Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson's Disease

Josef Mana[1], Ondrej Bezdicek[1], Filip Ruzicka[1], Anna Fecikova[1], Olga Klempirova[1], Tomas Nikolai[1], Tereza Uhrova[1], Evzen Ruzicka[1], Dusan Urgosik[2], and Robert Jech [1]

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

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

Author Note

Josef Mana https://orcid.org/0000-0002-7817-3978

Ondrej Bezdicek https://orcid.org/0000-0002-5108-0181

Robert Jech https://orcid.org/0000-0002-9732-8947

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: *Josef Mana*: Conceptualization, Data curation, Investigation, Formal analysis, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing; *Ondrej Bezdicek*: Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing; *Filip Ruzicka*: Investigation, Writing - review & editing; *Anna Fecikova*: Investigation; *Olga Klempirova*: Investigation; *Tomas Nikolai*: Investigation, Writing - review & editing; *Tereza Uhrova*: Investigation; *Evzen Ruzicka*: Conceptualization, Funding acquisition, Investigation, Writing - review & editing; *Dusan Urgosik*: Investigation; *Robert Jech*: Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Supervision, Writing - review & editing

Correspondence concerning this article should be addressed to Ondrej Bezdicek, Email: ondrej.bezdicek@gmail.com

Abstract

Cognitive decline represents a severe non-motor symptom of Parkinson's disease (PD) that can significantly reduce benefits of subthalamic deep brain stimulation (STN DBS). Here, we aimed to identify pre-surgery cognitive profile associated with faster post-surgery cognitive decline in STN DBS treated PD patients to characterize patients who could benefit from more monitoring during treatment course. A retrospective observational study of 126 PD patients treated by STN DBS combined with oral dopaminergic therapy followed for 3.54 years on average (SD = 2.32) with repeated assessments of cognition was conducted. Pre-surgery cognitive profile was obtained via a comprehensive neuropsychological examination. Data were analyzed using exploratory factor analysis for pre-surgery cognitive profile extraction and Bayesian generalized linear mixed models for description of the longitudinal cognitive outcome. Overall, we observed a mild annual cognitive decline of 0.90 points from a total of 144 points in the Mattis Dementia Rating Scale (95% posterior probability interval (PPI) [-1.19, -0.62]). Pre-surgery executive deficit predicted the rate of post-surgery cognitive decline (b = -0.39, 95% PPI [-0.63, -0.15]). The predictive utility of pre-surgery executive deficit resulted from summing small effects of several single test scores. Patients with PD treated with STN DBS experience only mild annual post-surgery cognitive decline. According to our data and models patients with worse long-term cognitive prognosis can be identified via pre-surgery examination of executive functions. Aggregating results from multiple executive tests to estimate cognitive prognosis of PD patients treated with STN DBS is likely superior to examining single test scores.

Keywords: Parkinson's disease, deep brain stimulation, cognition, longitudinal, latent variable analysis

Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson's Disease

Introduction

Bilateral subthalamic nucleus (STN) deep brain stimulation (DBS) is an advanced symptomatic treatment of Parkinson's disease (PD) that can successfully reduce motor symptoms and improve patients' quality of life (Armstrong & Okun, 2020; Bratsos et al., 2018). On the other hand, prior research revealed considerable heterogeneity in cognitive outcomes after STN DBS with a small to moderate post-surgery decline in verbal fluency and equivocal results for other cognitive domains (Combs et al., 2015; Mehanna et al., 2017; Parsons et al., 2006). The ability to predict which patients are likely to develop post-surgery cognitive decline can thus prove useful for patient selection and for guiding post-surgery patient monitoring. In this article, we aim to describe pre-surgery cognitive profile extractable from clinically available neuropsychological evaluation that indicates higher risk of long-term post-surgery cognitive decline in everyday clinical settings.

