Table of contents

# Supplementary material for “Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson’s Disease”

Josef Mana1, Ondrej Bezdicek1, Filip Ruzicka1, Anna Fecikova1, Olga Klempirova1, Tomas Nikolai1, Tereza Uhrova1, Evzen Ruzicka1, Dusan Urgosik2, and Robert Jech1

1Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine and General University Hospital in Prague, Charles University, Czech Republic

2Department of stereotactic and radiation neurosurgery, Na Homolce Hospital, Prague, Czech Republic

# Author Note

Correspondence concerning this article should be addressed to Josef Mana, Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine and General University Hospital in Prague, Charles University, Czech Republic, Email: josef.mana@protonmail.com

# Supplementary material for “Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson’s Disease”

In this supplementary material we present additional information to manuscript *“Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson’s Disease”* including further presentation of the results that was not included in the main text due to space constraints. All procedures described in this supplementary material are accompanied by R code used to implement the steps described herein and Stan code for Bayesian generalized linear mixed models (GLMMs) fitted during this project. The R code and Stan models as well as raw files containing all images and tables are available at <https://github.com/josefmana/dbs_longCOG>. Since the data used for model fitting in our study contain medical records of included patients, they are not publicly available for privacy reasons. Moreover, because the GLMMs reported in this article are exceedingly large for purposes of online storage (> 2 GB each), only the R and Stan codes are included.

# Pre-surgery cross-sectional exploratory factor analysis

## Data pre-processing

For exploratory factor analyses (EFAs) we first log transformed all response time-based tasks (i.e., Trail Making Test and Stroop test), then standardized (i.e., mean-centered and scaled by their in-sample standard deviation) all variables before applying multiple imputations for missing values. EFA was then fitted on each imputed data set via ordinary least squares to find the minimal residual (minres) solution. This procedure was repeated for three up to eight factor solutions.

## Supplementary presentation of results

Supplementary EFA results are presented in **?@tbl-perf** and [Figure 1](#fig-perf) (see below). **?@tbl-perf** presents numerical summary of fit indexes of each three to eight factor solutions across one hundred imputations. Note that Tucker-Lewis Index (TLI) was above the threshold implying good fit (TLI > 0.9) in only three out of four six-factor models, but it was above this threshold in all but three out of one hundred seven-factor models. Similar information is visually presented in [Figure 1](#fig-perf) which depicts density plots of TLI and upper 90% confidence interval boundary of root-mean-square-error approximation (RMSEA) of all models across imputations. This clear improvement in fit of seven- compared to six-factor model, only modest improvement of eight- compared to seven-factor model, and overall theoretical plausibility of factors identified by the seven-factor model led us to retain seven factors for further analyses.

| Model | TLI | RMSEA | RMSEA 90% CI (upper bound) | Total variance accounted for | upper bound RMSEA < 0.08 (%) | TLI > 0.90 (%) |
| --- | --- | --- | --- | --- | --- | --- |
| 3-factor | 0.68 (0.03) | 0.09 (0.00) | 0.11 (0.00) | 0.38 (0.01) | 0 | 0 |
| 4-factor | 0.81 (0.03) | 0.07 (0.01) | 0.09 (0.00) | 0.44 (0.01) | 6 | 0 |
| 5-factor | 0.87 (0.03) | 0.06 (0.01) | 0.08 (0.01) | 0.48 (0.01) | 68 | 16 |
| 6-factor | 0.92 (0.03) | 0.04 (0.01) | 0.07 (0.01) | 0.52 (0.01) | 96 | 74 |
| 7-factor | 0.96 (0.03) | 0.03 (0.01) | 0.06 (0.01) | 0.55 (0.01) | 99 | 97 |
| 8-factor | 0.99 (0.03) | 0.02 (0.01) | 0.05 (0.01) | 0.58 (0.01) | 100 | 100 |
| Values represent mean (SD) or percentages if indicated in brackets. | | | | | | |
| TLI Tucker-Lewis Index. RMSEA root-mean-square-error approximation. CI confidence interval | | | | | | |

**?(caption)**

|  |
| --- |
| Figure 1: Factor analyses fit indexes. Density plots of (A) Tucker-Lewis Index (TLI) and (B) upper boundary of 90% confidence interval (CI) of the root-mean-square-error approximation for three- to eight-factor solutions of factor analysis of pre-surgery cognitive profile. Density plots are taken over one hundred imputed datasets. Vertical lines represent boundaries of good fit according to TLI (i.e., TLI > 0.9) and RMSEA (i.e., RMSEA < 0.08). |

