Updated Criteria for the Diagnostic Procedure for Parkinson’s Disease Dementia on Level I and their Validity in Deep Brain Stimulation Cohort

Martina Mana, Josef Mana, Tereza Uhrova, Robert Jech, and Ondrej Bezdicek

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

Author Note

Martina Mana https://orcid.org/0009-0007-4665-3946

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

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

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

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Martina Mana: Conceptualization, Data curation, Writing - original draft; Josef Mana: Conceptualization, Data curation, Investigation, Formal analysis, Software, Methodology, Project administration, Validation, Writing - original draft; Tereza Uhrova: Investigation; Robert Jech: Funding acquisition, Resources, Writing - review & editing; Ondrej Bezdicek: Investigation, Data curation, Funding acquisition, Conceptualization, Project administration, Supervision, Writing - original draft

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

Updated Criteria for the Diagnostic Procedure for Parkinson's Disease Dementia on Level I and their Validity in Deep Brain Stimulation Cohort

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder characterised by a progressive onset of motor symptoms, including rigidity, bradykinesia, postural instability and resting tremor. Beyond motor symptoms, patients routinely suffer from non-motor impairments (Postuma et al., 2015), such as cognitive decline. This decline might peak in Parkinson's disease dementia (PDD) in a subset of patients (Meireles & Massano, 2012).

Despite its clinical relevance, the diagnosis of PDD remains complex. The original International Parkinson and Movement Disorder Society (MDS) diagnostic criteria (Dubois et al., 2007), first formalized in 2007, were heavily influenced by frameworks established primarily for Alzheimer's disease (AD). Although these criteria provided a valuable initial foundation, they lacked the specificity required to capture the distinct pathophysiological and cognitive features of PD-related dementia (Emre et al., 2007; Yamashita et al., 2023).

To enhance diagnostic accuracy, the MDS introduced a two-level system for Parkinson's disease dementia. Level I criteria are designed for clinical use, relying on brief cognitive assessments and clinical judgment. Level II criteria involve comprehensive neuropsychological testing across multiple domains, providing greater diagnostic detail for research or specialized settings (Emre et al., 2007).

One notable feature of the original Level I criteria was the provision of an algorithm that allowed for flexibility in test selection (Dubois et al., 2007). Specifically, clinicians could choose between months reversed or seven backwards for attention assessment, lexical fluency or clock drawing for executive function evaluation. Moreover, the critaria included MMSE pentagons for evaluation of visuospatial ability, and three-word recall for memory assessment. Agreement across different criteria allows for the parallel computation of inter-rater reliability , which, in turn, facilitates the calculation of construct validity (Conway et al., 1995) further strengthening the diagnostic framework for PDD.

According to a recent metanalysis, the epidemiological estimates of PDD prevalence among individuals with PD vary widely, ranging from 14% to 55%, depending on methodological criteria employed (Sousa et al., 2022). Moreover, factors such as patients' sex (Cereda et al., 2016), age and disease duration appear to modulate the risk of cognitive decline (Oh et al., 2016; Rana et al., 2011).

Currently, efforts are focused on refining the PDD diagnostic framework to improve its consistency and applicability in both research and clinical contexts (Kulisevsky et al., 2024).

The present study evaluates the diagnostic concordance between the original Level I PDD criteria, as established by the MDS Task Force (Dubois et al., 2007; Emre et al., 2007) and criteria inspired by the recent call for change (Kulisevsky et al., 2024) within a cohort of PD patients considered for DBS. Furthermore, both sets of criteria are compared to PDD diagnosed on Level II. The study aims to address following research objectives (RO): (RO1) To estimate the prevalence of Parkinson's disease dementia (PDD) and evaluate the diagnostic variability and concordance across different PDD criteria. (RO2) To identify specific diagnostic components that contribute to variability in PDD classification across the applied criteria.

Methods

Participants

This study retrospectively analyzed clinical data from a cohort of patients with PD at the General University Hospital in Prague. All patients were diagnosed with idiopathic PD by a movement disorder specialist according to the MDS Clinical Diagnostic Criteria for PD (Postuma et al., 2015). Clinical records spanning August 2014 to February 2025 were examined. All participants underwent neuropsychological evaluation conducted by a trained clinical psychologist (OB) as part of standard preoperative cognitive assessments for DBS eligibility at the General University Hospital in Prague.

Ethical approval for the study protocol was obtained from the Ethics Committee of the General University Hospital in Prague. Informed consent was secured from all patients prior to their neuropsychological assessments, in adherence to ethical research guidelines.

