Cognitive predictors of cognitive decline in Parkinson's disease treated by subthalamic deep brain stimulation

DISSERTATION THESIS

Mgr. Josef Mana

Table of contents

6 Conclusions   23

7 Summary   23

8 Souhrn   23

9 References   23

10 List of Publications   23

Supplement   23

1 Introduction

1.1 Parkinson's disease

1.1.1 Motor and nonmotor symptoms

1.1.2 Neuropsychologic evaluation in PD

1.1.3 Theories of cognitive deficit in PD

1.2 Deep brain stimulation

1.2.1 DBS for treatment of motor symptoms

1.2.2 Non-motor side effects of DBS

1.2.3 Cognitive performance in DBS treated patients

1.3 Measuring cognitive functions

1.3.1 Clasical test theory

1.3.2 Item response theory

1.3.3 Causal versus descriptive targets of inference

2 Research Aims

The primary aim of this thesis is to describe pre-surgery cognitive profile of STN DBS treated patients with PD that is prognostic of faster long-term post-surgery rate of cognitive decline. In other words, the thesis ought to indicate which cognitive functions are likely to be impaired already at pre-surgery neuropsychologic assessment in patients that go on to show relatively faster post-surgery cognitive decline. As the answer to this research question is symmetric, this thesis should also indicate which cognitive functions are likely to be relatively unimpaired at pre-surgery assessment in patients that enjoy good long-term post-surgery cognitive performance.

Secondary aims are to enhance our description of pre-surgery cognitive profile prognostic of post-surgery cognitive decline in STN DBS treated PD patients by describing pre-surgery magnetic resonance imaging associated with post-surgery cognitive decline, and to breach the gap between cognitive deficit measured in laboratory settings and its impact on everyday life

by examining how PD patients' performance of daily living change after initiating STN DBS treatment.

To achieve these research goals, we begin by showing that psychologically meaningful differences between patients with and without cognitive impairment can be even in principle observed by studying differences in verbal and non-verbal memory learning curves of PD patients with and without diagnosed PD-MCI (Havlík et al. 2020). Next, we present a longitudinal study that includes data of 126 PD patients repeatedly screened for cognitive deficit while being treated by STN DBS (Mana et al. 2024). Discussion of this study will comprise the majority of the thesis as it directly addresses its primary research aim. Finally, we build upon this study by addressing the secondary aims of exploring pre-surgery structural connectivity profile in MRI of patients who experience more severe post-surgery cognitive decline (Filip et al. 2024), and relating the objective post-surgery cognitive performance assessed in laboratory settings to patients' subjective difficulty in performing cognitively demanding instrumental activities of daily living (Bezdicek et al. 2022).

2.1 Study 1: Learning Curve in Verbal and Non-verbal Memory of Patients with Parkinson's Disease

Declarative memory is one of the cognitive domains that may be impaired even in non-demented PD patients (Bezdicek et al. 2018; Domellöf et al. 2015; Curtis et al. 2019). When taking into account potential mechanisms causing memory deficits in PD which may be either executive (such as retrieval deficit hypothesis) or associative (such as the associative binding hypothesis) (Bezdicek et al. 2019; Brønnick et al. 2011; Chiaravalloti et al. 2014), we can expect there to be differences in immediate recall (i.e., *retention*) and learning over trials (i.e., *slope* or *learning curve*) PD-related deficits. To improve our understanding of these distinctions, Study 1 aims to address following research questions: *RQ1.1)* How do PD patients with and without diagnosis of MCI differ from healthy adults in their visual and verbal memory retention, *RQ1.2)* How do PD patients with and without diagnosis of MCI differ from healthy adults in their visual and verbal memory learning curves? And *RQ1.3)* Do differences in retention and learning curve between PD patients with and without diagnosis of MCI and healthy adults vary according to sensory domain?

2.2 Study 2: Preoperative Cognitive Profile Predictive of Cognitive Decline after Subthalamic Deep Brain Stimulation in Parkinson's Disease

STN DBS in PD patients has been associated with heterogenous cognitive outcomes with prior studies reporting findings ranging from small to moderate post-surgery decline in verbal fluency to 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. The majority of prior studies predicting longitudinal post-surgery cognitive

decline employed pre-surgery/post-surgery design with change scores as their dependent variable (Gruber et al. 2019; Kim et al. 2014) which have a drawback of confounding true change with measurement error (Singer and Willett 2003). Furthermore, the focus on change scores allows researchers to estimate group-level post-surgery changes describing their sample but ignores patient-level variability which is necessary to generalize findings beyond the sample (Yarkoni 2020). In Study 2, we aim to predict cognitive true score changes after STN-DBS leveraging a data set that includes three or more observations in large enough number of patients to estimate both group-level post-surgery cognitive decline to describe our sample as well as patient-level variability to provide predictions for other similar samples. To this end, we asked the following research questions: *RQ2.1)* What is the size of expected long-term rate of cognitive decline after STN DBS in PD patients? *RQ2.2)* What is the pre-surgery cognitive profile that is predictive of long-term post-surgery cognitive decline in STN DBS treated PD patients?

2.3 Study 3: Structural and microstructural predictors of cognitive decline in deep brain stimulation of subthalamic nucleus in Parkinson's disease

Another, increasingly popular modality for predicting post-surgery cognitive decline is pre-surgery profile of anatomy and structural or functional connectivity of patients' brain in MRI (Blume et al. 2017; Costentin et al. 2019; Planche et al. 2018). Consequently, we follow our results of Study 2 up with longitudinal examination of STN DBS treated patients with PD that also underwent diffusion weighted imaging (DWI) and structural MRI before surgery. The research question was *RQ3.1)* What is the pre-surgery profile of structural integrity and microstructural connectivity in MRI that is predictive of long-term post-surgery cognitive decline in STN DBS treated PD patients?

2.4 Study 4: The Instrumental Activities of Daily Living in Parkinson's Disease Patients Treated by Subthalamic Deep Brain Stimulation

Whereas studies 1-3 focus on describing patients' in-laboratory cognitive performance, Study 4 aims to bridge the gap between cognitive deficit detectable by objective cognitive testing and patients' subjective assessment of its impact on everyday living. Activities of daily living (ADL) play a crucial role in this aspect as absence or presence of cognitively caused ADL deficit differentiates between PD-MCI and PDD (Dubois et al. 2007; Litvan et al. 2012). Specifically, cognitively demanding instrumental ADL (IADL) such as following instructions or doing more than one thing at a time may be impaired in PD indicating progression of cognitive decline (Brennan et al. 2016a). In Study 4 we aim to document post-surgery IADL changes of PD patients and estimate causal effect of dopaminergic medication level as a potentially easy-to-intervene-on factor to moderate post-surgery IADL. Following research questons were addressed in this study: *RQ4.1)* What is the size of change in self-reported IADL one year after STN DBS compared to pre-surgery IADL level in PD patients? *RQ4.2)* What is the size

of one year post-surgery self-reported IADL change that can be attributed to time and STN DBS effects rather than other post-surgery factors? *RQ4.3)* How does one year post-surgery self-reported IADL change in response to adjusting levels of dopaminergic medication?

3 Methods

3.1 Learning curve in verbal and non-verbal memory of patients with Parkinson's disease

3.1.1 Participants

The study involved 60 patients with PD recruited from the Movement Disorders Center, Department of Neurology at First Faculty of Medicine and General University Hospital in Prague, and 60 age and sex matched healthy adults recruited for the National Normative Study of Cognitive Determinants of Healthy Aging (Štěpánková et al. 2015). The exclusion criteria were as follows: PDD according to Movement Disorder Society criteria (Dubois et al. 2007; Emre et al. 2007), atypical or secondary parkinsonism, severe or unstable depression, psychotic symptoms (hal- lucinations or delusions) including those caused by medication, anticholinergic medications, and other medical or neurological conditions potentially resulting in cognitive impairment (e.g., history of seizure, stroke, or head trauma). All patients were examined in the "on" motor state. Patients were further divided to patients with normal cognition (PD-NC) and patients with mild cognitive impairment (PD-MCI) according to their performance on test battery described below.

