Association of obstructive sleep apnea with brain volummetry and cognition in de novo Parkinson's disease

Josef Mana

2024-07-30

Methods

Neuropsychological assessment

All patients (PD) and healthy controls (HC) were administered a battery of neuropsychological tests at enrollment to the study and a subsample of participants were administered the same tests at re-test four years after the enrollment. The battery included assessment of (i) declarative memory via Rey Auditory Verbal Learing Test (RAVLT) (Bezdicek et al. 2014; Frydrychová et al. 2018); (ii) attention via Trail Making Test, part A (TMT-A) (Bezdicek et al. 2012; Bezdicek, Stepankova, et al. 2017), and dot colour naming (PST-D) as well as naming colour of neutral words (PST-W) conditions from Prague Stroop Test (Bezdicek et al. 2015); (iii) executive function via Trail Making Test, part B (Bezdicek et al. 2012; Bezdicek, Stepankova, et al. 2017), and Prague Stroop Test, interference condition (i.e., naming colour of contrasting colour words, PST-C) (Bezdicek et al. 2015); and (iv) processing speed via Grooved Pegboard Test (GPT) (Kløve 1963). The patients were further examined using tests from the standard International Parkinson and Movement Disorder Society (MDS) neuropsychological battery at Level II for mild cognitive impairment in Parkinson's disease (PD-MCI) (Litvan et al. 2012; Bezdicek, Sulc, et al. 2017). The Czech normative calculator established by Bezdicek, Sulc, et al. (2017) was used to assign PD-MCI diagnosis to each PD patient separately at enrollment and retest. Finally, all participants were administered Montreal cognitive assessment (MoCA) (Kopecek et al. 2017; Nasreddine et al. 2005) annually for cognitive screening.

Statistical analyses

All demographic (age, education, gender) and descriptive clinical (BMI, BDI-II, MDS-UPDRS III, disease duration) variables were described by their mean and standard deviation if continuous and frequency if nominal separately for HC OSA-, HC OSA+, PD OSA-, and PD OSA+ groups of participants, and compared by Gaussian (continuous variables) or logistic (binary variables) regression with group (PD vs HC), OSA (OSA+ vs OSA-) and their interaction as predictors. As we did not aim to control type I error rate in these analyses, decision threshold for claiming statistically significant difference in demographic and decscriptive clinical variables was set at $p < .05$ without adjustment for multiple comparisons.

Cortical thickness

4 Fíla

Subcortical volummetry

The strength of association of subcortical structures' volume with group (PD vs HC) and OSA (OSA+ vs OSA-) was evaluated by a set of univariate linear regressions fitted to data via QR decomposition. Each subcortical structure's volume was regressed on group, OSA and their interaction as exposures of interest, and years of age, gender and TIV as additive linear covariates. All continuous variables were standardised (i.e., mean-centred and scaled by their in-sample standard deviation) before entering the analysis, and nominal variables were entered via sum coding. Each regression was tested for deviation from normality of residuals via the Shapiro-Wilk test and deviation from homoscedasticity via the Breusch-Pagan test. Within the classical Neyman-Pearson hypothesis testing framework, we selected a decision threshold for rejecting the hypothesis of zero association between subcortical structure's volume and group/OSA by adjusting for 5% false discovery rate (FDR) (Benjamini and Hochberg 1995) in tests of all main effects and interactions of interest (i.e., coefficients related to the group and OSA variables) across all subcortical structures examined in this section. The test statistics consisted of t-values for each relevant regression coefficient calculated by lm() function in R software for statistical computing (R Core Team 2024). The primary estimand of interest, i.e., the interaction between group and OSA, was further characterised by calculating simple main effects of OSA within each group and the difference between these simple main effects via the avg_comparisons() function in the "marginaleffects" R package (Arel-Bundock, Greifer, and Heiss Forthcoming).

Based on the results of the primary analyses outlined above, we further explored association of hippocampal areas with group and OSA in a series of post-hoc univariate regressions with identical specification as described above. The only difference between the primary analyses of the gross subcortical structures described above, and the analysis of hippocampal areas was that in the latter case, we focused on the group/OSA interaction only. The decision threshold

for rejecting the hypothesis of zero group/OSA interaction was set according to adjustment for 5% FDR within this set of analyses.

Finally, to reflect current recommendations of the American Statistical Association which advised against basing scientific conclusions on whether a p-value passes a specific threshold (Wasserstein and Lazar 2016), we also calculated Shannon information (i.e., s-values) $-log_2(p)$. S-value is cognitive tool to help researchers intuitively evaluate strength of evidence against a null hypothesis contained in the results as equivalent to the number consecutive "heads" tosses that would provide the same amount of evidence against the null hypothesis that the coin is fair (Greenland 2019; Cole, Edwards, and Greenland 2021).