Studies addressing the task of predicting post-surgery cognitive decline in STN DBS treated PD patients can be broadly divided to two groups, randomized controlled trials (RCT) and long-term observational studies. In a typical RCT, patients are randomized to treatment and placebo groups and outcomes are compared in a full factorial design (evaluating interactions between group and time of assessment as the estimand of interest). Courtesy of their experimental control RCTs allow for causal inference and are well suited for providing guidelines for patient selection. However, even though RCTs are regarded as a gold standard for causal inference, it is ethically unacceptable to deny DBS treatment for PD patients for longer time intervals than necessary. Long-term (i.e., more than three years after surgery) outcomes can thus be best described by observational studies. While observational studies usually do not allow for causal inference and are not well suited for guiding patient selection due to a lack of proper control group and resulting collider bias (Cinelli et al., 2022), they are well suited for description of patients' long-term outcomes. Longitudinal observational studies can serve as a basis for selecting high-risk STN DBS treated patients that would benefit from increased monitoring.

Previous longitudinal observational studies reported that PD patients treated with STN DBS showing pre-surgery deficit in attention and executive functions are at risk of faster post- surgery cognitive decline or developing dementia (Bove et al., 2020; Gruber et al., 2019; Kim et al., 2014; Kishore et al., 2019; Smeding et al., 2009). However, previous studies aimed at identifying any possible pre-surgery predictors of post-surgery cognitive decline accepting high false positive error rates in the process. In this study, we complement prior findings by identifying a sparse solution to the problem of identifying pre-surgery cognitive profile that is predictive of long-term post-surgery cognitive decline in naturalistic clinical settings. In other words, we aim to describe a minimal significant pre-surgery cognitive profile that predicts higher rate of post-surgery cognitive decline in a sample derived from everyday clinical practice.

In a typical observational study aiming to determine pre-surgery risk factors of post- surgery cognitive decline the authors employ the following two-step procedure. In the first step, a series of separate univariate analyses for each potential predictor is conducted to pre-select variables for further analysis. In the second step, predictors that achieved an arbitrary threshold (e.g., $p < 0.05$) are used to predict the cognitive decline in a subsequent multiple regression model (Bove et al., 2020; Gruber et al., 2019; Kim et al., 2014; Smeding et al., 2009). This procedure can lead to false positive error rates that are magnitudes higher than the expected nominal five percent. To overcome this shortcoming, we apply to our data the Bayesian Lasso regression, a method developed for identifying small amount of significant predictors out of a larger pool of possible predictors such as results from a comprehensive neuropsychological battery (Park & Casella, 2008).

Another way to achieve sparsity in prediction of post-surgery cognitive decline is to reduce the number of potential predictors. In the context of neuropsychological assessment this can be accomplished straightforwardly via a latent variable approach such as factor analysis that statistically extracts commonalities across several cognitive tasks. Added benefit of employing such a procedure to pre-surgery predictors is that latent variable approaches can reduce the impact of the task impurity problem – the observation that any cognitive task involves several cognitive functions at once (Burgess, 2014; Whitney & Hinson, 2010).

Overall, in this study we aimed to derive a sparse solution to the task of identifying pre- surgery cognitive profile predictive of long-term post-surgery cognitive decline in STN DBS treated PD patients. In other words, instead of identifying any pre-surgery cognitive variables that can be predictive of post-surgery decline, we aimed to identify only the most likely predictive ones. To this end, we asked the following research questions: *RQ1)* What is the size of expected long-term rate of cognitive decline after STN DBS in PD patients? *RQ2)* What is the pre-surgery cognitive profile that is predictive of long-term post-surgery cognitive decline in STN DBS treated PD? To answer these questions, we analyzed data of retrospectively sampled longitudinally followed STN DBS treated PD patients with a single pre-surgery comprehensive neuropsychological assessment and up to five post-surgery cognitive screening assessments.

Materials and methods

Participants

The data of all patients diagnosed with idiopathic PD following United Kingdom Parkinson's Disease Society Brain Bank Criteria (Hughes et al., 1992) that underwent cognitive evaluation for STN DBS treatment at General University Hospital in Prague between years 2000 and 2020 were retrospectively gathered from clinical records and considered for inclusion in the study. Patients with atypical parkinsonian syndromes, dementia, depression at the time of pre- surgery assessment (according to an independent psychiatric evaluation), recurrent psychotic conditions or a gait disorder despite optimal dopaminergic therapy during pre-surgery assessment were not implanted and were thus not included in the study. Furthermore, only patients who underwent pre-surgery and at least one post-surgery assessment were included. All included patients were treated via continuous bilateral STN DBS in conjunction with dopaminergic therapy. Bilateral STN DBS implantation was performed as previously described (Jech et al., 2006; Jech et al., 2012; Ugosik et al., 2011). All patients provided signed informed consent and the study was approved by the General University Hospital Ethics Committee in Prague, Czech Republic.