3.1.2 Measures

All participants were examined with the Montreal cognitive assessment (MoCA) screening test for signs of overall cognitive deterioration (Kopecek et al. 2017; Nasreddine et al. 2005). The healthy control group was further examined via a complex test battery (Štěpánková et al. 2015) whereas PD patients underwent a standardized battery for PD-MCI according to the Movement Disorder Society Task Force Level II criteria (Bezdicek, Sulc, et al. 2017; Bezdicek, Nikolai, et al. 2017; Litvan et al. 2012). For the purposes of the current study, only the Czech versions of Brief Visuospatial Memory Test (BVMT-R) (Benedict 1997) and Rey Auditory Verbal Learning Test (RAVLT) (Bezdicek et al. 2014) were analysed.

The BVMT-R is a test of visual and spatial declarative memory consisting of a grid of six figures for the participant to remember and draw after 10 seconds of exposure. The stimulus sheet is presented to the participant three times resulting in three 0-12 scores (maximum two point per figure) representing visuospatial *retention* (the first trial) as well as visuospatial *learning curve* (difference between succesive further trials). The test further includes delayed free recall and delayed recognition trials (Benedict 1997) which were not analysed in this thesis.

The RAVLT is a test of verbal declarative memory consisting of a list of 15 words presented in five consecutive trials always followed by the immediate recall. Consequently, the data consist of five 0-15 scores (one point for each word correctly recalled) representing verbal *retention* (the first trial) as well as verbal *learning curve* (difference between succesive further trials). The test further includes interfering list recall, post-interference recall, delayed free recall, delayed recognition, and delayed recognition with forced-choice (Bezdicek et al. 2014; Frydrychová et al. 2018) which were not analysed in this thesis.

3.1.3 Statistical analyses

RAVLT and BVMT-R data were analyzed using Bayesian generalized linear mixed models (GLMMs) (Tuerlinckx et al. 2006; Gelman and Hill 2006; McElreath 2020). Single trial scores were used as outcomes for separate RAVLT and BVMT-R GLMMs with two level of predictors: (i) natural logarithm of trial order, group (HC, PD-NC and PD-MCI) and their interaction on a group level, and (ii) varying participant-specific intercepts at the participant level. Outcome variables as well as trial order were treated as continuous. To account for expected asymmetry in the data, we used skewed Gaussian measurement error model for both outcome variables. Improper flat priors over reals were set-up for population-level parameters and half student-t priors with 3 degrees of freedom for group-level parameters.

To evaluate the memory profile of PD patients in RAVLT and BVMT-R, we first estimated difference between group-specific marginal means across trials (main effects contrasts). Although difference in these marginal means indicates potential memory deficit in some of the groups, it does not indicate whether the deficit is due to impaired *retention* or impaired *learning curve*. We thus also compared between-group differences in marginal means of the first trial performance (simple effect contrasts) as a measure of *retention*, and between-group differences in marginal trends of the logarithmic trial order parameter (interaction contrasts) as a mesure of *learning curve*. All estimates were described by their 89% highest density posterior intervals (HDPI) and compared via the Probability of Direction (*pd*) as an index of effect existence. Marginal means were also compared via percentage in region of practical equivalence (*ROPE*) as an index of effect significance. *ROPE* was set to an interval ranging from -0.1 to 0.1 of the standard deviation of the outcome variable according to the Czech normative data. The results were interpreted following reporting guidelines for Bayesian analyses as articulated by Makowski et al. (2019).

All GLMMs were fitted using via Stan's (version 2.32.2) build-in Hamiltonian Monte Carlo (HMC) sampler accessed via R software for statistical computing version 4.3.3 using package "brms" (Bürkner 2017; R Core Team 2024; Stan Development Team 2020). Four parallel chains were run each for 2,000 iterations for each GLMM. The first 1,000 iterations served as a warm-up and were discarded. Convergence was checked numerically by inspection of the $\hat{R}s$ and visually by inspection of trace plots. Full analysis code is available at https://github.com/josefmana/pd_learCUR.git.

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

3.2.1 Participants

The study involved 126 patients with idiopathic PD following United Kingdom Parkinson's Disease Society Brain Bank Criteria (Hughes et al. 1992) that underwent surgery for STN DBS treatment at the Movement Disorders Center, Department of Neurology at First Faculty of Medicine and General University Hospital in Prague between years 2000 and 2020 and were repeatedly screened for overall cognitive performance in ensuing years. Exclusion criteria were contingent upon patients being suitable candidates for STN DBS treatment and followed the Core assessment program for surgical interventional therapies in Parkinson's disease (CAPSIT) protocol (Defer et al. 1999), consequently, patients with atypical parkinsonian syndromes, dementia, depression at the time of pre-surgery assessment, recurrent psychotic conditions or a gait disorder despite optimal dopaminergic therapy during pre-surgery assessment were not implanted and were thus not included into the study.

3.2.2 Neropsychological examination

Pre-surgery neuropsychological assessment examined the following cognitive domains: (i) attention via Trail Making Test, part A (TMT-A) (Bezdicek et al. 2012; Bezdicek, Stepankova, et al. 2017; Partington and 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 via Trail Making Test, part B (TMT-B) (Bezdicek et al. 2012; Bezdicek, Stepankova, et al. 2017; Partington and Leiter 1949) for set shifting, Tower of London task (TOL) (Michalec et al. 2017; Shallice 1982) for planning, 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), and Controlled Oral Word Association Test (COWAT, letters K + P) (Nikolai et al. 2015) for mental flexibility; (iii) language via 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 via Digit Span (DS) from WAIS-III (Wechsler 2010) as well as letter-number sequencing (LNS) (Wechsler 2011) and Spatial Span (SS) from Wechsler Memory Scale, third edition (WMS-III) (Wechsler 2011) for auditory and spatial working memory respectively; and (v) memory via Rey Auditory Verbal Learning Test (RAVLT) (Bezdicek et al. 2014; Frydrychová et al. 2018) for explicit verbal learning and memory, and WMS-III Family Pictures (FP) for visuo-spatial memory (Wechsler 2011). Furthermore, anxiety was assessed with the State-Trait Anxiety Inventory for the state (STAI-X1) and trait (STAI-X2) anxiety (Spielberger et al. 1983).

Patients' longitudinal cognitive state was assessed pre-surgery and at several times post-surgery using Mattis Dementia Rating Scale, second edition (DRS-2) (Bezdicek, Michalec, et al. 2015;

Jurica, Leitten, and Mattis 2001). Moreover, subjective depressive symptoms were assessed with Beck Depression Inventory, second edition (BDI-II) (Beck, Steer, and Brown 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. Finally, The levodopa equivalent daily dose (LEDD) was calculated at each assessment time-point according to Tomlinson et al. (2010).

3.2.3 Estimands

Theoretical estimands linked to each research question of this study and their mapping to statistical estimators according to framework of Lundberg, Johnson, and Stewart (2021) are presented in Table 1. Regarding our *RQ2.1*, we aimed to estimate the expected cognitive decline on two levels of generalisation: (i) the current sample and (ii) a population of patients selected for DBS treatment via the CAPSIT-protocol criteria (Defer et al. 1999). Whereas virtually all previous studies examining long-term cognitive changes after STN DBS constraint their conclusions to sample-level estimates as even studies employing GLMMs elected to report fixed-effects only (Boel et al. 2016; Pal et al. 2022), in this study we leveraged the hierarchical structure of GLMMs to provide both sample- and population-level estimates. To allow for this generalisation, we assume exchangeability between patients selected via CAPSIT criteria to the extend that can be quantified by patient-level variance estimated from our sample (see Yarkoni 2020). Empirical estimands were the same unit-specific quantities as those presented in Table 1, conditional on patient being selected for the study (based on geographical and exclusion criteria described above). Importantly, all three estimands are descriptive, not causal.

Table 1: Mapping of research questions to estimands to qunatities to be estimated in the study.

