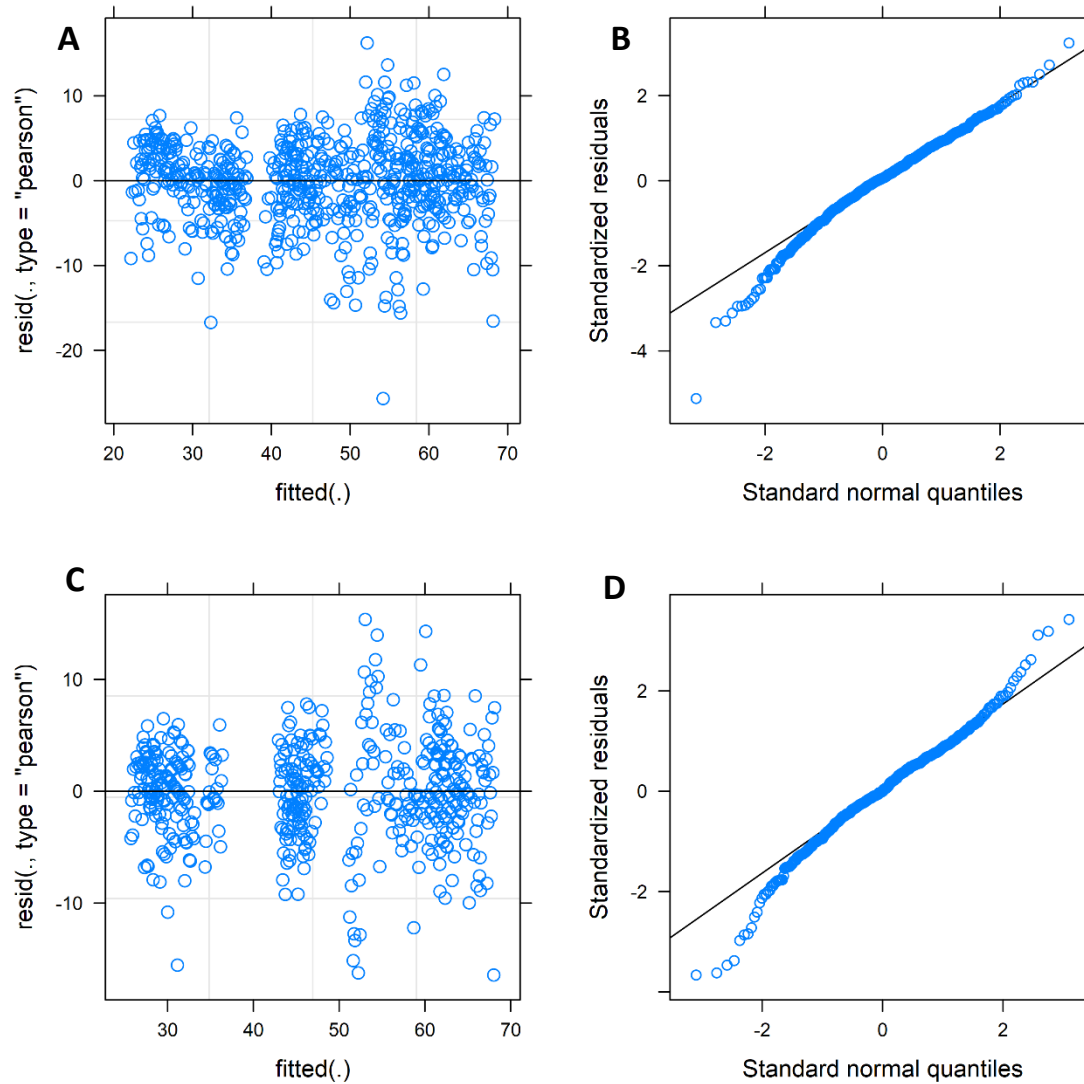
Coefficient path and mean RMSE plots for app SDMT elastic net regressions, with and without tapping score as covariate. **(A-B)** Plots for app SDMT with tapping. **(C-D)** Plots for app SDMT without tapping.