Assessments

All patients underwent a comprehensive pre-surgery assessment including neuropsychological and neurological examinations. The patients were followed up post-surgery with similar examination protocol at varying time intervals according to their options. Post- surgery, patients were first contacted one year after the surgery and every two years afterwards. The pre-surgery assessment was performed with the usual dopaminergic therapy (ON medication). In the post-surgery assessment, patients were examined in the ON medication condition and STN DBS ON with optimal stimulation parameters.

Pre-surgery neuropsychological measures

The neuropsychological assessment was arranged analogously to the standard International Parkinson and Movement Disorder Society (MDS) neuropsychological battery at Level II for mild cognitive impairment in Parkinson's disease (PD-MCI) (Bezdicek, Sulc, et al., 2017; Litvan et al., 2012). The battery consisted of 10 tests in 5 cognitive domains: (i) attention: Trail Making Test, part A (TMT-A) (Bezdicek et al., 2012; Bezdicek, Stepankova, et al., 2017; Partington & Leiter, 1949) and dot color naming condition from Prague Stroop Test (PST-D) (Bezdicek, Lukavsky, et al., 2015) for sustained visual attention; (ii) executive functions: Trail Making Test, part B (TMT-B) (Bezdicek et al., 2012; Bezdicek, Stepankova, et al., 2017; Partington & Leiter, 1949) for set shifting, and Tower of London task (TOL) (Michalec et al., 2017; Shallice, 1982) for planning; (iii) language: Similarities (Sim.) from Wechsler Adult Intelligence Scale, third revision (WAIS-III) (Wechsler, 2010) for conceptualization, and category verbal fluency test (CFT, category Animals) (Nikolai et al., 2015) for speeded word production; (iv) working memory: Digit Span backward (DS-B) from WAIS-III (Wechsler, 2010) and Spatial Span backward (SS-B) from Wechsler Memory Scale, third edition (WMS-III) (Wechsler, 2011) for auditory and spatial working memory respectively; and (v) memory: Rey Auditory Verbal Learning Test delayed recall (RAVLT-DR) (Bezdicek et al., 2014; Frydrychová et al., 2018) for explicit verbal learning and memory, and WMS-III Family Pictures delayed recall (FP-DR) for visuo-spatial memory (Wechsler, 2011). Furthermore, we administered the following tests beyond the battery: Prague Stroop Test, naming color of neutral words (PST-W) and interference condition (i.e., naming color of contrasting color words, PST-C) for sensitivity to interference (Bezdicek, Lukavsky, et al., 2015), Controlled Oral Word Association Test (COWAT, letters K + P) (Nikolai et al., 2015) for mental flexibility, and WMS-III letter-number sequencing (LNS) (Wechsler, 2011) for working memory. Finally, anxiety was assessed with the State-Trait Anxiety Inventory for the state (STAI-X1) and trait (STAI-X2) anxiety (Spielberger et al., 1983).

Longitudinal neuropsychological measures

Patients' longitudinal cognitive state was assessed pre-surgery and post-surgery with MDS battery at Level I using Mattis Dementia Rating Scale, second edition (DRS-2) (Bezdicek, Michalec, et al., 2015; Jurica et al., 2001). DRS-2 is a routinely employed cognitive screening measure in PD that has been shown to have acceptable discriminative performance for PD-MCI in Czech population with both sensitivity estimated to be around 0.8 (Bezdicek, Michalec, et al., 2015; Mazancova et al., 2020). Furthermore, subjective depressive symptoms were assessed with Beck Depression Inventory, second edition (BDI-II) (Beck et al., 1996; Ciharova et al., 2020) at each assessment. BDI-II was not used for pre-surgery exclusion due to depression which was instead ascertained by an independent neuropsychiatric evaluation.