Neuropsychological Assessment

Cognitive performance was evaluated at both Level I (abbreviated assessment) and Level II (comprehensive assessment) according to the standard MDS neuropsychology battery for Parkinson's Disease Mild Cognitive Impairment (PD-MCI) Bezdicek et al. (2017). Level I was assessed using the Mini-Mental State Examination (MMSE) (Folstein et al., 1975; Stepankova et al., 2015) and the Montreal Cognitive Assessment (MoCA) (Kopecek et al., 2016; Nasreddine et al., 2005). The neuropsychological assessment at Level II covered five cognitive domains, each evaluated through specific tests as follows: attention and working memory assessed using Trail Making Test Part A (TMT-A) (Bezdicek et al., 2012; Reitan, 2004), and WAIS Digit Span Backward (WAIS DSB) (Wechsler, 1997); executive function evaluated via Categorical Verbal Fluency (CF) (Benton et al., 1989)[1], and subtest from the Prague Stroop Test – Colors (PST-C) (Bezdíček et al., 2021); language measured with the WAIS Similarities subtest (Wechsler, 1997), and the Boston Naming Test (BNT-60) (Kaplan et al., 1983); memory examined using the Rey Auditory Verbal Learning Test (RAVLT) (Frydrychová et al., 2018; Rey, 1964)[2] for delayed recall, and the Brief Visual Memory Test–Revised (BVMTR) (Benedict, 1997; Havlík et al., 2020) for *delayed recall*[3], or WAIS Family Pictures subtest (Wechsler, 1997) *including delayed recall and forced choice recognition*[4]; visuospatial function assessed through the Judgment of Line Orientation Test (JoLO) (Benton et al., 1983), and Clock Drawing Test (CLOX) (Royall et al., 1998).

In addition to the core cognitive assessments, *tasks such as*[5] the Clock Drawing Test (CDT) and Letter Fluency tasks were included to capture domain-specific impairments. The classification of Parkinson's Disease Dementia (PDD) based on Level I criteria was determined using established scoring thresholds from the original criteria (Dubois et al., 2007) with corresponding MoCA equivalents.

To assess functional impairment, the Functional Activities Questionnaire (FAQ) (Bezdicek et al., 2016; Pfeffer et al., 1982) was administered. Additionally, neuropsychiatric

[1] **JM: This one instead - https://doi.org/10.14735/amcsnn2015292**

[2] **JM: Do not forget this - https://doi.org/10.1080/13825585.2013.865699**

[3] **JM: The same label is used for RAVLT-DR and BVMT-R-DR here, we should sort it out.**

[4] **JM: We did not use any recognition. Should tidy this section such that it includes only indexes used by our study.**

[5] **JM: Stating such as "such as" are usually not appropritate in Methods section of a scientific study, because all methods should be reported exactly.**

status was evaluated using the Beck Depression Inventory-II (BDI-II) (Beck et al., 1996; Ciharova et al., 2020) and State-Trait Anxiety Inventory (STAI) (Mullner et al., 1980; Spielberger et al., 1983). Psychotic symptoms were assessed through structured psychiatric interviews conducted by a trained psychiatrist.

Diagnostic algorithms for probable Parkinson's Disease Dementia

In this study, we applied three distinct sets of diagnostic algorithms for probable PDD at Level I. The first set was based on the original framework (Dubois et al., 2007), which utilized the Mini-Mental State Examination (MMSE) as a global cognitive screening tool, supplemented by assessments of attention, executive function, visuospatial abilities, and memory. The second set of algorithms was based on the recent call for change of dementia diagnostic guidelines (Kulisevsky et al., 2024), which advocates for more sensitive cognitive domain assessments in the context of PD. This updated approach incorporated specific items from the Montreal Cognitive Assessment (MoCA). The third approach applied the Czech version of the shortened Montreal Cognitive Assessment (sMoCA) (Bezdicek et al., 2020), a time-efficient modification designed to measure global cognitive performance using a reduced testing protocol that omits items providing redundant information. Lastly, the fourth approach followed the Level II protocol for diagnosis of PDD and Mild Cognitive Impairment in PD (PD-MCI) (Dubois et al., 2007; Litvan et al., 2012). The Level II methodology, including the use of a regression-based normative scoring approach, has been detailed in a prior study (Bezdicek et al., 2017). In this study, the thresholds for cognitive impairment at Level II were set at $z \leq -1.5$. All non-cognitive criteria of probable PDD (i.e., diagnosis of PD that developed before dementia and absence of Major Depression, delirium or other abnormalities that obscure diagnosis) held true for all patients in the sample according to the psychiatric and neurological examinations.

For each of these diagnostic approaches, we applied two operationalizations of deficits in Instrumental Activities of Daily Living (IADL). First, we utilized FAQ item 9, which approximates the pill questionnaire from the original criteria (Dubois et al., 2007) employing a cut-off score of 2 points or higher. Second, we applied the entire Functional Activities Questionnaire (FAQ) as suggested in the call for change (Kulisevsky et al., 2024), employing a cut-off score of 7 points or higher based on Czech normative data (Bezdíček et al., 2011). These methodologies resulted in a total of 68 algorithms, which were distributed across different diagnostic criteria: 4 MMSE-based, 60 MoCA-based, 2 sMoCA-based, and 2 based on the Level II battery (see Table 1 and Appendix Table A1 for the exact specification of each algorithm).