# Longitudinal generalized linear mixed models

## Data pre-processing

To simplify the process of choosing appropriate prior distributions and minimize multicollinearity, all variables were standardized (i.e., mean-centered and scaled by their in-sample standard deviation) before the analyses. The only variable that was not pre-processed this way was time after surgery. This variable was entered into all models in its raw scale (i.e., years after surgery) shifted forward by a median time of pre-surgery assessment (i.e., 0.30 years). Consequently, model intercepts represent estimates of patients’ cognitive performance in Mattis Dementia Rating Scale (DRS-2) at pre-surgery assessment (0.30 years before surgery) and time slopes represent DRS-2 annual post-surgery cognitive decline. Before they were entered into the models, all pre-surgery cognitive factors and test scores were coded such that higher values indicated poorer performance. Parameters associated with these variables (see [Figure 3](#fig-robtest), [Figure 4](#fig-robfact), **?@tbl-posttest**, **?@tbl-postfact** as well as Figure 4 in the main text) thus represent an effect of a (relative) pre-surgery deficit in a corresponding latent cognitive factor or manifest cognitive test score on prediction of pre-surgery DRS-2 (the parameters) and post-surgery annual decline in DRS-2 (the parameters). Negative parameter values imply that a pre-surgery cognitive deficit unfavorably affects the outcome and vice versa for positive parameter values.

## Posterior predictive check

To validate the in-sample fit of our predictive models, we computed models’ “predictions” for each included patient and compared these predictions to observed values (see [Figure 2](#fig-ppc)). Note that since one of the advantages of multilevel modelling is partial pooling, i.e., shrinking parameter estimates towards each other and thus down-weighting the effect of influential outliers to reduce overfitting, the model is neither expected nor required to replicate observed values exactly. Our models show reasonable fit to most patients with clear shrinkage in case of outliers (for instance patient S045 in [Figure 2](#fig-ppc)). Furthermore, while the “test scores” and the “factor scores” model provide similar posterior predictions for our patients, the “test scores” model was evidently more influenced by outlying values to a small degree (for instance patients S023, S107 or S124).

|  |
| --- |
| Figure 2: Posterior predictive checks. Posterior predictions of included patients’ performance according to the predictive generalized linear mixed models (GLMMs) reported in the main text. Lines represent expected (median) performance, shades represent 95% posterior probability intervals (PPIs) of the performance according to the GLMMs, dots represent observed values. |

## Supplementary presentation of results

In **?@tbl-posttest** we present numerical summary of group-level posterior parameters of the “test scores” model while in **?@tbl-postfact** we present numerical summary of group-level posterior parameters of the “factor scores” model which supplement the information presented in Figure 3 in the main text. Since only the interaction terms (i.e., the parameters) comprised empirical estimands for our query (*RQ2*), the remaining parameters were omitted from the main text.