### Mixed-effects model for calculating ICC with sitting number

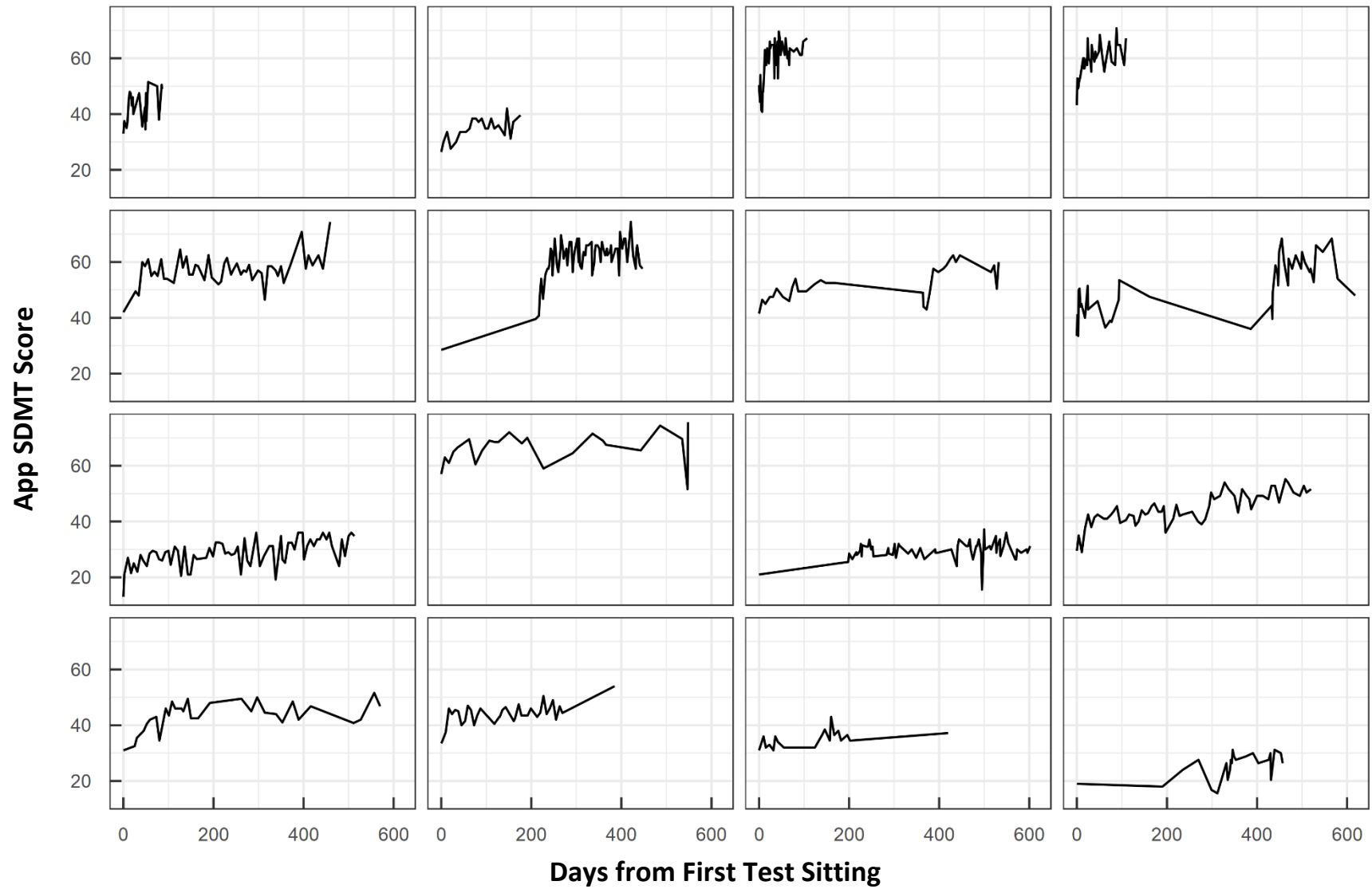
$$\text{app SDMT score} = \text{test sitting} + (1 \mid \text{PID})$$

Supplementary Figure 9



Diagnostic plots for the mixed-effects model that calculate ICC based on scores across number of sittings. **(A-B)** Diagnostic plots for model that includes all longitudinal data. **(C-D)** Diagnostic plots for model that includes only post-learning data.