[Insert Table 1 here]

Statistical Analyses

Following the framework proposed by Lundberg et al. (2021), in this study we explicitly connect our research objectives and their corresponding theoretical (i.e., targets of inference) and empirical (i.e., data-driven) estimands to statistical estimates. The theoretical estimand refers to a unit-specific quantity defined over a target population and represents the ideal quantity that would address the research question under optimal conditions, such as access to complete population data or perfect experimental control. In contrast, the empirical estimand corresponds to the quantity that is actually computable using the available dataset, given real world constraints. The full description of the study's estimands and their relation to our research objectives is presented in the Appendix (see Table A2).

To address study objectives, we started by repeatedly assigning each patient the diagnosis of probable PDD based on each PDD algorithm listed in Table 1 (see also Table A1) resulting in a 204 (patients) × 68 (operationalizations) matrix where each cell indicates whether a patient (row) meets criteria for probable PDD according to an algorithm (column)[6]. PDD rate estimates were computed as $\frac{N_{PDD}}{N_{total}}$ separately for each algortihm. The predictive value of age and sex was then evaluated by fitting a set of logistic regressions, one for each algorithm for probable PDD, whereby the probable PDD was predicted by age, sex and their interaction.

Next, a set of two class cross-tabulations with associated statistics was computed for each pair of algorithms via the `confusionMatrix()` function from the R package *caret* (Kuhn, 2008). For each pair of algorithms, the analysis was repeated twice such that each

[6] **JM: This could and should be shared most likely, as long, as we anonymize properly. Let's ask Oto Mestek and the NPO team how and if is it possible.**

variable of the pair served once as the reference and once as the predictor. Following measures were used to evaluate pairwise concordance between different algorithms for probable PDD: 1) Cohen's κ with its 95% confidence interval (CI) computed via the `cohen.kappa()` function from the R package *psych* (William Revelle, 2024); 2) Accuracy (i.e., the proportion of correct predictions, both true positives and true negatives, among the total number of cases) with its 95% CI; 3) Sensitivity/Recall (i.e., the proportion of true positives); and 4) Specificity (i.e., the proportion of true negatives).[7]

Finally, the No Information Rate (NIR) was calculated for each pair of algorithms. NIR is the accuracy that could be obtained by always predicting the majority class and in our case it is equivalent to the complement of the PDD rate estimate according to the reference algorithm. Accuracy of prediction was compared to the NIR via a one-sided Exact Binomial Test as implemented by the `binom.test()` R stats function. Reference/predictor pairs associated with p < .05 were considered to show significantly better accuracy than NIR. In other words, for reference/predictor pairs associated with p < .05, we conclude that knowing the probable PDD status according to the predictor algorithm helps to estimate the probable PDD status according to the reference algorithm and the two algorithms thus show substantial concordance.

Data wrangling and visualizations were done in the *tidyverse* package (Wickham et al., 2019) and tables were formatted in the *gt* package (Iannone et al., 2024). All analyses were conducted within the R (version 4.3.3) software environment for statistical computing (R Core Team, 2024). The software code supporting this article is available at https://github.com/josefmana/DemCr1t.git.[8]

[7] Unlike Cohen's κ, Accuracy, Sensitivity and Specificity are not symmetrical, i.e., their value depends on which variable is considered reference and which is considered predictor. Consequently, we report these values twice for each pair of algorithms. Note that the Sensitivity of a reference/predictor pair corresponds to the Positive Predictive Value if their roles were reversed. The same relationship holds true between the Specificity and the Negative Predictive Value.

[8] **JM: Do not forget to make it public before submitted!**

Results

Sample Description

A total of 204 patients were included. The sample included 126 (62%) men, with an average of 58.96 (SD = 8.35) years of age, 13.75 (SD = 3.07) years of education, 10.79 (SD = 4.23) years of disease duration, 37.43 (SD = 13.04) Unified Parkinson Disease Rating Scale (UPDRS), part III in medication OFF state and 15.97 (SD = 8.25) UPDRS III in medication ON state. Cognitive characteristics of the sample are summarized in Table 2.

[Insert Table 2 here]

PDD Rate Estimates

Algorithm-wise rate of PDD estimates are presented in Table A3. On average, estimated PDD rate was 6.03% (SD = 3.46, Md = 3.94, range 2.00-16.75). Notably, the estimates were substantially lower when FAQ item 9 was used as a criterion of IADL deficit (M = 3.11% SD = 0.48, Md = 2.96, range 2.00-3.94) compared to using total FAQ score criterion (M = 8.95% SD = 2.54, Md = 9.11, range 3.50-16.75) as demonstrated also in Figure 1. Neither age nor sex or their interaction (*p*s ≥ .110) reliably predicted probable PDD classification across algorithms (see Figure A1 and Figure A2).

