# Supplement

**Stability over time of psychiatric ratings and questionnaire results, but not cognitive performance in control individuals**

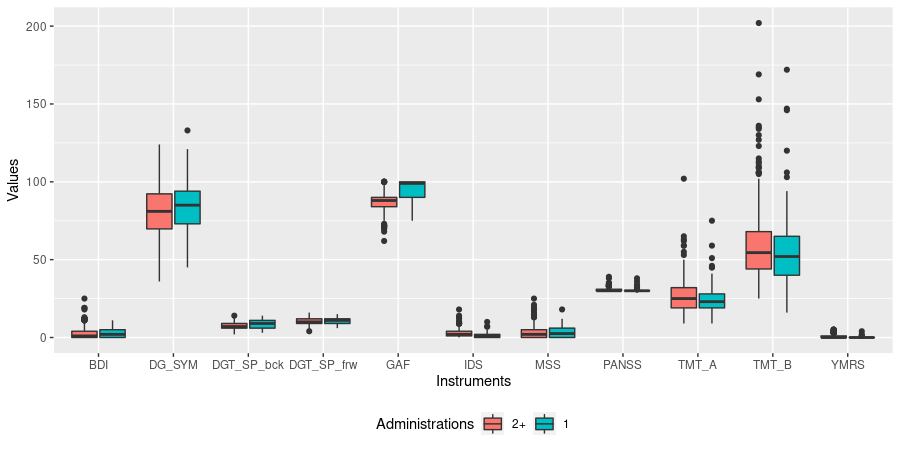
Stahl, Katharina, Adorjan, Kristina, Anderson-Schmidt, Heike, Budde, Monika, Comes, Ashley L., Gade, Katrin, Heilbronner, Maria, Kalman, Janos L., Klöhn-Saghatolislam, Farahnaz, Oraki Kohshour, Mojtaba, Papiol, Sergi, Reich-Erkelenz, Daniela, Schaupp, Sabrina K., Schulte, Eva C., Senner, Fanny, Vogl, Thomas, Wiltfang, Jens, Reininghaus, Eva, Falkai, Peter, Schulze, Thomas G., Bickeböller, Heike, and Heilbronner, Urs

## Dropout analysis

We investigated differences between the participants with one and those with two or more administrations. Since we expect the performance in cognitive tests to improve in later administrations, we only took the data from the first. Note that VLMT was only performed as part of the second administration, which is why we cannot include it. Figure S1 presents the data, stratified according to administrations.

Our main concern was a dropout bias, with particular regard to those participants dropping out whose performance was worse. The results pointed rather towards the other direction, if any difference could be observed at all. We tested for other differences between the groups and discovered that the participants dropping out were also significantly younger than the other participants (p=0.03426). We believe that age is the reason for differences in performance, where younger participants perform better, and we are not overly concerned about systematic differences between the groups. Therefore, we assume missing at random.

Figure S1: Boxplots of performance for instruments at the first administration comparing participants with one and two or more administrations.



## Intra-person variability

We identified people with higher variability with the following procedure:

For each cognitive test and each individual, we computed the difference between administrations and then derived a decisive interval. This is well known for outlier detection, using the quartiles (q0.25 andq0.75) and the interquartile range (IQR).

I = [q0.25 − 1.5⋅IQR; q0.75 + 1.5⋅IQR]

Every individual with at least one particularly large change falling outside of this interval (and thus representing a potential outlier in variability), was labeled “Higher Variability”. This led to the dichotomization of the dataset into the two groups of higher and lower variability:

|  |  |  |
| --- | --- | --- |
| Table S1: Dichotomization of the dataset according to assigned variability group, stratified by sex. | | |
|  | Lower Variability | Higher Variability |
| Female | 121 | 77 |
| Male | 66 | 62 |
| Both Sexes | 187 | 139 |

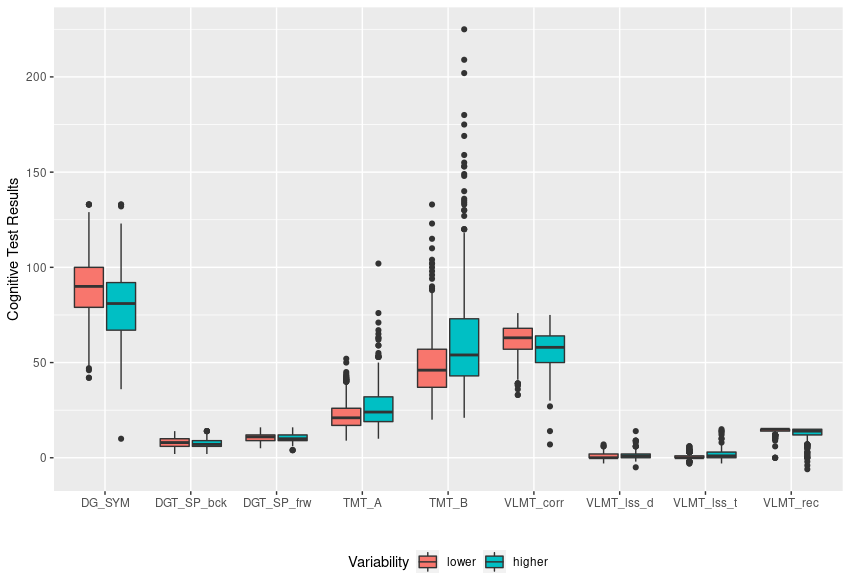
We then tested with a Wilcoxon rank sum test for each variable whether the lower variability group performed better than the higher variability group. Table S2 presents the results.

|  |  |
| --- | --- |
| Table S2: Results of Wilcoxon rank sum tests for difference in performance according to assigned variability group. neg indicates outcome variables for which a lower value is unfavorable. For all other variables, a higher value is unfavorable. | |
| Test | p-Value |
| DG\_SYMneg | 0.0000 |
| DGT\_SP\_bckneg | 0.0266 |
| DGT\_SP\_frwneg | 0.2584 |
| TMT\_A | 0.0000 |
| TMT\_B | 0.0000 |
| VLMT\_corrneg | 0.0000 |
| VLMT\_lss\_d | 0.0002 |
| VLMT\_lss\_t | 0.0000 |
| VLMT\_recneg | 0.0000 |

.

The group with higher variability performed significantly worse in eight out of nine cognitive tests.

The participants with higher variability are significantly older (p-value < 10-6,, median age lower variability: 27 years, higher variability: 41 years)



S2: Boxplots of performance in cognitive tests comparing participants with lower and higher variability. Please refer to Table S2 for the direction of better performance.