Neurological examination

Patients' motor state was assessed with part three of the Movement Disorders Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS III) in medication ON and medication OFF state during the pre-surgery levodopa test. Scores of patients who underwent the older version of the Unified Parkinson's Disease Rating Scale (UPDRS III) were converted to the MDS-UPDRS III scale using the method described by Hentz et al. (2015). The levodopa equivalent daily dose (LEDD) was calculated at each assessment time-point according to Tomlinson et al. (2010).

Statistical analyses

Deriving pre-surgery cognitive profile

Latent cognitive factors were extracted from the data via an exploratory factor analysis (EFA) with varimax rotation using ordinary least squares to find the minimum residual solution (Harman & Jones, 1966). We opted for the orthogonal varimax rotation because: (i) extracting orthogonal factors can be statistically advantageous in later steps of our analysis due to reducing multicollinearity, and (ii) in the framework of PD-MCI, it is considered desirable to describe patients' cognitive profile by factors or tests that are independent of each other (Litvan et al., 2012).

All pre-surgery cognitive tests listed above were entered into EFA as input variables (see Supplementary Materials for the exact processing pipeline). Missing observations were multiply imputed using a parametric bootstrap via the "missMDA" R package to create one hundred imputed data sets We then computed EFA with three up to eight factors via the "psych" R package (Josse & Husson, 2016; R Core Team, 2022; Revelle, 2022) using each imputed data set. Within each imputed data set, factor scores for each patient were calculated using the regression method (Thomson, 1951).

We based the number of extracted factors on a combination of the root-mean-square error approximation (RMSEA), Tucker-Lewis Index (TLI), and consistency of each factor model across imputations. TLI is a measure of a goodness-of-fit such that higher values of TLI imply better fit and values exceeding 0.90 are considered to indicate a good model fit. On the other hand, RMSEA is a measure of badness-of-fit such that lower values imply better fit with values less than 0.08 indicating an adequate model fit (Browne & Cudeck, 1992). A model was considered consistent if the model identified similar factors across imputed data sets.

Describing and predicting post-surgery cognitive decline

Longitudinal data were analyzed using Bayesian generalized linear mixed models (GLMMs). Whereas commonly used analysis of change scores in the pre-test/post-test study design (Combs et al., 2015; Kim et al., 2014; Parsons et al., 2006) confounds true change with measurement error (Singer & Willett, 2003), GLMMs overcome this issue by estimating both group-level (i.e., “fixed effect”) as well as patient-level (i.e., “random effect”) parameters. Furthermore, modelling patient-level effects results in partial pooling of parameter estimates (shifting parameter estimates towards each other), which reduces the influence of outliers and facilitates reliable group-level inference (Gelman et al., 2012; Tuerlinckx et al., 2006).

To describe the rate of post-surgery cognitive decline, we estimated a GLMM with longitudinal DRS-2 performance as an outcome predicted by the time after surgery on the group-level and correlated patient-specific intercepts and slopes on the patient-level. Since the group-level slope of this model represents the expected rate of cognitive decline after STN DBS, it constituted the empirical estimand (Lundberg et al., 2021) for *RQ1.* To evaluate suitability of the linear model we compared it to an equivalent non-linear model that estimated post-surgery cognitive trajectory via tensor product smooths (Wood et al., 2012). Both models were fitted using non-informative improper flat priors to ensure that their parameters are informed primarily by the data.

Two GLMMs were estimated to evaluate predictive utility of pre-surgery cognitive profile. The longitudinal DRS-2 performance was predicted on a group-level by post-surgery time slopes varying by either patients’ pre-surgery cognitive tests’ scores (the “test scores” model) or patients’ pre-surgery latent cognitive factors’ scores extracted from the EFA reported above (the “factor scores” model). Both models further included correlated patient-level intercepts and slopes. To check robustness of our findings we compared the results to estimates of GLMMs that also included group-level effects of age, LEDD and BDI-II (and their interaction with the time after surgery) to adjust for potentially confounding effects of aging, dopaminergic medication, and depressive symptoms.