[Insert Figure 1 here]

Concordance between Algorithms

Results of the analyses of prediction Accuracy, Cohen's κ, Sensitivity and Specificity are presented in Figure 2, Figure A3, Figure A4 and Figure A5 respectively. **Numerical results are available at ...** [9]**.** Generally, algorithms that employed the same operationalization of IADL deficit showed substantial pairwise concordance, however, algorithms that operationalized IADL deficit differently did not. Whereas among algorithms with identical IADL deficit operationalization, the agreement judged by Cohen's κ was moderately high (operationalization by FAQ total score: $\kappa = 0.75$, SD = 0.13; operationalization by FAQ item 9: $\kappa = 0.86$, SD = 0.09), among algorithms that differ in IADL deficit operationalization it was low ($\kappa = 0.34$, SD = 0.08).

[9] **JM: ... as some kind of Supplementary Table, ideally html but Excel file would work as well.**

[Insert Figure 2 here]

Prediction of Level II Criteria

For easier interpretability of our results, we next examine cases where Level II algorithms served as a reference and Level I algorithms as a predictor. Table 3 shows five Level I algorithms with the highest and five with the lowest accuracy in predicting Level II classification of probable PDD.

When IADL deficit was defined by total FAQ score, the Level II estimate of PDD rate was 10.71%. All five Level I algorithms that approximated the Level II classification most accurately were MoCA-based and defined Executive Function deficit by Clock drawing rather than Lexical fluency test. On the other hand, two out of the five Level I algorithms with the lowest accuracy were MMSE-based, whereas the remaining three were MoCA-based and defined Executive Function deficit by Lexical fluency test.

When IADL deficit was defined by FAQ item 9 score, the Level II estimate of PDD rate was 3.57%. Overall, the difference between the most accurate and the least accurate Level I algorithms was lower than in the case of IADL deficit being defined by FAQ total score (see Table 3). The five most accurate algorithms were all MoCA-based, defined Executive Function deficit by Clock drawing (with threshold < 2) and in majority of cases defined Language deficit by Animal naming. Two out of the five Level I algorithms with the lowest accuracy were MMSE-based, whereas the remaining three were MoCA-based and defined Executive Function deficit by Clock drawing (with threshold < 3) and Language deficit by Abstraction.

Finally, if the predictors are sorted by their balanced accuracy (i.e., average of sensitivity and specificity) instead of raw accuracy, the results are similar with the exception that for prediction of Level II with total FAQ score algorithm for probable PDD, the highest balanced accuracy was achieved by the sMoCA algorithm with 0.95 sensitivity and 0.93 specificity (see Table A4).

[Insert Table 3 here]

Discussion

This study evaluated the applicability and validity of diagnostic frameworks for diagnosing probable PDD within a cohort of patients considered for deep brain stimulation

(DBS). Our results demonstrate that diagnostic outcomes are markedly influenced by the chosen type of operationalization, particularly in relation to the assessment of the cognitive domains and IADLs.

Variability in Prevalence Estimates

One of the key findings of this study was the broad range of estimated PDD rate depending on the operationalization strategy. As seen in Table 3, the rate of PDD presence estimates varied from 2.00% to 16.75%, with the highest being derived from a combination of the sMoCA and the total FAQ score combination. When only FAQ item 9 was used to determine IADL deficits, the rate estimates were significantly lower (M = 3.09%, SD = 0.48), underscoring significant influence of functional assessment choice on diagnostic outcomes.

These findings are lower in comparison with previous research demonstrating wide variability in reported PDD rate among PD patients. For instance, retrospective study reported a PDD rate of 19.7% (Rana et al., 2011), while a clinical investigation by Aarsland et al. found even higher rate of around 30% (**?**). A recent complex meta-analysis synthesizing global data placed the pooled rate at 26.30% (Sousa et al., 2022). Compared to these estimates, our study reports generally lower prevalence rates, likely reflecting differences in sample characteristics, diagnostic criteria, and methodology.

Notably, our cohort consisted of patients being evaluated for DBS, a procedure typically reserved for individuals with relatively preserved cognitive function, which inherently biases against higher PDD prevalence rate. These comparisons emphasize that the observed lower prevalence in our sample is likely attributable to the preselection of cognitively intact individuals for DBS consideration, as well as to methodological variations such as the use of brief screening tool and differing IADL operationalizations.[10]

Cognitive and Clinical Context [11]

The relatively low prevalence of PDD across operationalizations may also reflect the relatively preserved cognitive profile of the DBS cohort, as evident in Table 2. The mean MoCA (M = 24.07, SD = 3.48) and MMSE (M = 26.69, SD = 2.22) scores suggest that on

[10] **Keep it?**

[11] **Keep this section?**

average the patients were functioning at a globally intact level. This likely reflects the pre-selection of cognitively preserved individuals for DBS, in line with standard eligibility criteria (**?**).

Memory performance, such as RAVLT and BVMTR delayed recall scores, also pointed to only mild deficits, particularly in domains central to the PDD vs. PD-MCI distinction. This aligns with previous findings that Level I criteria may overestimate dementia in patients with subtle impairments unless operational thresholds are rigorously defined (Aarsland et al., 2021; Bezdicek et al., 2016).