| Parameter | \*b\* | 95% PPI | Pr(\*b\*<0) |
| --- | --- | --- | --- |
| Global intercept (α) | | | |
| Intercept | 140.17 | [139.58, 140.75] | 0.000 |
| Baseline correlates (β) | | | |
| TMT-A | 0.00 | [-0.35, 0.37] | 0.486 |
| TMT-B | -0.26 | [-0.86, 0.16] | 0.884 |
| DS-F | -0.05 | [-0.50, 0.28] | 0.655 |
| DS-B | -0.07 | [-0.54, 0.26] | 0.699 |
| LNS | -0.19 | [-0.77, 0.18] | 0.844 |
| SS-F | -0.05 | [-0.51, 0.29] | 0.654 |
| SS-B | -0.10 | [-0.57, 0.23] | 0.741 |
| TOL | -0.07 | [-0.53, 0.26] | 0.700 |
| PST-D | 0.03 | [-0.31, 0.47] | 0.394 |
| PST-W | 0.00 | [-0.38, 0.41] | 0.499 |
| PST-C | -0.35 | [-0.94, 0.10] | 0.935 |
| COWAT | 0.00 | [-0.36, 0.37] | 0.498 |
| CFT | -0.15 | [-0.72, 0.21] | 0.801 |
| Sim. | -0.19 | [-0.76, 0.16] | 0.848 |
| RAVLT-IR | -0.06 | [-0.51, 0.29] | 0.660 |
| RAVLT-B | -0.35 | [-0.96, 0.10] | 0.936 |
| RAVLT-DR | 0.06 | [-0.28, 0.54] | 0.330 |
| RAVLT-Rec50 | -0.01 | [-0.40, 0.35] | 0.537 |
| RAVLT-Rec15 | -0.11 | [-0.59, 0.21] | 0.769 |
| FP-IR | -0.06 | [-0.54, 0.30] | 0.669 |
| FP-DR | -0.04 | [-0.52, 0.34] | 0.617 |
| STAI-X1 | -0.00 | [-0.36, 0.36] | 0.509 |
| STAI-X2 | 0.01 | [-0.35, 0.38] | 0.472 |
| Time-dependent effects (𝛿) | | | |
| Time | -0.72 | [-0.98, -0.47] | 1.000 |
| TMT-A × Time | -0.08 | [-0.32, 0.11] | 0.810 |
| TMT-B × Time | -0.15 | [-0.45, 0.09] | 0.887 |
| DS-F × Time | 0.12 | [-0.08, 0.35] | 0.130 |
| DS-B × Time | 0.07 | [-0.14, 0.32] | 0.242 |
| LNS × Time | 0.07 | [-0.15, 0.33] | 0.255 |
| SS-F × Time | 0.23 | [-0.06, 0.57] | 0.057 |
| SS-B × Time | -0.11 | [-0.39, 0.11] | 0.838 |
| TOL × Time | -0.05 | [-0.27, 0.14] | 0.696 |
| PST-D × Time | -0.01 | [-0.27, 0.22] | 0.561 |
| PST-W × Time | -0.15 | [-0.44, 0.08] | 0.896 |
| PST-C × Time | -0.09 | [-0.35, 0.12] | 0.811 |
| COWAT × Time | -0.14 | [-0.36, 0.05] | 0.922 |
| CFT × Time | -0.02 | [-0.24, 0.20] | 0.581 |
| Sim. × Time | 0.08 | [-0.13, 0.34] | 0.227 |
| RAVLT-IR × Time | 0.00 | [-0.23, 0.24] | 0.478 |
| RAVLT-B × Time | 0.02 | [-0.17, 0.24] | 0.392 |
| RAVLT-DR × Time | 0.07 | [-0.13, 0.31] | 0.228 |
| RAVLT-Rec50 × Time | -0.03 | [-0.27, 0.17] | 0.632 |
| RAVLT-Rec15 × Time | -0.00 | [-0.22, 0.21] | 0.521 |
| FP-IR × Time | -0.03 | [-0.34, 0.26] | 0.594 |
| FP-DR × Time | -0.05 | [-0.39, 0.22] | 0.684 |
| STAI-X1 × Time | -0.00 | [-0.20, 0.18] | 0.521 |
| STAI-X2 × Time | -0.00 | [-0.21, 0.20] | 0.510 |
| All cognitive predictors were scaled such that negative values mean negative effect of pre-surgery deficit on longitudinal cognitive trajectory. | | | |
| b: parameter value point estimate (posterior media); PPI: posterior probability interval; Pr(b < 0): probability that a parameter is negative, i.e., probability that the predictor has a negative effect on the outcome (this quantity does not apply to Intercept where it cannot be interpreted but it is reported for completeness); ×: statistical interaction term; STAI-X1: State-Trait Anxiety Inventory, the state version; STAI-X2: State-Trait Anxiety Inventory, the trait version; TMT-A: Trail Making Test, part A; TMT-B: Trail Making Test, part B; DS-F: Digit Span forward; DS-B: Digit Span backward; LNS: letter-number sequencing; SS-F: Spatial Span forward; SS-B: Spatial Span backward; TOL: Tower of London task; PST-D: Prague Stroop Test, dot color naming; PST-W: Prague Stroop Test, word color naming; PST-C: Prague Stroop Test, interference condition; COWAT: Controlled Oral Word Association Test; CFT: category fluency test; Sim.: Similarities; RAVLT-IR: Rey Auditory Verbal Learning Test, immediate recall; RAVLT-B: Rey Auditory Verbal Learning Test, recall of the interference set; RAVLT-DR: Rey Auditory Verbal Learning Test, delayed recall; RAVLT-Rec50: Rey Auditory Verbal Learning Test, delayed recognition from 50 items (15 correct answers + 35 distractors); RAVLT-Rec15: Rey Auditory Verbal Learning Test, delayed recognition, number of correctly identified from 15 items; FP-IR: Family Pictures, immediate recall; FP-DR: Family Pictures, delayed recall. | | | |