Since previous long-term studies demonstrated that a subset of PD patients treated with STN DBS can develop dementia which may lead to heavy tails in the data distribution of cognitive test scores, we modelled the data distribution with Student-t instead of Gaussian likelihood. Furthermore, because the outcome DRS-2 has a maximum of 144 points which is achieved by a large proportion of healthy people (Bezdicek, Michalec, et al., 2015), we used the right-censored version of Student-t to account for the ceiling effect. Models' likelihoods had following specification:

$$P(DRS_i = DRS_{max}) = 1 - T(\vartheta, \mu_i, \sigma), for DRS_i \in N_{max}, N_{max} = i : drs_i = drs_{max}$$

$$DRS_i \sim\ t(\vartheta, \mu_i, \sigma), for DRS_i \in N_1, N_1 = i : drs_i < drs_{max}$$

$$\mu_i = \alpha + \delta_{time} time_i + \sum_{j=1}^{m} (\beta_{predictor[j]} predictor_{[j]i} + \delta_{predictor[j]} time_i predictor_{[j]i}) + \bar{\alpha}_{id[i]} + \bar{\delta}_{id[i]} time_i$$

$i = 1 \ldots n$, where n is the total number of assessments across all patients, m is the total number of pre-surgery predictors, DRS_{max} is the maximal attainable score in DRS-2 (i.e., a raw score of 144), *T()* is the Student-t cumulative distribution function, *t()* is the Student-t probability density function, $time_i$ is the time from surgery at assessment i, $predictor_{[j]i}$ is the pre-surgery cognitive score in the predictor (i.e., either a test or latent factor) j of the patient evaluated at assessment i, and the remaining terms denote model parameters. Empirical estimands relating to *RQ2* comprised of the two sets of $\delta_{predictor[j]}$ representing the expected prognostic value of single pre-surgery cognitive tests and latent cognitive factors.

We specified equivalent prior distributions for model parameters of both the "test scores" and the "factor scores" models. We used the Bayesian Lasso priors for all group-level parameters barring the intercept. This prior is the Bayesian equivalent of the Lasso method for performing variable selection and allows for fitting models with a large number of potentially collinear predictors All remaining parameters were given weakly informative priors to ensure that models' estimates fall within the range of measurable values of the outcome (see https://github.com/josefmana/dbs_longCOG for the R and Stan code).

Model description and statistical testing

Effects were described by medians and 95% highest density posterior probability intervals (PPIs) of corresponding model parameters. A 95% PPI can be interpreted such that a given parameter lies within this interval with 95% probability. Models were compared via the expected log pointwise predictive density (ELPD) computed via the leave-one-out cross-validation (LOO-CV) as approximated by the Pareto-smoothed importance sampling (PSIS) (Vehtari et al., 2015). The ELPD difference ($ELPD_{dif}$) and its 95% frequentist confidence interval (CI) were used to decide whether predictive performance of compared models statistically significantly differs (i.e., the 95% CI excludes zero). To identify influential observations, we calculated a Pareto-k diagnostic and looked for observations with Pareto-k > 0.7 which can be considered problematic (Bürkner et al., 2020; Vehtari et al., 2015).

Evaluating false positive error rates

To validate the assumption that our analysis provides lower false positive rates than the commonly used two-step procedure we conducted series of simulations with a data set structure equivalent to that observed in our data. Patients' outcome was generated as a normally distributed random variable with unit standard deviation and mean depending on average annual rate of cognitive decline and patient-specific random deviations. Moreover, for each patient we generated a set of potential predictors including either seven independent variables, twenty-three independent variables or twenty-three covaried variables representing our analysis of the predictive utility of seven latent cognitive factors and twenty-three observed cognitive test scores respectively. Covariance structure in the case of covaried predictors was based on the structure of the battery described above with predictors that represented test measures belonging to the same superordinate task having Pearson's correlation of 0.7 (thus sharing approximately half of the variance) and zero otherwise (see Figure S5 in Supplementary materials). The simulations were set-up such that there was no effect of any predictor on the outcome. Subsequently, we generated one hundred data sets which were then fitted via the two-step procedure and Bayesian Lasso. For each procedure, the number of statistically significant interactions between time and any of the predictors were recorded to estimate the amount of false positive errors these procedures produce under the null hypothesis.