Research question	Estimand (unit specific quantity)	Estimand (population)	Statistical estimator
What is the size of expected long-term rate of cognitive decline after STN DBS in PD patients?	Difference between expected post-surgery cognitive performance and expected cognitive performance k years before	Current sample	$\mu_i = \alpha + \delta_{time}time_i$
		CAPSIT based selected patients	$\mu_i = \alpha + \delta_{time}time_i + \alpha_{id[i]} + \delta_{id[i]}time_i$

Research question	Estimand (unit specific quantity)	Estimand (population)	Statistical estimator
What is the pre-surgery cognitive profile that is predictive of long-term post-surgery cognitive decline in STN DBS treated PD patients?	Difference between expected post-surgery cognitive decline of a patient with fixed level of pre-surgery performance across all cognitive factors and expected post-surgery cognitive decline of patients with performance that is one unit smaller in a single cognitive factor but equal to this patient's performance otherwise	Current sample	$\mu_i = \alpha + \delta_{time}time_i + \sum_j factor_{[j]i}(\beta_{factor[j]} + \delta_{factor[j]}time_i)$

3.2.4 Statistical analyses

Pre-surgery cognitive battery was pre-processed via via an exploratory factor analysis (EFA) with varimax rotation using ordinary least squares to find the minimum residual solution (Harman and Jones 1966). All pre-surgery cognitive tests were entered into EFA as input variables. 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 (R Core Team 2024; Josse and Husson 2016; 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). The number of extracted factors was based 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 and Cudeck 1992). A model was considered consistent if it identified similar factors across imputed data sets.

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. The group-level slope of this model constituted statistical estimate of the sample version of our *RQ2.1* estimand (i.e., the expected annual cognitive decline in the sample). To arrive at statistical estimate of the population version of our *RQ2.1* estimand (i.e., the expected annual cognitive decline in a population of patients selected for surgery using CAPSIT-protocol criteria) we used the model to predict expected post-surgery cognitive decline at one year post-surgery intervals compared to a pre-surgery assessment using both group- and patient-level parameters.

To evaluate predictive utility of pre-surgery cognitive profile, we estimated further two GLMMs. 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. Since DRS-2 scores may include significant outliers, we used Studdent-t instead of Gaussian measurement error model. 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), the right-censored version of Student-t was used to account for the ceiling effect. Estimands relating to *RQ2.2* comprised of the two sets of interaction coefficients 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, most importantly the Bayesian Lasso priors for were used all group-level parameters barring the intercept (Park and Casella 2008).

Estimates were described by full posterior distributions, medians and 95% highest density posterior probability intervals (HDPIs) of corresponding model parameters or predictions as appropriate. When presenting results for the second version of *RQ2.1* estimand, we report medians and 90% equal-tailed posterior probability intervals (ETIs) instead. A 90% ETI can be interpreted such that a given parameter or prediction lies with 5% probability above its upper bound and with 5% probability below its lower bound.

All GLMMs were fitted using via Stan's (version 2.32.2) build-in HMC sampler accessed via R version 4.3.3 using package "brms" (Bürkner 2017; R Core Team 2024; 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 $\hat{R}s$ and visually by inspection of trace plots. R package "tidyverse" was for data operations, "tidybayes" was used for operations with model posteriors, and "ggridges," and "patchwork" were used for plotting (Wickham et al. 2019; Kay 2023; Wickham 2016; Wilke 2024; Pedersen 2020). Full analysis code is available at https://github.com/josefmana/dbs_cogPRED.git.

3.3 Structural and microstructural predictors of cognitive decline in deep brain stimulation of subthalamic nucleus in Parkinson's disease

3.3.1 Participants

The study involved 72 patients with PD diagnosed according to the criteria for clinically established PD defined by the Movement Disorders Society (Postuma et al. 2015) that were indicated for STN DBS. Exclusion criteria were general contraindications to MRI examination (see Study 2), substantial vascular or space occupying brain lesions or a neurological or psychiatric disorder other than PD and its related complications. The examination of cognition (via DRS-2) was performed before the STN DBS implantation and then in the years 1, 3 and 5 after

the surgery with the last available assessment, i.e. with the longest follow-up duration, being used to calculate the DRS-2 change per year ($\Delta\text{DRS-2} = \frac{\text{DRS-2}_{post} - \text{DRS-2}_{pre}}{\text{Years post-surgery}}$). Patients with ΔDRS-2 of –2 or less were labelled as cognitive decline (CD) group, the remaining patients were considered cognitively stable (CS).[1]

Pre-surgery MRI acquisition was performed using a 3T MAGNETOM Skyra scanner (Siemens, Erlangen, Germany). A T1-weighted (T1w) scan was acquired with magnetisation-prepared rapid gradient echo (MPRAGE) sequence, 1.0-mm isotropic resolution, repetition time (TR) = 2,200 ms, inversion time (TI) = 900 ms, echo time (TE) = 2.43 ms, and flip angle (FA) = 8°. The protocol further included DWI with voxel size $2.0 \times 2.0 \times 2.0$ mm^3, TR = 9,000 ms, TE = 94 ms, FA = 90°, single b-value of 1100 s/mm^2, and 30 directions with 5 additional b0 images, acquired with antero-posterior phase encoding direction. Post-surgery T1w scan with MPRAGE sequence, 1.0 mm isotropic resolution, TR = 2,140 ms, TI = 1,100 ms, TE = 3.93 ms, and FA == 15° acquired using a 1.5 T MAGNETOM Avanto scanner (Siemens, Erlangen, Germany) was utilised to estimate the position of the DBS electrode.

3.3.2 MRI data processing

For a full MRI data processing pipeline, see the source article (Filip et al. 2024). Shortly, the goal of MRI processing was to (i) transform the raw T1w images from their native space to standardized Montreal Neurilogical Institute (MNI) space (Grabner et al. 2006) to guide connectivity analyses and estimate subcortical grey matter volumes, (ii) extract anatomical connectivity metrics from DWI images, and (iii) extract cortical thickness estimates. Pre-processing steps followed the minimal preprocessing pipelines for the Human Connectome Project leading to set of standard Connectivity Informatics Technology Initiative (CIFTI) files in grayordinate space (Glasser et al. 2013). This process resulted in extracting following predictor of interest: (i) fractional anisotropy (FA) and mean diffusivity (MD) as proxies of microstructural connectivity, and (ii) cortical thickness and subcortical grey matter (based on 69 subcortical regions of interest) as proxies of macrostructural integrity. Lastly, Lead-DBS software version 2.5.3 (Horn and Kühn 2015; Horn et al. 2019) was utilized to determine the position of DBS leads and active contacts with DISTAL subcortical atlas for STN compart-mentalization (Ewert et al. 2018). The overlap of volume of affected tissue at the time of last recorded cognitive assessment (VAT) and the entire STN as well as its motor, associative, and limbic components separately was calculated, providing four overlap volumes for each side.

3.3.3 Statistical analyses

Outcome data were descibed separately for CS and CD groups. Means of continuous variables were compared between groups using two-sample, two-tailed T-tests whereas frequency tables

---

[1]This choice was based on the reasoning that patient that would scored at maximal 144/144 points before sugery would with 2 points/year decline reach the optimal threshold for PD-MCI according to the Czech normative study (Bezdicek, Michalec, et al. 2015) at three-year post-surgery mark.

of nominal variables were were compared using Fisher's exact test. Differences were considered statistically significant if their q-value was lower than .05 after adjusting for 5% False Discovery Rate (FDR)(Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001). To analyse microstructural and macrostructural correlates of pre-surgery cognitive state and post-surgery cognitive decline, two sets of General Linear Models (GLMs) were fitted with region-specific microstructural (FA and MD) and macrostructural (cortical thickness and subcortical grey matter volume) measures as outcomes, pre-surgery DRS-2 score or group (CD versus CS) as primary predictors, and age, sex and disease duration as covariates. Statistical significance of resulting regression coefficients of primary predictors was decided based on non-parametric analysis as implemented in the Permutation Analysis of Linear Models package with 10,000 permutations and FDR correction over the number of parcels separately for each modality (Winkler et al. 2014). Results were considered significant at adjusted q-value $< .05$ and parcel cluster size equal or above 2 to eliminate singleton cortical parcels.

3.4 The Instrumental Activities of Daily Living in Parkinson's Disease Patients Treated by Subthalamic Deep Brain Stimulation

3.4.1 Participants

The study involved 32 patients with PD diagnosed according to the criteria for clinically established PD defined by the Movement Disorders Society (Postuma et al. 2015) that were indicated for STN DBS with identical exclusion criteria as in Study 2 and Study 3 (except for the MRI-specific exclusion criteria of Study 3). All PD patients were under dopaminergic therapy (i.e., levodopa, dopamine agonist, or a combination of them), and levodopa's equivalent daily dose (LEDD) for each patient was calculated before and after surgery (Tomlinson et al. 2010).