**?(caption)**

| Parameter | \*b\* | 95% PPI | Pr(\*b\*<0) |
| --- | --- | --- | --- |
| Global intercept (α) | | | |
| Intercept | 140.25 | [139.68, 140.84] | 0.000 |
| Baseline correlates (β) | | | |
| EF/Att. | -0.19 | [-0.78, 0.28] | 0.790 |
| EM | -0.17 | [-0.71, 0.25] | 0.787 |
| VWM | -0.92 | [-1.68, -0.11] | 0.991 |
| VM | -0.35 | [-1.02, 0.19] | 0.889 |
| SS | -0.73 | [-1.39, -0.03] | 0.985 |
| An. | -0.06 | [-0.59, 0.40] | 0.613 |
| SWM | -0.32 | [-1.05, 0.23] | 0.861 |
| Time-dependent effects (𝛿) | | | |
| Time | -0.75 | [-0.99, -0.51] | 1.000 |
| EF/Att. × Time | -0.39 | [-0.63, -0.15] | 0.999 |
| EM × Time | -0.00 | [-0.22, 0.22] | 0.510 |
| VWM × Time | 0.17 | [-0.09, 0.45] | 0.099 |
| VM × Time | -0.17 | [-0.44, 0.10] | 0.888 |
| SS × Time | -0.14 | [-0.47, 0.18] | 0.779 |
| An. × Time | -0.00 | [-0.21, 0.21] | 0.504 |
| SWM × Time | 0.06 | [-0.33, 0.41] | 0.367 |
| All cognitive predictors were scaled such that negative values mean negative effect of pre-surgery deficit on longitudinal cognitive trajectory. | | | |
| b: parameter value point estimate (posterior media); PPI: posterior probability interval; Pr(b < 0): probability that a parameter is negative, i.e., probability that the predictor has a negative effect on the outcome (this quantity does not apply to Intercept where it cannot be interpreted but it is reported for completeness); ×: statistical interaction term; EF/Att.: Executive functions/Attention; EM: Episodic memory; VWM: Verbal working memory; VM: Visuospatial memory; SS: Set shifting; An: Anxiety; SWM: Spatial working memory. | | | |

**?(caption)**

## Robustness checks

To confirm that our results are robust to effects of aging, dopaminergic medication, and depressive symptoms, we carried out a robustness check by fitting parallel “test scores” and “factor scores” models with additional group-level predictors age, levodopa equivalent daily dose (LEDD) and Beck Depression Inventory (BDI-II). Stan code for each of these models with the exact specification is available at https://github.com/josefmana/dbs\_longCOG and is equivalent to the models reported in the main text. For this reason, we do not present their mathematical definitions here in sake of brevity.

Side-to-side comparison of models’ group-level parameters’ posterior summaries is presented in [Figure 3](#fig-robtest) for the “test scores” model and in **?@fig-postfact** for the “factor scores” model. All models arrived at similar posteriors implying our results are robust to effects of aging, dopaminergic medication, and depressive symptoms. Importantly, the empirical estimands relating to *RQ2* (i.e., the time-dependent effects represented by the parameters) are similar across models leading to identical substantive conclusions.

|  |
| --- |
| Figure 3: Posteriors medians and 95% posterior probability intervals (PPIs) of group-level effects from the longitudinal generalized linear mixed model predicting post-surgery cognitive decline by pre-surgery cognitive test scores without (“test scores”) and with adjustment for covariates (“test scores (with covariates)”). All cognitive predictors were scaled such that negative values mean negative effect of pre-surgery deficit on longitudinal cognitive trajectory. See main text for acronyms. |

|  |
| --- |
| Figure 4: Robustness check (the “factor scores” model). Posteriors medians and 95% posterior probability intervals (PPIs) of group-level effects from the longitudinal generalized linear mixed model predicting post-surgery cognitive decline by pre-surgery latent cognitive factor scores without (“factor scores”) and with adjustment for covariates (“factor scores (with covariates)”). All cognitive predictors were scaled such that negative values mean negative effect of pre-surgery deficit on longitudinal cognitive trajectory. See main text for acronyms. |