Transparency and openness

All GLMMs were fitted using via Stan's (version 2.21.0) build-in Hamiltonian Monte Carlo sampler accessed via R version 4.2.0 using package "brms" (Bürkner, 2017; R Core Team, 2022; Stan Development Team, 2020). Four parallel chains were run each for 2,500 iterations for each GLMM. The first 500 iterations served as a warm-up and were discarded. Convergence was checked numerically by inspection of the Rs and visually by inspection of trace plots. We used R packages "tidyverse" and "dplyr" for data operations, "tidybayes" for operation with model posteriors, and "DiagrammeR, "ggplot2" and "patchwork" for plotting (Iannone, 2022; Pedersen, 2020; Wickham, 2016). This study's design and its analysis were not pre-registered. The data are not publicly available due to privacy or ethical restrictions. The computer code used in our data analysis as well as synthetic data and replicable code for simulations to estimate false positive error rates can be accessed at https://github.com/josefmana/dbs_longCOG.

Results

Discussion

References

Armstrong, M. J., & Okun, M. S. (2020). Diagnosis and Treatment of Parkinson Disease. *JAMA*, *323*(6), 548. https://doi.org/10.1001/jama.2019.22360

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck depression inventory–II.* American Psychological Association (APA). https://doi.org/10.1037/t00742-000

Bezdicek, O., Lukavsky, J., Stepankova, H., Nikolai, T., Axelrod, B. N., Michalec, J., Rika, E., & Kopecek, M. (2015). The Prague Stroop Test: Normative standards in older Czech adults and discriminative validity for mild cognitive impairment in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *37*(8), 794–807. https://doi.org/10.1080/13803395.2015.1057106

Bezdicek, O., Michalec, J., Nikolai, T., Havránková, P., Roth, J., Jech, R., & Rika, E. (2015). Clinical Validity of the Mattis Dementia Rating Scale in Differentiating Mild Cognitive Impairment in Parkinson's Disease and Normative Data. *Dementia and*

Geriatric Cognitive Disorders, *39*(5-6), 303–311. https://doi.org/10.1159/000375365

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech Version of the Trail Making Test: Normative Data and Clinical Utility. *Archives of Clinical Neuropsychology*, *27*(8), 906–914. https://doi.org/10.1093/arclin/acs084

Bezdicek, O., Stepankova, H., Axelrod, B. N., Nikolai, T., Sulc, Z., Jech, R., Rika, E., & Kopecek, M. (2017). Clinimetric validity of the Trail Making Test Czech version in Parkinson's disease and normative data for older adults. *The Clinical Neuropsychologist*, *31*(sup1), 42–60. https://doi.org/10.1080/13854046.2017.1324045

Bezdicek, O., Stepankova, H., Moták, L., Axelrod, B. N., Woodard, J. L., Preiss, M., Nikolai, T., Rika, E., & Poreh, A. (2014). Czech version of Rey Auditory Verbal Learning test: Normative data. *Aging, Neuropsychology, and Cognition*, *21*(6), 693–721. https://doi.org/10.1080/13825585.2013.865699

Bezdicek, O., Sulc, Z., Nikolai, T., Stepankova, H., Kopecek, M., Jech, R., & Rika, E. (2017). A parsimonious scoring and normative calculator for the Parkinson's disease mild cognitive impairment battery. *The Clinical Neuropsychologist*, *31*(6-7), 1231–1247. https://doi.org/10.1080/13854046.2017.1293161

Bove, F., Fraix, V., Cavallieri, F., Schmitt, E., Lhommée, E., Bichon, A., Meoni, S., Pélissier, P., Kistner, A., Chevrier, E., Ardouin, C., Limousin, P., Krack, P., Benabid, A. L., Chabardès, S., Seigneuret, E., Castrioto, A., & Moro, E. (2020). Dementia and subthalamic deep brain stimulation in Parkinson disease. *Neurology*, *95*(4). https://doi.org/10.1212/wnl.0000000000009822

Bratsos, S. P., Karponis, D., & Saleh, S. N. (2018). Efficacy and Safety of Deep Brain Stimulation in the Treatment of Parkinson's Disease: A Systematic Review and Meta-analysis of Randomized Controlled Trials. *Cureus*. https://doi.org/10.7759/cureus.3474

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit.

Sociological Methods & Research, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Burgess, P. W. (2014). Theory and Methodology in Executive Function Research. In P. Rabbitt (Ed.), *Methodology of Frontal and Executive Function* (pp. 87–121). Psychology Press.