3.4.2 Assessments

Both pre-surgery and post-surgery neuropsychological assessment was performed in accordance with published reccomendations (Kubu 2018) and included cognitive screening via Mattis Dementia Rating Scale, second edition (DRS-2) (Bezdicek, Michalec, et al. 2015; Jurica, Leitten, and Mattis 2001), screening of depressive symptoms via Beck Depression Inventory, second edition (BDI-II) (Beck, Steer, and Brown 1996; Ciharova et al. 2020), and the Penn Parkinson's Daily Activities Questionnaire (PDAQ) as a measure of instrumental activities of daily living (IADL).The PDAQ is a brief self-report tool consisting of fifteen items selected by Item Response Theory (IRT)-based statistics from a larger pool of items asking patients about the level of diffucilties they experience with cognitively demanding IADL on a five point Likert scale ranging from 0 ("cannot do") to 4 ("no difficulty") (Brennan et al. 2016a, 2016b). Finally, during the comprehensive pre-surgery assessment for STN DBS patient selection and post-surgery control assessment, motor function was evaluated via MDS-UPDRS III administered

by a trained movement disorders neurologist, and psychiatric symptoms were evaluated by a neuropsychiatrist with specialisation in movement disorders to assess risky neuropsychiatric complications.

3.4.3 Causal assumptions

The causal assumptions of Study 4 are represented in the form of a directed acyclic graph (DAG) depicted in Figure 1 (panel A). Full description of this graphical model is presented in the source article (Bezdicek et al. 2022). Briefly, the assumptions are that post-surgery responses to PDAQ are determined by their pre-surgery level, time-locked clinical characteristics (DRS-2, BDI-II, LEDD), patient- and item-specific characteristics, and DBS itself which is in turn determined by pre-surgery patient's cognitive, affective and medication profiles, all of which are used by clinicians to decide whether to treat the patient with STN DBS. The double-headed arrow between BDI_{pre} and DBS idicates common cause of these nodes, namely underlying depressive syndrome can both inform the psychiatrist about contraindication to DBS treatment and increase BDI-II score.[2] The only difference between the model presented in Figure 1 and the source paper is that here we added further edges from the patient node to DBS node as well as all clinical characteristics on top of its edge to PDAQ. This change is meant to represent that we assume patient-specific time-invariant characteristics (such as disease type or genetic profile) to affect not only PDAQ responses but also all the other variables in our model. Nonetheless, adding edges from the patient node did not change the adjustment sets needed to answer our research questions in any way compared to the source article.

Importantly, to answer *RQ4.2* and *RQ4.3*, we can use the DAG presented in Figure 1 to use the back-door criterion to derive *adjustment sets*, i.e., the set of covariates that, if conditioned on, allow for interpretation of statistical modelling resuls as causal (Pearl 2009; McElreath 2020; Cinelli, Forney, and Pearl 2022). Applying the back-door criterion, we arrive at adjustment sets presented in panels B and C of Figure 1 for *RQ4.2* and *RQ4.3* respectively. Variables to be adjusted for are represent by squares and the backdoors being closed by this adjustment are depicted as light grey edges in the figure. Although panel C in Figure 1 do not include item node into the adjustment set, item-level parameters were still included into analysis because they represent competing causes and their inclusion thus will not lead to bias while potentially improving statistical efficiency (see Model 8 in Cinelli, Forney, and Pearl 2022).

3.4.4 Statistical analyses

The data were analyzed using a set of GLMMs with responses to each item of PDAQ as an outcome, patient-specific and item-specific varying predictors, and a structure of group-level parameters dependend on research question. For *RQ4.1*, only the time of assessment (pre- vs

---

[2]Note that the decision for exclusion from STN DBS treatment for current depression is not based on BDI-II (which is adminstered by neuropsychologists in our institution), but by an independent neuropsychiatric evaluation at our institution.
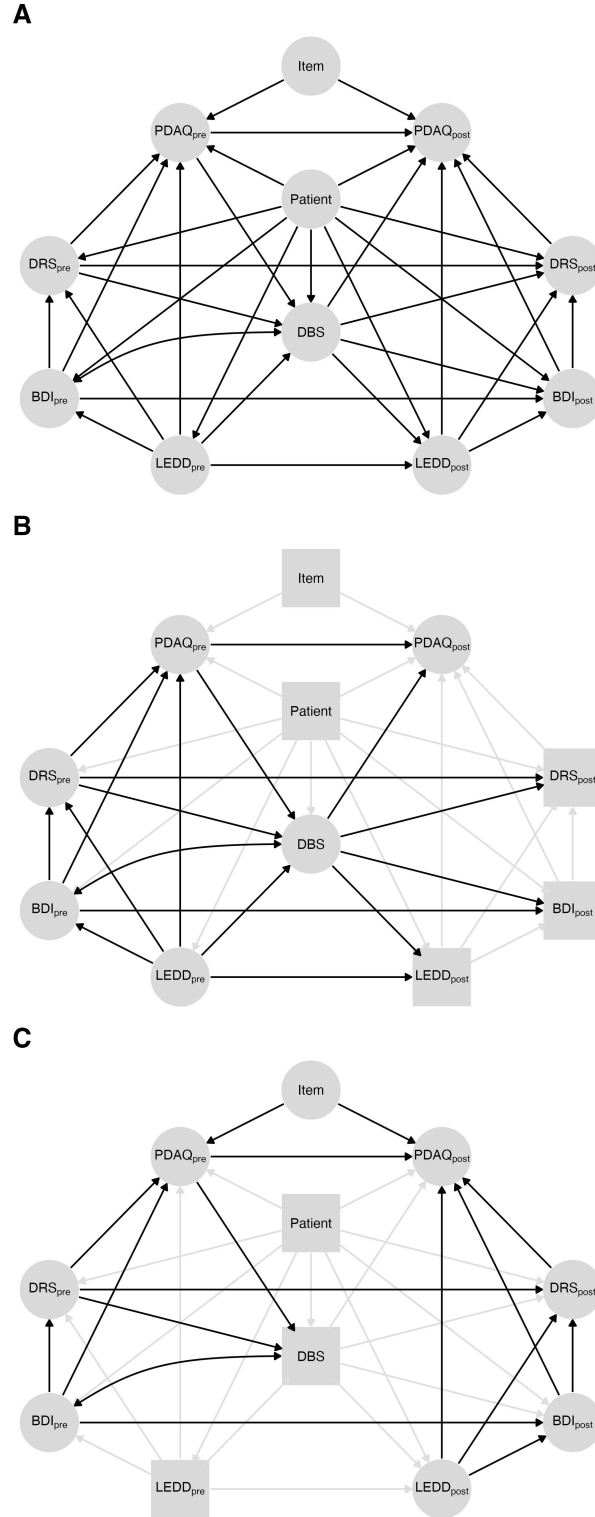
Figure 1: Directed acyclic graph representing causal assumptions of relationships between variables of interest in the current study. The panels represent base causal model (A) as well as model after adjusting for covariates to extract estimate of direct post-surgery change (B), and total effect of post-surgery LEDD on post-surgery IADL.

post-surgery) was used to predict mean group-level responses. Following panel B of Figure 1, the time of assessment as well as DRS-2, BDI-II, LEDD and their interactions with the time of assessment were used to predict group-level responses in model for *RQ4.2*. Finally, following panel C of Figure 1, the time of assessment, LEDD and their interaction were used to predict group-level responses in model for *RQ4.3*. Across all models, the response variable, i.e., the answer to each single PDAQ item on 5-point Likert scale, was modeled using the *ordered-logit* response function (also called *cummulative logit model* or *graded response model* in the literature, Samejima 1995; Liddell and Kruschke 2018; Bürkner and Vuorre 2019; McElreath 2020).

We used Student- t priors with zero mean, a scale of 2.5 and three degrees of freedom for Intercepts and random effects' variance components and regularizing Normal priors with zero mean and standard deviation of 0.5 for the fixed effects. GLMM parameters were described on a logit scale by their medians, 95% highest density posterior probability intervals (PPIs) and the probability of being positive (i.e., the probability that a predictor has a positive effect on IADL).