Implications for DBS Eligibility and Clinical Practice

Our findings also bear direct implications for clinical decision-making, especially in the context of surgical candidacy for DBS. Given that PDD remains a contraindication for DBS, the observed diagnostic instability could result in disparate treatment decisions based solely on which cognitive criteria are employed. Thus, harmonization of assessment procedures and operational cutoffs is essential to ensure equitable access to surgical treatment while maintaining diagnostic rigor.

Further, our data support the use of multidimensional assessment strategies, such as Level II neuropsychological batteries or regression-based normative comparisons, as confirmatory tools in diagnostically ambiguous cases. These approaches may enhance diagnostic precision and mitigate the risks of both false positives and unjustified treatment exclusion.

Constraints on generalizability

This study's generalizability is limited by its retrospective design and the homogeneity of the DBS candidate cohort, which may not reflect broader PD populations with varying cognitive profiles. Additionally, reliance on Czech normative data may restrict international applicability, although it represents real-world clinical standards within the region (**?**).

Future Directions

Future research should focus on prospective validation of updated Level I criteria against longitudinal functional outcomes and neurobiological markers. Additionally, the development of adaptive diagnostic algorithms that can integrate performance across cognitive, functional, and psychiatric domains may enhance diagnostic sensitivity while reducing the risk of over-classification.

Conclusions

This study systematically investigated the application of multiple Level I diagnostic criteria for Parkinson's disease dementia (PDD) within a cohort of patients considered for deep brain stimulation (DBS). The findings reveal substantial variability in prevalence estimates, strongly influenced by the choice of cognitive screening instruments and the operationalization of functional impairment. The divergence observed across operationalizations demonstrates the sensitivity of diagnostic outcomes to seemingly minor methodological choices.

The proposed revisions to the diagnostic framework (Kulisevsky et al., 2024) criteria offer enhanced sensitivity by leveraging MoCA-based components and broader IADL assessments, the use must be cautiously calibrated to prevent over-diagnosis in populations with mild or borderline cognitive deficits. Conversely, overly conservative criteria, such as reliance on pill questionnaire (i.e. FAQ item 9 equivalence) may fail to detect meaningful functional decline and thus under-identify true cases of probable PDD.

Ultimately, this study contributes to the improving landscape of PDD diagnostics by offering empirical evidence for the refinement of Level I criteria and reinforcing the value of psychometric rigor in clinical neuropsychology. Future work should extend this validation to longitudinal trajectories and integrate neurobiological correlates, ensuring that cognitive criteria remain both scientifically grounded and clinically actionable.

References

Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers*, *7*(1). https://doi.org/10.1038/s41572-021-00280-3

Beck, A. T., Steer, R. A., & Brown, G. (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*. Psychological Assessment Resources.

Benton, A. L., Hamsher, K. D., & Sivan, A. B. (1989). *Multilingual aphasia examination*.

AJA Associates.

Benton, A. L., Varney, N. R., & Hamsher, K. (1983). Visuospatial judgment: A clinical test. *Archives of Neurology*, *40*(3), 429–432.

Bezdicek, O., M., Č., Moore, T. M., Georgi, H. S., Sulc, Z., Wolk, D. A., Weintraub, D. A., Moberg, P. J., Jech, R., Kopecek, M., & Roalf, D. R. (2020). Determining a short form Montreal Cognitive Assessment (s-MoCA) Czech version: Validity in mild cognitive impairment Parkinson's disease and cross-cultural comparison. *Assessment*, *27*(8), 1960–1970. https://doi.org/10.1177/1073191118778896

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., Martinec Novakova, L., & Kopecek, M. (2016). Toward the processing speed theory of activities of daily living in healthy aging: Normative data of the functional activities questionnaire. *Aging Clinical and Experimental Research*, *28*, 239–247.

Bezdicek, O., Sulc, Z., Nikolai, T., Stepankova, H., Kopecek, M., Jech, R., & Růžička, 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

Bezdíček, O., Georgi, H., Nikolai, T., & Kopeček, M. (2021). *Pražská verze stroopova testu*. Karolinum. https://karolinum.cz/en/books/bezdicek-prazska-verze-stroopova-testu-25158

Bezdíček, O., Lukavský, J., & Preiss, M. (2011). Functional activities questionnaire, czech version – a validation study. *Česká a Slovenská Neurologie a Neurochirurgie*, *74*(107), 36–42. https://www.csnn.eu/casopisy/ceska-slovenska-neurologie/2011-1/validizacni-studie-ceske-verze-dotazniku-faq-34140

Cereda, E., Cilia, R., Klersy, C., Siri, C., Pozzi, B., Reali, E., Colombo, A., Zecchinelli, A. L., Mariani, C. B., Tesei, S., Canesi, M., Sacilotto, G., Meucci, N., Zini, M., Isaias, I. U., Barichella, M., Cassani, E., Goldwurm, S., & Pezzoli, G. (2016). Dementia in parkinson's disease: Is male gender a risk factor? *Parkinsonism & Related Disorders*, *26*, 67–72. https://doi.org/10.1016/j.parkreldis.2016.02.024

Ciharova, M., Cígler, H., Dostálová, V., Šivicová, 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