Bürkner, P.-C. (2017). **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., Gabry, J., & Vehtari, A. (2020). Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. *Computational Statistics*, *36*(2), 1243–1261. https://doi.org/10.1007/s00180-020-01045-4

Ciharova, M., Cígler, H., Dostálová, V., ivicová, G., & Bezdicek, O. (2020). Beck depression inventory, second edition, Czech version: demographic correlates, factor structure and comparison with foreign data. *International Journal of Psychiatry in Clinical Practice*, *24*(4), 371–379. https://doi.org/10.1080/13651501.2020.1775854

Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, 004912412210995. https://doi.org/10.1177/00491241221099552

Combs, H. L., Folley, B. S., Berry, D. T. R., Segerstrom, S. C., Han, D. Y., Anderson-Mooney, A. J., Walls, B. D., & Horne, C. van. (2015). Cognition and Depression Following Deep Brain Stimulation of the Subthalamic Nucleus and Globus Pallidus Pars Internus in Parkinson's Disease: A Meta-Analysis. *Neuropsychology Review*, *25*(4), 439–454. https://doi.org/10.1007/s11065-015-9302-0

Frydrychová, Z., Kopeek, M., Bezdicek, O., & Georgi Stepankova, H. (2018). Czech normative study of the Revised Rey Auditory Verbal Learning Test (RAVLT) in older adults. *Ceskoslovenska Psychologie*, *62*(4), 330–349.

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry

About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213

Gruber, D., Calmbach, L., Kühn, A. A., Krause, P., Kopp, U. A., Schneider, G.-H., & Kupsch, A. (2019). Longterm outcome of cognition, affective state, and quality of life following subthalamic deep brain stimulation in Parkinson's disease. *Journal of Neural Transmission*, *126*(3), 309–318. https://doi.org/10.1007/s00702-019-01972-7

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrica*, *31*(3), 351–368.

Hentz, J. G., Mehta, S. H., Shill, H. A., Driver-Dunckley, E., Beach, T. G., & Adler, C. H. (2015). Simplified conversion method for unified Parkinson's disease rating scale motor examinations. *Movement Disorders*, *30*(14), 1967–1970. https://doi.org/10.1002/mds.26435

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*(3), 181–184. https://doi.org/10.1136/jnnp.55.3.181

Iannone, R. (2022). *DiagrammeR: Graph/network visualization.* https://CRAN.R-project.org/package=DiagrammeR

Jech, R., Mueller, K., Urgoík, D., Sieger, T., Holiga, ., Rika, F., Duek, P., Havránková, P., Vymazal, J., & Rika, E. (2012). The Subthalamic Microlesion Story in Parkinson's Disease: Electrode Insertion-Related Motor Improvement with Relative Cortico-Subcortical Hypoactivation in fMRI. *PLoS ONE*, *7*(11), e49056. https://doi.org/10.1371/journal.pone.0049056

Jech, R., Ruzicka, E., Ugosik, D., Serranova, T., Volfova, M., Novakova, O., Roth, J., Dusek, P., & Mecir, P. (2006). Deep brain stimulation of the subthalamic nucleus affects resting EEG and visual evoked potentials in Parkinson's disease. *Clinical Neurophysiology*, *117*(5), 1017–1028. https://doi.org/10.1016/j.clinph.2006.01.009

Josse, J., & Husson, F. (2016). *{missMDA}: A package for handling missing values in multivariate data analysis*. *70*. https://doi.org/10.18637/jss.v070.i01

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). *Dementia rating scale-2 (DRS-2) professional manual*. Psychological Assessment Resources.

Kim, H.-J., Jeon, B. S., Paek, S. H., Lee, K.-M., Kim, J.-Y., Lee, J.-Y., Kim, H. J., Yun, J. Y., Kim, Y. E., Yang, H.-J., & Ehm, G. (2014). Long-term cognitive outcome of bilateral subthalamic deep brain stimulation in Parkinson's disease. *Journal of Neurology*, *261*(6), 1090–1096. https://doi.org/10.1007/s00415-014-7321-z

Kishore, A., Krishnan, S., Pisharady, K., Rajan, R., Sarma, S., & Sarma, P. (2019). Predictors of dementia-free survival after bilateral subthalamic deep brain stimulation for Parkinson's disease. *Neurology India*, *67*(2), 459. https://doi.org/10.4103/0028-3886.258056