All GLMMs were fitted using via Stan's (version 2.32.2) build-in HMC sampler accessed via R version 4.3.3 using package "brms" (Bürkner 2017; R Core Team 2024; Stan Development Team 2020). Four parallel chains were run each for 1,500 iterations for each GLMM with the first 500 iterations being discarded as a warm-up. Convergence was checked numerically by inspection of the $\hat{R}s$ and visually by inspection of trace plots. R packages "tidyverse," "tidybayes," "ggridges," and "patchwork" were used similarly to Study 2 (Wickham et al. 2019; Kay 2023; Wickham 2016; Wilke 2024; Pedersen 2020). Full analysis code is available at https://github.com/josefmana/dbs_postopIADL.git.

4 Results

4.1 Learning curve in verbal and non-verbal memory of patients with Parkinson's disease

4.1.1 Sample characteristics

In total, 60 HC participants and 60 patients with PD were out of which 25 were diagnosed with PD-MCI were included into the study. Demographic and clinical characteristics of the sample are presented in Table 2.

Table 2: Demographic characteristics of normative and control sample.

Variable	HC (N = 60)	PD-NC (N = 35)	PD-MCI (N = 25)
Age (years)	61.92 (3.98)	59.43 (8.62)	62.00 (9.71)
Education (years)	14.07 (2.57)	15.87 (3.13)	13.40 (2.89)

17

Sex (% male)	43.33	60.00	56.00
PD duration (years)	-	6.43 (6.22)	8.64 (6.10)
LEDD (mg)	-	840.88 (805.11)	1061.67 (653.08)
MoCA (range 0-30)	26.32 (2.30)	26.31 (1.64)	24.16 (3.10)
UPDRS III (range 0-132)	-	21.09 (12.03)	25.96 (13.97)

BDI-II: Beck Depression Rating Scale, second edition; HC: healthy control group; LEDD: levodopa equivalent daily dose; MoCA: Montreal Cognitive Assessment; N: number of observations; PD: Parkinson's Disease; PD-MCI: Mild Cognitive Impairment in Parkinson's Disease; PD-NC: Normal Cognition in Parkinson's Disease; UPDRS III: Unified Parkinson's Disease Rating Scale, motor part; all values represent mean (standard deviation) for continuous and percentages for nominal variables.

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

4.2.1 Sample characteristics

Baseline demographic and clinical characteristics as well as stimulation parameters of the sample are presented in Table 3 whereas average pre-surgery cognitive profile of the sample is presented in Table 4. Mean duration of a follow-up after the surgery was 3.54 years (SD = 2.32, median = 3.07, range = 0.72–11.38) with a median number of 3 assessments per patient (range = 2–6) (see also Figure 2).

Table 3: Clinical characteristics of the sample of included patients

	N	Md	Min.Max	M	SD
Baseline characteristics					
Age at surgery (years)	126	58	40-76	57.25	7.96
Education (years)	117	13	10-23	14.26	2.91
Sex (males)	83 (66 %)	-	-	-	-
Disease duration at surgery (years)	125	11	4-30	11.67	4.05
LEDD (mg)	114	1614	400-4138	1696.88	672.33
Levodopa test (% response)	93	54	20-81	52.64	12.81
MDS-UPDRS III (ON medication)	105	21	7-46	21.78	7.57
MDS-UPDRS III (OFF medication)	100	45	24-81	45.79	10.93
Stimulation parameters[1]					
Current right (mA)	67	2.1	0.6-4.3	2.14	0.71
Current left (mA)	67	2.3	1.0-3.9	2.35	0.68
Voltage right (V)	59	3.0	1.4-5.3	3.00	0.65

	N	Md	Min.Max	M	SD
Voltage left (V)	59	2.9	0.5-5.7	2.87	0.74
Pulse duration right (µs)	126	60.0	52.0-120.0	73.98	17.14
Pulse duration left (µs)	126	60.0	30.0-120.0	71.57	16.15
Frequency right (Hz)	126	130.0	60.0-210.0	128.42	12.44
Frequency left (Hz)	126	130.0	60.0-160.0	127.89	11.14

[1]Each measurement of each electrode considered independently. For stimulation parameters, column N indicate number of patients with current/voltage mode of stimulation.
N: number of observations; Md: median; M: mean; SD: standard deviation; MDS-UPDRS III: Movement Disorder Society Unified Parkinson's Disease Rating Scale, motor part; LEDD: levodopa equivalent daily dose; Levodopa test: a percentage change of the MDS-UPDRS III score from medication OFF to medication ON state during the levodopa test as described in the main text; V: Volts; mA: milliampere; s: microseconds; Hz: Hertz.

Table 4: Pre-surgery neuropsychological measures of included patients

Test	N	Md	Min.Max	M	SD
DRS-2 (range 0-144)	126	141	129-144	139.77	3.68
BDI-II (range 0-63)	122	8	0-28	9.28	5.95
STAI-X1 (range 20-80)	104	37	23-63	38.27	8.66
STAI-X2 (range 20-80)	104	39	22-62	39.52	8.11
TMT-A (secs)	125	41	18-122	43.15	15.85
TMT-B (secs)	124	102	39-334	119.01	54.96
DS-F (range 0-16)	113	8	5-16	8.94	2.02
DS-B (range 0-14)	113	6	2-11	6.21	1.80
LNS (range 0-21)	97	8	2-13	7.85	2.46
SS-F (range 0-16)	110	8	4-14	7.54	1.74
SS-B (range 0-16)	110	7	2-11	6.97	1.69
TOL (range 0-108)	118	78	46-90	74.93	9.81
PST-D (secs)	124	13	8-20	13.09	2.37
PST-W (secs)	124	15	10-25	15.72	2.97
PST-C (secs)	124	28	14-57	29.35	9.15
COWAT (total words)	125	32	12-57	32.35	9.05
CFT (words/min.)	89	22	3-39	22.55	7.10
Sim. (range 0-28)	94	22	8-28	21.61	4.35
RAVLT-IR (range 0-75)	108	44	20-64	43.80	8.39
RAVLT-B (range 0-15)	108	5	0-8	4.71	1.45
RAVLT-DR (range 0-15)	108	8	3-14	8.37	2.49
RAVLT-Rec50 (range 0-50)	105	46	33-50	45.10	3.49
RAVLT-Rec15 (range 0-15)	107	14	9-15	13.32	1.54
FP-IR (range 0-64)	74	32	15-55	32.04	10.21