Supplementary Figure 10

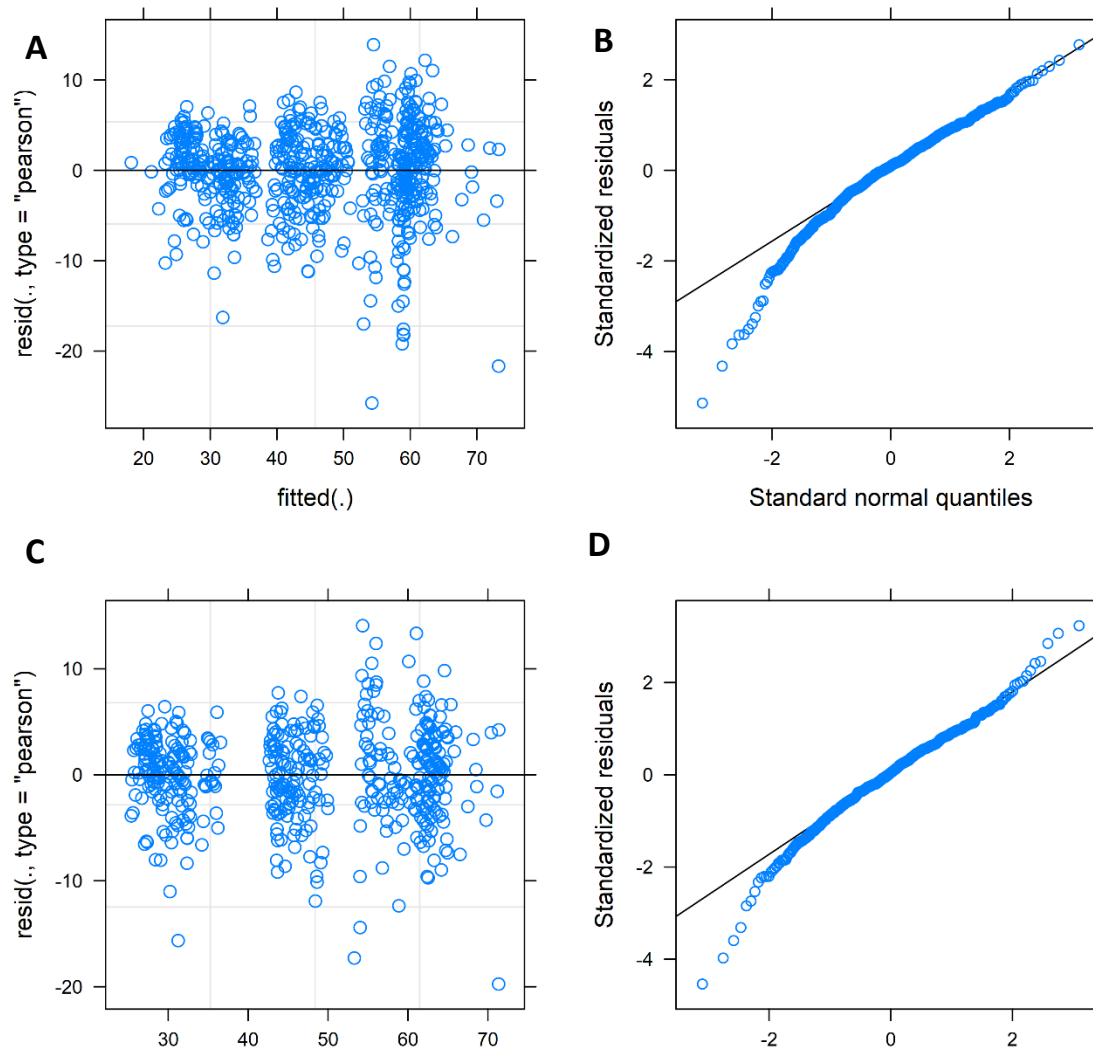


Longitudinal app SDMT scores plotted against days from first test sitting of sitting number.

### Mixed-effects model for calculating ICC days difference

$$\text{app SDMT score} = \text{days since first sitting} + (1 | \text{PID})$$

Supplementary Figure 11



Diagnostic plots for the mixed-effects model that calculate ICC based on scores across different days. **(A-B)** Diagnostic plots for model that includes all longitudinal data. **(C-D)** Diagnostic plots for model that includes only post-learning data (determined using the model in suppl. 9). ICC with all data: 0.88. ICC post-learning: 0.91.

To access all data and analyses scripts, go to:  
<https://github.com/vanessatmorgan/Bielekova-Lab-Code>