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A Meta-Analysis of Interrater and Internal Consistency Reliability of Selection Interviews. *Journal of Applied Psychology*, *80*(5), 565. https://doi.org/10.1037/0021-9010.80.5.565

Dubois, B., Burn, D., Goetz, C., Aarsland, D., Brown, R. G., Broe, G. A., Dickson, D., Duyckaerts, C., Cummings, J., Gauthier, S., Korczyn, A., Lees, A., Levy, R., Litvan, I., Mizuno, Y., McKeith, I. G., Olanow, C. W., Poewe, W., Sampaio, C., … Emre, M. (2007). Diagnostic procedures for Parkinson's disease dementia: Recommendations from the movement disorder society task force. *Movement Disorders*, *22*(16), 2314–2324. https://doi.org/10.1002/mds.21844

Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., … Dubois, B. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders*, *22*(12), 1689–1707. https://doi.org/10.1002/mds.21507

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Frydrychová, Z., Kopeček, M., Bezdíček, O., & Štěpánková, J. (2018). České normy pro revidovaný reyův auditorně-verbální test učení (RAVLT) pro populaci starších osob. *Československá Psychologie*, *62*(4), 330–349. https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.02da6923-6394-4d8d-b018-dd6e3b2503be

Havlík, F., Mana, J., Dušek, P., Jech, R., Ržička, E., Kopeček, M., Georgi, H., & Bezdicek, O. (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.

Iannone, R., Cheng, J., Schloerke, B., Hughes, E., Lauer, A., Seo, J., Brevoort, K., & Roy, O. (2024). *Gt: Easily create presentation-ready display tables*. https://CRAN.R-project.org/package=gt

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston naming test*. Lea & Febiger.

Kopecek, M., Stepankova, H., Lukavsky, J., Ripova, D., Nikolai, T., & Bezdicek, O. (2016). 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

Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kulisevsky, J., Litvan, I., Weintraub, D., Goldman, J. G., Tröster, A. I., Lewis, S. J., Parkinson, I., Group, M. D. S. P.-M. S., et al. (2024). A call for change: Updating the operational definition for dementia in parkinson's disease. *Movement Disorders Clinical Practice*.

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

Meireles, J., & Massano, J. (2012). Cognitive impairment and dementia in parkinsons disease: Clinical features, diagnosis, and management. *Frontiers in Neurology*, *3*. https://doi.org/10.3389/fneur.2012.00088

Mullner, J., Ruisl, I., & Farkas, G. (1980). *Dotaznik na meranie uzkosti a uzkostlivosti - STAI*. Bratislava: Psychodiagnostické a didaktické testy.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Oh, Y.-S., Kim, J.-S., Park, I.-S., Shim, Y.-S., Song, I.-U., Park, J.-W., Lee, P.-H., Lyoo, C.-H., Ahn, T.-B., Ma, H.-I., et al. (2016). Prevalence and treatment pattern of p arkinson's disease dementia in k orea. *Geriatrics & Gerontology International*, *16*(2), 230–236.

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of Functional Activities in Older Adults in the Community. *Journal of Gerontology*, *37*(3), 323–329. https://doi.org/10.1093/geronj/37.3.323

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, *30*(12), 1591–1601. https://doi.org/10.1002/mds.26424

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rana, A. Q., Yousuf, M. S., Naz, S., & Qa'aty, N. (2011). Prevalence and relation of dementia to various factors in parkinson's disease. *Psychiatry and Clinical Neurosciences*, *65*(4), 317–321. https://doi.org/10.1111/j.1440-1819.2011.02291.x

Reitan, R. (2004). The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology*, *19*(2), 281–288. https://doi.org/10.1016/s0887-6177(03)00042-8

Rey, A. (1964). *L'examen clinique en psychologie (the clinical psychological examination)*. Presses Universitaires de France.

Royall, D. R., Cordes, J. A., & Polk, M. (1998). CLOX: an executive clock drawing task. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*(5), 588–594. https://doi.org/10.1136/jnnp.64.5.588

Sousa, C. S. e, Alarcão, J., Martins, I. P., & Ferreira, J. J. (2022). Frequency of dementia in parkinson's disease: A systematic review and meta-analysis. *Journal of the Neurological*

Sciences, *432*, 120077.

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.

Stepankova, H., Nikolai, T., Lukavsky, J., Bezdicek, O., Vrajova, M., & Kopecek, M. (2015). Mini-Mental State Examination – česká normativní studie. *Ceska a slovenska neurologie a neurochirurgie*, *78*(111), 57–63.

Wechsler, D. (1997). *Wechsler adult intelligence scale—third edition (WAIS-III)*. Psychological Corporation. https://books.google.cz/books/about/Wais_III_Wechsler_Adult_Intelligence_Sca.html?id=qTCuGQAACAAJ

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

William Revelle. (2024). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Yamashita, K. Y., Bhoopatiraju, S., & Silverglate, G. T., Bret D. aGrossberg. (2023). Biomarkers in Parkinson's disease: A state of the art review. *Biomarkers in Neuropsychiatry*, *9*, 100074. https://doi.org/10.1016/j.bionps.2023.100074

Zemanová, N., Bezdíček, O., Michalec, J., Nikolai, T., Roth, J., Jech, R., & Ržička, E. (2016). Validační studie české verze bostonského testu pojmenování. *Česká a Slovenská Neurologie a Neurochirurgie*, *79*(112), 3. https://doi.org/10.14735/amcsnn2016307

Table 1

Summary of probable PDD operationalizations compared in the study.