Litvan, I., Goldman, J. G., Tröster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., Mollenhauer, B., Adler, C. H., Marder, K., Williams-Gray, C. H., Aarsland, D., Kulisevsky, J., Rodriguez-Oroz, M. C., Burn, D. J., Barker, R. A., & Emre, M. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: *Movement* Disorder Society Task Force guidelines. *Movement Disorders*, *27*(3), 349–356. https://doi.org/10.1002/mds.24893

Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187

Mazancova, A. F., Rika, E., Jech, R., & Bezdicek, O. (2020). Test the Best: Classification Accuracies of Four Cognitive Rating Scales for Parkinson's Disease Mild Cognitive Impairment. *Archives of Clinical Neuropsychology*, *35*(7), 1069–1077. https://doi.org/10.1093/arclin/acaa039

Mehanna, R., Bajwa, J. A., Fernandez, H., & Wagle Shukla, A. A. (2017). Cognitive Impact of Deep Brain Stimulation on Parkinson's Disease Patients. *Parkinson's Disease*, *2017*, 1–15. https://doi.org/10.1155/2017/3085140

Michalec, J., Bezdicek, O., Nikolai, T., Harsa, P., Jech, R., Silhan, P., Hyza, M., Ruzicka, E., & Shallice, T. (2017). A Comparative Study of Tower of London Scoring Systems and Normative Data. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acw111

Nikolai, T., Stepankova, H., Michalec, J., Bezdicek, O., Horáková, K., Marková, H., Ruzicka, E., & Kopecek, M. (2015). Tests of verbal fluency, czech normative study in older patients. *eská a Slovenská Neurologie a Neurochirurgie*, *78/111*(3), 292–299. https://doi.org/10.14735/amcsnn2015292

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337

Parsons, T. D., Rogers, S. A., Braaten, A. J., Woods, S. P., & Tröster, A. I. (2006). Cognitive sequelae of subthalamic nucleus deep brain stimulation in Parkinson's disease: a meta-analysis. *The Lancet Neurology*, *5*(7), 578–588. https://doi.org/10.1016/s1474-4422(06)70475-6

Partington, J. E., & Leiter, R. G. (1949). Partington's Pathways Test. *Psychological Service Center Journal*, *1*, 11–20.

Pedersen, T. L. (2020). *Patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality research*. https://CRAN.R-project.org/package=psych

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199–209. https://doi.org/10.1098/rstb.1982.0082

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Smeding, H. M. M., Speelman, J. D., Huizenga, H. M., Schuurman, P. R., & Schmand, B. (2009). Predictors of cognitive and psychosocial outcome after STN DBS in

Parkinson's Disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *82*(7), 754–760. https://doi.org/10.1136/jnnp.2007.140012

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stan Development Team. (2020). *Stan modeling language users guide and reference manual, version 2.21.0*. http://mc-stan.org/

Thomson, G. (1951). *The factorial analysis of human ability*. University of London Press.

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders*, *25*(15), 2649–2653. https://doi.org/10.1002/mds.23429

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 225–255. https://doi.org/10.1348/000711005x79857

Ugosik, D., Jech, R., Ruzicka, E., Ruzicka, F., Liscák, R., & Vladyka, V. (2011). Deep brain stimulation in movement disorders: a Prague-center experience. *Casopis Lekaru Ceskych*, *150*(4-5), 223–228.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2015). *Pareto smoothed importance sampling*. https://doi.org/10.48550/ARXIV.1507.02646

Wechsler, D. (2010). *Wechsler adult intelligence scale - third revision*. Hogrefe - Testcentrum.

Wechsler, D. (2011). *Wechsler memory scale -third edition abbreviated*. Hogrefe - Testcentrum.

Whitney, P., & Hinson, J. M. (2010). *Measurement of cognition in studies of sleep deprivation* (pp. 37–48). Elsevier. https://doi.org/10.1016/b978-0-444-53702-7.00003-8

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*.

https://ggplot2.tidyverse.org

Wood, S. N., Scheipl, F., & Faraway, J. J. (2012). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360. https://doi.org/10.1007/s11222-012-9314-z