			FP-DR (range 0-64)	74	32	13-55	31.91	9.97

N: number of observations; Md: median; M: mean; SD: standard deviation; DRS-2: Dementia Rating Scale, second edition; BDI-II: Beck Depression Rating Scale, second edition; 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; Secs: seconds; Total words: word count in two minutes (one minute per each letter P and K); words/min.: word count in one minute time limit.
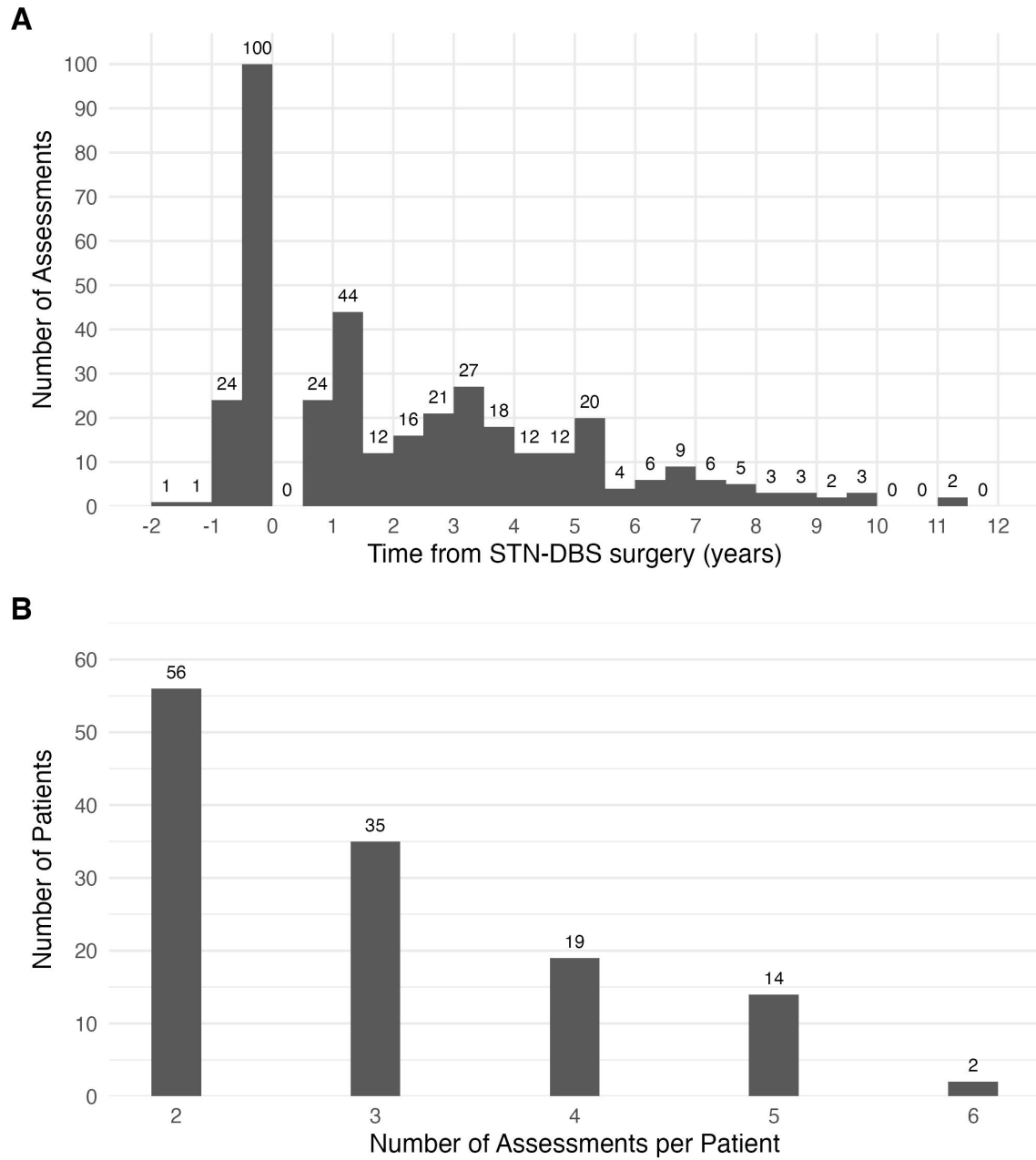
Figure 2: Distribution of assessments. Distribution of (A) follow-up years and (B) number of assessments per patient for N = 126 patients. Negative values on horizontal axis in (A) represent pre-surgery assessments, number of assessments in (B) includes one pre-surgery and various number of post-surgery assessments.

Beck, Aaron T., R. A. Steer, and G. Brown. 1996. "Beck Depression Inventory–II." American Psychological Association (APA). https://doi.org/10.1037/t00742-000.

Benedict, R H B. 1997. *Brief Visuospatial Memory Test- Revised: Professional Manual*. PAR.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *The Annals of Statistics* 29 (4): 1165–88. http://www.jstor.org/stable/2674075.

Bezdicek, Ondrej, Tommaso Ballarini, Herman Buschke, Filip Růžička, Jan Roth, Franziska Albrecht, Evžen Růžička, Karsten Mueller, Matthias L Schroeter, and Robert Jech. 2019. "Memory Impairment in Parkinson's Disease: The Retrieval Versus Associative Deficit Hypothesis Revisited and Reconciled." *Neuropsychology* 33 (3): 391–405. https://doi.org/10.1037/neu0000503.

Bezdicek, Ondrej, Tommaso Ballarini, Filip Růžička, Jan Roth, Karsten Mueller, Robert Jech, and Matthias L. Schroeter. 2018. "Mild Cognitive Impairment Disrupts Attention Network Connectivity in Parkinson's Disease: A Combined Multimodal MRI and Meta-Analytical Study." *Neuropsychologia* 112: 105–15. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2018.03.011.

Bezdicek, Ondrej, Jiri Lukavsky, Hana Stepankova, Tomas Nikolai, Bradley N. Axelrod, Jiri Michalec, Evžen Růžička, and Miloslav Kopecek. 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, Ondrej, Josef Mana, Filip Růžička, Filip Havlik, Anna Fečíková, Tereza Uhrová, Evžen Růžička, Dušan Urgošík, and Robert Jech. 2022. "The Instrumental Activities of Daily Living in Parkinson's Disease Patients Treated by Subthalamic Deep Brain Stimulation." *Frontiers in Aging Neuroscience* 14. https://doi.org/10.3389/fnagi.2022.886491.

Bezdicek, Ondrej, Jiri Michalec, Tomas Nikolai, Petra Havránková, Jan Roth, Robert Jech, and Evžen Růžička. 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): 303–11. https://doi.org/10.1159/000375365.

Bezdicek, Ondrej, L. Motak, B. N. Axelrod, M. Preiss, T. Nikolai, M. Vyhnalek, A. Poreh, and E. Ruzicka. 2012. "Czech Version of the Trail Making Test: Normative Data and Clinical Utility." *Archives of Clinical Neuropsychology* 27 (8): 906–14. https://doi.org/10.1093/arclin/acs084.

Bezdicek, Ondrej, Tomas Nikolai, Jiri Michalec, Filip Růžička, Petra Havránková, Jan Roth, Robert Jech, and Evžen Růžička. 2017. "The Diagnostic Accuracy of Parkinson's Disease Mild Cognitive Impairment Battery Using the Movement Disorder Society Task Force Criteria." *Movement Disorders Clinical Practice* 4 (2): 237–44. https://doi.org/https://doi.org/10.1002/mdc3.12391.

Bezdicek, Ondrej, Hana Stepankova, Bradley N. Axelrod, Tomas Nikolai, Zdenek Sulc, Robert Jech, Evžen Růžička, and Miloslav Kopecek. 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, Ondrej, Hana Stepankova, Ladislav Moták, Bradley N. Axelrod, John L. Woodard, Marek Preiss, Tomáš Nikolai, Evžen Růžička, and Amir Poreh. 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, Ondrej, Zdenek Sulc, Tomas Nikolai, Hana Stepankova, Miloslav Kopecek, Robert Jech, and Evžen Růžička. 2017. "A Parsimonious Scoring and Normative Calculator for the Parkinson's Disease Mild Cognitive Impairment Battery." *The Clinical Neuropsychologist* 31 (6-7): 1231–47. https://doi.org/10.1080/13854046.2017.1293161.

Blume, J., M. Lange, E. Rothenfusser, C. Doenitz, U. Bogdahn, A. Brawanski, and J. Schlaier. 2017. "The Impact of White Matter Lesions on the Cognitive Outcome of Subthalamic Nucleus Deep Brain Stimulation in Parkinson's Disease." *Clinical Neurology and Neurosurgery* 159: 87–92. https://doi.org/https://doi.org/10.1016/j.clineuro.2017.05.023.

Boel, Judith A., Vincent J. J. Odekerken, Ben A. Schmand, Gert J. Geurtsen, Danielle C. Cath, Martijn Figee, Pepijn van den Munckhof, et al. 2016. "Cognitive and Psychiatric Outcome 3 Years After Globus Pallidus Pars Interna or Subthalamic Nucleus Deep Brain Stimulation for Parkinson's Disease." *Parkinsonism & Related Disorders* 33: 90–95. https://doi.org/https://doi.org/10.1016/j.parkreldis.2016.09.018.

Brennan, Laura, Andrew Siderowf, Jonathan D. Rubright, Jacqueline Rick, Nabila Dahodwala, John E. Duda, Howard Hurtig, et al. 2016a. "The Penn Parkinson's Daily Activities Questionnaire-15: Psychometric Properties of a Brief Assessment of Cognitive Instrumental Activities of Daily Living in Parkinson's Disease." *Parkinsonism & Related Disorders* 25: 21–26. https://doi.org/https://doi.org/10.1016/j.parkreldis.2016.02.020.

———, et al. 2016b. "The Penn Parkinson's Daily Activities Questionnaire-15: Psychometric Properties of a Brief Assessment of Cognitive Instrumental Activities of Daily Living in Parkinson's Disease." *Parkinsonism & Related Disorders* 25: 21–26. https://doi.org/https://doi.org/10.1016/j.parkreldis.2016.02.020.