Type	Global functioning	Attention	Executive Function
MMSE-based	MMSE < 26	Sevens backwards < 4	Clock drawing < 2 OR Lexical fluency (S) <
MoCA-based	MoCA < 27	Sevens backwards < 3	Clock drawing < {2, 3} OR Lexical fluency (K
sMoca-based	sMoCA < 13	-	-
Level II	-	TMT A & WAIS DSB	CF A & PST C

[a] The visual memory was evaluated based on WMS-III Family Pictures or BVMTR depending on which test was used in the assessment. This lead to no missing values because each patient underwent assessment via one of these tests.

Note. MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; sMoCA: short version of the MoCA; TMT A: Trail Making Test, Part A; WAIS DSB: Wechsler Adult Intelligence Scale Digit Span, Backwards; CF A: Categorical Verbal Fluency, Animals; PST C: Prague Stroop Test, Colours; WAIS Similarities: Wechsler Adult Intelligence Scale, Similarities; BNT 60: Boston Naming Test; RAVLT delayed recall: Rey Auditory Verbal Learning Test, Delayed Recall; BVMTR delayed recall: Brief Verbal Memory Test, Delayed Recall; WMS-III Family Pictures: Wechsler Memory Scale Family Pictures; JoLO: Boston Judgement of Line Orientation; CLOX: Clock Drawing Test. The OR operator implies that exactly one of the criteria listed is utilized within a single operationalization; the & operator implies that both criteria are used at the same time within a single operationalization; each threshold value within the set brackets { } was used to define probable PDD once in combination with all the other criteria on the same row.

Table 2

Cognitive characteristics of the sample.

	N	Md	Min-max	M	SD
MMSE					
Total score (Range 0-30)	203	27	15-30	26.69	2.22
Sevens (Range 0-5)	1/2/8/20/34/139	-	-	-	-
VF S (Number of Words per Minute)	202	15	1-34	14.95	5.80
Clock Drawing (Range 0-2)	26/91/86	-	-	-	-
Pentagons (Range 0-1)	187 (92%)	-	-	-	-
Three words (Range 0-3)	5/14/60/124	-	-	-	-
MoCA					
Total score (Range 0-30)	203	24	9-30	24.07	3.48
sMoCA total score (Range 0-16)	203	11	1-16	11.26	2.74
Sevens (Range 0-3)	1/2/29/171	-	-	-	-
VF K (Number of Words per Minute)	204	16	0-29	15.50	5.34
Clock drawing (Range 0-3)	24/83/96	-	-	-	-
Cube drawing (Range 0-1)	164 (81%)	-	-	-	-
Five words (Range 0-5)	69/19/29/39/22/25	-	-	-	-
Animal naming (Range 0-3)	10/193	-	-	-	-
Abstraction (Range 0-2)	7/72/124	-	-	-	-
Affect					
BDI (Range 0-63)	203	10	0-34	10.79	7.02
STAI X1 (Range 0-80)	188	39	20-72	38.96	8.93
STAI X2 (Range 0-80)	186	40	22-63	40.38	7.77
IADL					
FAQ (Range 0-30)	203	2	0-25	4.05	4.89
FAQ 9 (Range 0-1)	144/48/10/1	-	-	-	-
Screening					
DRS-II (Range0-144)	202	139	115-144	138.00	5.26

Table 3

Level I algorithms for probable PDD as predictors of Level II classification as the reference.

Level II (1)^a						Level II (2)^b				
Predictor	κ	Accuracy	p	Sensitivity	Specificity	Predictor	κ	Accuracy	p	Ser
			Top five							
MoCA (3)	0.64	0.94	.010	0.52	0.99	MoCA (12)	0.85	0.99	.028	
MoCA (51)	0.70	0.94	.010	0.71	0.97	MoCA (16)	0.85	0.99	.028	
MoCA (11)	0.63	0.94	.019	0.57	0.98	MoCA (2)	0.83	0.99	.028	
MoCA (13)	0.68	0.94	.019	0.71	0.97	MoCA (20)	0.85	0.99	.028	
MoCA (15)	0.63	0.94	.019	0.57	0.98	MoCA (4)	0.83	0.99	.028	
			Bottom five							
MMSE (1)	0.53	0.93	.061	0.43	0.99	MoCA (50)	0.70	0.98	.168	
MoCA (25)	0.57	0.92	.098	0.57	0.97	MoCA (54)	0.70	0.98	.168	
MoCA (35)	0.53	0.92	.098	0.48	0.98	MoCA (58)	0.70	0.98	.168	
MMSE (3)	0.41	0.92	.143	0.30	0.99	MMSE (2)	0.66	0.98	.168	
MoCA (39)	0.51	0.92	.148	0.48	0.97	MMSE (4)	0.53	0.97	.296	