Cognitive variables

The distribution of cognitive performance conditional on group, OSA, and measurement occasion (enrollment vs retest) was evaluated using a set of independent Bayesian linear mixed models (LMMs) with cognitive test scores regressed on group, OSA, occasion, and their interactions as fixed effects on group-level and participant-specific random intercepts. All outcomes were standardised before entering the analysis and response time variables were log-transformed before standardising. Contrasts described by Rouder et al. (2012) as implemented in contr.equalprior() function from the "bayestestR" R package (Makowski, Ben-Shachar, and Lüdecke 2019) were specified for all group-level parameters to ensure that all pairwise prior differences are centred around zero. Weakly informative priors, i.e. *Normal(0,1)* for group-level parameters and *Exponential(1)* for participant-level and residual variances, were specified to ensure the model converges to reasonable parameter values. Employing LMMs allowed us to use all data without requiring the participants to have both enrollment and retest measurements.

To ensure the results are informed by data, prior and likelihood sensitivity of posterior estimates was evaluated via the powerscale_sensitivity() function from "priorsense" R package (Kallioinen et al. 2023). Moreover, to evaluate model fit, posterior predictive means and SDs of the model within each combination of group, OSA, and assessment were visually compared to values observed in the data. All LMMs were fitted using Stan's (version 2.21.0) build-in Hamiltonian Monte Carlo 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 2,000 iterations for each LMM with the first 1,000 iterations serving as a warm-up. Convergence was checked numerically by inspection of the $\widehat{R}s$ and visually by inspection of trace plots.

After establishing model fit, the results were summarised by computing pairwise comparisons based on main effects, two-way interactions and three-way interactions between group, OSA and occasion. These comparisons were then described by their medians, 95% equal-tailed posterior probability intervals (PPIs), and probability of direction (*pd*, i.e., the certainty associated with the most probable direction of the effect) on the original outcome scale. Since

Table 1: Demographic and descriptive clinical variables

this portion of our analysis was purely exploratory and was not set-up to formally test any hypothesis, we did not set any decision threshold regarding reported posterior comparisons but instead interpreted *pd* as continous index of effect existence probability following guidelines of Makowski et al. (2019) whereby *pd* 95% indicates uncertain effect, $pd > 95\%$ indicates possibly existing effect, $pd > 97\%$ indicates likely existing effect, $pd > 99\%$ indicates probably existing effect, and $pd > 99.9\%$ indicates almost certainly existing effect.

Results

Sample description

Cortical thickness

4 Fíla

Subcortical volummetry

Cognitive variables

Appendix

References

Arel-Bundock, Vincent, Noah Greifer, and Andrew Heiss. Forthcoming. "How to Interpret Statistical Models Using marginaleffects in R and Python." *Journal of Statistical Software*, Forthcoming.

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.

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, 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, 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.

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.

Cole, Stephen R, Jessie K Edwards, and Sander Greenland. 2021. "Surprise!" *American Journal of Epidemiology* 190 (2): 191–93.

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.

Greenland, Sander. 2019. "Valid p-Values Behave Exactly as They Should: Some Misleading Criticisms of p-Values and Their Resolution with s-Values." *The American Statistician* 73 (sup1): 106–14.

Kallioinen, Noa, Topi Paananen, Paul-Christian Bürkner, and Aki Vehtari. 2023. "Detecting and Diagnosing Prior and Likelihood Sensitivity with Power-Scaling." *Statistics and Computing* 34. https://doi.org/10.1007/s11222-023-10366-5.

Kløve, Hallgrim. 1963. "Clinical Neuropsychology." *Medical Clinics of North America* 47 (6): 1647–58. https://doi.org/https://doi.org/10.1016/S0025-7125(16)33515-5.

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.

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.

Makowski, Dominique, Mattan S Ben-Shachar, SH Annabel Chen, and Daniel Lüdecke. 2019. "Indices of Effect Existence and Significance in the Bayesian Framework." *Frontiers in Psychology* 10: 2767.

Makowski, Dominique, Mattan S. Ben-Shachar, and Daniel Lüdecke. 2019. "bayestestR: Describing Effects and Their Uncertainty, Existence and Significance Within the Bayesian Framework." *Journal of Open Source Software* 4 (40): 1541. https://doi.org/10.21105/joss.01541.

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.

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

Rouder, Jeffrey N, Richard D Morey, Paul L Speckman, and Jordan M Province. 2012. "Default Bayes Factors for ANOVA Designs." *Journal of Mathematical Psychology* 56 (5): 356–74.

Stan Development Team. 2020. "Stan Modeling Language Users Guide and Reference Manual, Version 2.21.0." http://mc-stan.org/.

Wasserstein, Ronald L, and Nicole A Lazar. 2016. "The ASA Statement on p-Values: Context, Process, and Purpose." *The American Statistician*. Taylor & Francis.