Brønnick, Kolbjørn, Guido Alves, Dag Aarsland, Ole-Bjørn Tysnes, and Jan Petter Larsen. 2011. "Verbal Memory in Drug-Naive, Newly Diagnosed Parkinson's Disease. The Retrieval Deficit Hypothesis Revisited." *Neuropsychology* 25 (1): 114—124. https://doi.org/10.1037/a0020857.

Browne, Michael W., and Robert Cudeck. 1992. "Alternative Ways of Assessing Model Fit." *Sociological Methods & Research* 21 (2): 230–58. https://doi.org/10.1177/0049124192021002005.

Bürkner, Paul-Christian. 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, Paul-Christian, and Matti Vuorre. 2019. "Ordinal Regression Models in Psychology: A Tutorial." *Advances in Methods and Practices in Psychological Science* 2 (1): 77–101. https://doi.org/10.1177/2515245918823199.

Chiaravalloti, Nancy D., Naroa Ibarretxe-Bilbao, John DeLuca, Olga Rusu, Javier Pena, Inés García-Gorostiaga, and Natalia Ojeda. 2014. "The Source of the Memory Impairment in

Parkinson's Disease: Acquisition Versus Retrieval." *Movement Disorders* 29 (6): 765–71. https://doi.org/https://doi.org/10.1002/mds.25842.

Ciharova, Marketa, Hynek Cígler, Veronika Dostálová, Gabriela Šivicová, and Ondrej Bezdicek. 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): 371–79. https://doi.org/10.1080/13651501.2020.1775854.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2022. "A Crash Course in Good and Bad Controls." *Sociological Methods & Research*, May, 004912412210995. https://doi.org/10.1177/00491241221099552.

Combs, Hannah L., Bradley S. Folley, David T. R. Berry, Suzanne C. Segerstrom, Dong Y. Han, Amelia J. Anderson-Mooney, Brittany D. Walls, and Craig van Horne. 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): 439–54. https://doi.org/10.1007/s11065-015-9302-0.

Costentin, Guillaume, Stéphane Derrey, Emmanuel Gérardin, Yohann Cruypeninck, Thibaut Pressat-Laffouilhere, Youssef Anouar, David Wallon, Floriane Le Goff, Marie-Laure Welter, and David Maltête. 2019. "White Matter Tracts Lesions and Decline of Verbal Fluency After Deep Brain Stimulation in Parkinson's Disease." *Human Brain Mapping* 40 (9): 2561–70. https://doi.org/https://doi.org/10.1002/hbm.24544.

Curtis, Ashley F., Mario Masellis, Richard Camicioli, Heather Davidson, and Mary C. Tierney. 2019. "Cognitive Profile of Non-Demented Parkinson's Disease: Meta-Analysis of Domain and Sex-Specific Deficits." *Parkinsonism & Related Disorders* 60: 32–42. https://doi.org/https://doi.org/10.1016/j.parkreldis.2018.10.014.

Defer, Gilles-Louis, Hakan Widner, Rose-Marie Marié, Philippe Rémy, and Marc Levivier. 1999. "Core Assessment Program for Surgical Interventional Therapies in Parkinson's Disease (CAPSIT-PD)." *Movement Disorders* 14 (4): 572–84. https://doi.org/https://doi.org/10.1002/1531-8257(199907)14:4%3C572::AID-MDS1005%3E3.0.CO;2-C.

Domellöf, M. E., U. Ekman, L. Forsgren, and E. Elgh. 2015. "Cognitive Function in the Early Phase of Parkinson's Disease, a Five-Year Follow-up." *Acta Neurologica Scandinavica* 132 (2): 79–88. https://doi.org/https://doi.org/10.1111/ane.12375.

Dubois, Bruno, David Burn, Christopher Goetz, Dag Aarsland, Richard G. Brown, Gerald A. Broe, Dennis Dickson, et al. 2007. "Diagnostic Procedures for Parkinson's Disease Dementia: Recommendations from the Movement Disorder Society Task Force." *Movement Disorders* 22 (16): 2314–24. https://doi.org/https://doi.org/10.1002/mds.21844.

Emre, Murat, Dag Aarsland, Richard Brown, David J. Burn, Charles Duyckaerts, Yoshikino Mizuno, Gerald Anthony Broe, et al. 2007. "Clinical Diagnostic Criteria for Dementia Associated with Parkinson's Disease." *Movement Disorders* 22 (12): 1689–1707. https://doi.org/https://doi.org/10.1002/mds.21507.

Ewert, Siobhan, Philip Plettig, Ningfei Li, M. Mallar Chakravarty, D. Louis Collins, Todd M. Herrington, Andrea A. Kühn, and Andreas Horn. 2018. "Toward Defining Deep Brain Stimulation Targets in MNI Space: A Subcortical Atlas Based on Multimodal MRI, Histology and Structural Connectivity." *NeuroImage* 170: 271–82. https://doi.org/https:

//doi.org/10.1016/j.neuroimage.2017.05.015.

Filip, Pavel, Josef Mana, Andrej Lasica, Jiří Keller, Dušan Urgošík, Jaromír May, Karsten Mueller, Robert Jech, Ondrej Bezdicek, and Filip Růžička. 2024. "Structural and Microstructural Predictors of Cognitive Decline in Deep Brain Stimulation of Subthalamic Nucleus in Parkinson's Disease." *NeuroImage: Clinical*.

Frydrychová, Zuzana, Miloslav Kopeček, Ondrej Bezdicek, and Hana Georgi Stepankova. 2018. "Czech normative study of the Revised Rey Auditory Verbal Learning Test (RAVLT) in older adults." *Ceskoslovenska Psychologie* 62 (4): 330–49.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Analytical Methods for Social Research. Cambridge University Press.

Glasser, Matthew F., Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, et al. 2013. "The Minimal Preprocessing Pipelines for the Human Connectome Project." *NeuroImage* 80: 105–24. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.04.127.

Grabner, Günther, Andrew L. Janke, Marc M. Budge, David Smith, Jens Pruessner, and D. Louis Collins. 2006. "Symmetric Atlasing and Model Based Segmentation: An Application to the Hippocampus in Older Adults." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, edited by Mads Larsen Rasmusand Nielsen and Jon Sporring, 58–66. Berlin, Heidelberg: Springer Berlin Heidelberg.

Gruber, Doreen, Lisa Calmbach, Andrea A. Kühn, Patricia Krause, Ute A. Kopp, Gerd-Helge Schneider, and Andreas Kupsch. 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): 309–18. https://doi.org/10.1007/s00702-019-01972-7.

Harman, H. H., and W. H. Jones. 1966. "Factor Analysis by Minimizing Residuals (Minres)." *Psychometrica* 31 (3): 351–68.

Havlík, Filip, Josef Mana, Petr Dušek, Robert Jech, Evžen Růžička, Miloslav Kopeček, Hana Georgi, and Ondrej Bezdicek. 2020. "Brief Visuospatial Memory Test-Revised: Normative Data and Clinical Utility of Learning Indices in Parkinson's Disease." *Journal of Clinical and Experimental Neuropsychology* 42 (10): 1099–1110. https://doi.org/10.1080/13803395.2020.1845303.

Horn, Andreas, and Andrea A. Kühn. 2015. "Lead-DBS: A Toolbox for Deep Brain Stimulation Electrode Localizations and Visualizations." *NeuroImage* 107: 127–35. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.12.002.

Horn, Andreas, Ningfei Li, Till A. Dembek, Ari Kappel, Chadwick Boulay, Siobhan Ewert, Anna Tietze, et al. 2019. "Lead-DBS V2: Towards a Comprehensive Pipeline for Deep Brain Stimulation Imaging." *NeuroImage* 184: 293–316. https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.08.068.

Hughes, A J, S E Daniel, L Kilford, and A J Lees. 1992. "Accuracy of Clinical Diagnosis of Idiopathic Parkinson's Disease: A Clinico-Pathological Study of 100 Cases." *Journal of Neurology, Neurosurgery & Psychiatry* 55 (3): 181–84. https://doi.org/10.1136/jnnp.55.3.181.

Josse, Julie, and Francois Husson. 2016. "{missMDA}: A Package for Handling Missing Values in Multivariate Data Analysis" 70. https://doi.org/10.18637/jss.v070.i01.

Jurica, P. J., C. L. Leitten, and S. Mattis. 2001. *Dementia Rating Scale-2 (DRS-2) Professional Manual.* Psychological Assessment Resources.

Kay, Matthew. 2023. *tidybayes: Tidy Data and Geoms for Bayesian Models.* https://doi.org/10.5281/zenodo.1308151.

Kim, Han-Joon, Beom S. Jeon, Sun Ha Paek, Kyoung-Min Lee, Ji-Young Kim, Jee-Young Lee, Hee Jin Kim, et al. 2014. "Long-Term Cognitive Outcome of Bilateral Subthalamic Deep Brain Stimulation in Parkinson's Disease." *Journal of Neurology* 261 (6): 1090–96. https://doi.org/10.1007/s00415-014-7321-z.

Kopecek, Miloslav, Hana Stepankova, Jiri Lukavsky, Daniela Ripova, Tomas Nikolai, and Ondrej Bezdicek. 2017. "Montreal Cognitive Assessment (MoCA): Normative Data for Old and Very Old Czech Adults." *Applied Neuropsychology: Adult* 24 (1): 23–29. https://doi.org/10.1080/23279095.2015.1065261.

Kubu, Cynthia S. 2018. "The Role of a Neuropsychologist on a Movement Disorders Deep Brain Stimulation Team." *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists* 33 (3): 365—374. https://doi.org/10.1093/arclin/acx130.

Liddell, Torrin M., and John K. Kruschke. 2018. "Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong?" *Journal of Experimental Social Psychology* 79: 328–48. https://doi.org/https://doi.org/10.1016/j.jesp.2018.08.009.

Litvan, Irene, Jennifer G. Goldman, Alexander I. Tröster, Ben A. Schmand, Daniel Weintraub, Ronald C. Petersen, Brit Mollenhauer, et al. 2012. "Diagnostic Criteria for Mild Cognitive Impairment in Parkinson's Disease: *Movement* Disorder Society Task Force Guidelines." *Movement Disorders* 27 (3): 349–56. https://doi.org/10.1002/mds.24893.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86 (3): 532–65. https://doi.org/10.1177/00031224211004187.

Makowski, Dominique, Mattan S. Ben-Shachar, S. H. Annabel Chen, and Daniel Lüdecke. 2019. "Indices of Effect Existence and Significance in the Bayesian Framework." *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.02767.

Mana, Josef, Ondrej Bezdicek, Filip Růžička, Andrej Lasica, Anna Šmídová, Olga Klempířová, Tomáš Nikolai, et al. 2024. "Preoperative Cognitive Profile Predictive of Cognitive Decline After Subthalamic Deep Brain Stimulation in Parkinson's Disease." https://doi.org/10.22541/au.169661569.99978744/v1.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and STAN.* Analytical Methods for Social Research. Chapman; Hall/CRC. https://doi.org/https://doi.org/10.1201/9780429029608.

Mehanna, Raja, Jawad A. Bajwa, Hubert Fernandez, and Aparna Ashutosh Wagle Shukla. 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, Jiri, Ondrej Bezdicek, Tomas Nikolai, Pavel Harsa, Robert Jech, Petr Silhan, Martin Hyza, Evzen Ruzicka, and Tim Shallice. 2017. "A Comparative Study of Tower of London

Scoring Systems and Normative Data." *Archives of Clinical Neuropsychology*, January. https://doi.org/10.1093/arclin/acw111.

Nasreddine, Ziad S., Natalie A. Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L. Cummings, and Howard Chertkow. 2005. "The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment." *Journal of the American Geriatrics Society* 53 (4): 695–99. https://doi.org/https://doi.org/10.1111/j.1532-5415.2005.53221.x.

Nikolai, Tomáš, Hana Stepankova, Jiří Michalec, Ondřej Bezdicek, Karolína Horáková, Hana Marková, Evžen Ruzicka, and Miloslav Kopecek. 2015. "Tests of Verbal Fluency, Czech Normative Study in Older Patients." *Česká a Slovenská Neurologie a Neurochirurgie* 78/111 (3): 292–99. https://doi.org/10.14735/amcsnn2015292.

Pal, Gian, Graziella Mangone, Emily J. Hill, Bichun Ouyang, Yuanqing Liu, Vanessa Lythe, Debra Ehrlich, et al. 2022. "Parkinson Disease and Subthalamic Nucleus Deep Brain Stimulation: Cognitive Effects in GBA Mutation Carriers." *Annals of Neurology* 91 (3): 424–35. https://doi.org/https://doi.org/10.1002/ana.26302.

Park, Trevor, and George Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103 (482): 681–86. https://doi.org/10.1198/016214508000000337.

Parsons, Thomas D, Steven A Rogers, Alyssa J Braaten, Steven Paul Woods, and Alexander I Tröster. 2006. "Cognitive Sequelae of Subthalamic Nucleus Deep Brain Stimulation in Parkinson's Disease: A Meta-Analysis." *The Lancet Neurology* 5 (7): 578–88. https://doi.org/10.1016/s1474-4422(06)70475-6.

Partington, J E, and R G Leiter. 1949. "Partington's Pathways Test." *Psychological Service Center Journal* 1: 11–20.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference.* Cambridge University Press. https://doi.org/10.1017/cbo9780511803161.

Pedersen, Thomas Lin. 2020. "Patchwork: The Composer of Plots." https://CRAN.R-project.org/package=patchwork.

Planche, Vincent, Fanny Munsch, Bruno Pereira, Emmanuel de Schlichting, Tiphaine Vidal, Jerome Coste, Dominique Morand, et al. 2018. "Anatomical Predictors of Cognitive Decline After Subthalamic Stimulation in Parkinson's Disease." *Brain Structure &Amp; Function* 223 (7). https://doi.org/10.1007/s00429-018-1677-2.

Postuma, Ronald B., Daniela Berg, Matthew Stern, Werner Poewe, C. Warren Olanow, Wolfgang Oertel, José Obeso, et al. 2015. "MDS Clinical Diagnostic Criteria for Parkinson's Disease." *Movement Disorders* 30 (12): 1591–601. https://doi.org/https://doi.org/10.1002/mds.26424.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Revelle, William. 2022. "Psych: Procedures for Psychological, Psychometric, and Personality Research." https://CRAN.R-project.org/package=psych.

Samejima, F. 1995. "Acceleration Model in the Heterogeneous Case of the General Graded Response Model." *Psychometrika* 60 (4): 549–72. https://doi.org/https://doi.org/10.1007/BF02294328.

Shallice, Timothy. 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, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis.* Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001.

Spielberger, C. D., R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs. 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/.

Štěpánková, Hana, Ondřej Bezdíček, Tomáš Nikolai, Karolína Horáková, Jiří Lukavský, and Miloslav Kopeček. 2015. "National Normative Study of Cognitive Determinants of Healthy Ageing-status report." *E-Psychologie* 9 (1): 1689–1707. https://e-psycholog.eu/clanek/224.

Thomson, G. 1951. *The Factorial Analysis of Human Ability.* University of London Press.

Tomlinson, Claire L., Rebecca Stowe, Smitaa Patel, Caroline Rick, Richard Gray, and Carl E. Clarke. 2010. "Systematic Review of Levodopa Dose Equivalency Reporting in Parkinson's Disease." *Movement Disorders* 25 (15): 2649–53. https://doi.org/10.1002/mds.23429.

Tuerlinckx, Francis, Frank Rijmen, Geert Verbeke, and Paul De Boeck. 2006. "Statistical Inference in Generalized Linear Mixed Models: A Review." *British Journal of Mathematical and Statistical Psychology* 59 (2): 225–55. https://doi.org/10.1348/000711005x79857.

Wechsler, D. 2010. *Wechsler Adult Intelligence Scale - Third Revision.* Praha: Hogrefe - Testcentrum.

———. 2011. *Wechsler Memory Scale -Third Edition Abbreviated.* Praha: Hogrefe - Testcentrum.

Wickham, Hadley. 2016. "Ggplot2: Elegant Graphics for Data Analysis." https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wilke, Claus O. 2024. *Ggridges: Ridgeline Plots in 'Ggplot2'.* https://wilkelab.org/ggridges/.

Winkler, Anderson M, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols. 2014. "Permutation Inference for the General Linear Model." *NeuroImage* 92: 381–97. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.01.060.

Yarkoni, Tal. 2020. "The Generalizability Crisis." *Behavioral and Brain Sciences* 45 (December). https://doi.org/10.1017/s0140525x20001685.