[a]IADL deficite was defined as FAQ (total score) > 7

[b]IADL deficite was defined as FAQ (item 9) > 1

κ: Cohen's κ; p: p-value associated with a one-sided Exact Binomial Test comparing the Accuracy to the No Information Rate; The table shows five most accurate (Top five) and five least accurate (Bottom five) Level I algrithms for Parkinson's Disease Dementia (PDD) in predicting Level II classficiation of PDD. The algorithms were grouped by their definition of the deficit in Instrumental Activities of Daily Living (IADLs). The items comprising each listed algorithm can be found in Table A1.

Figure 1

Summary of the estimates of probable PDD rate across operationalizations under consideration. Vertical lines represent estimates arrived at by using sMoCA (dotted) or Level II (dashed) with FAQ item 9 (orange) or FAQ total score (blue) as criteria for probable PDD.

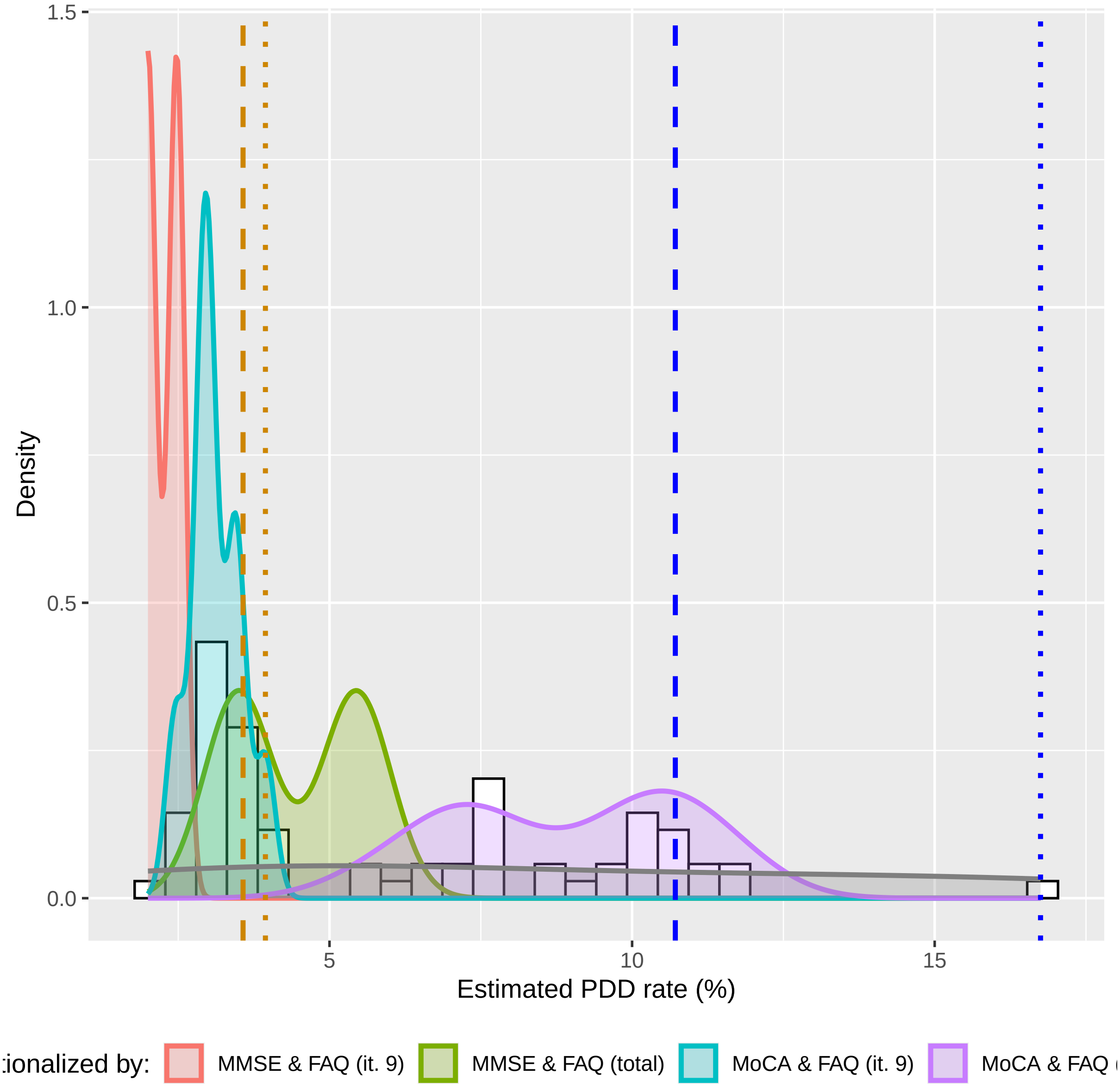
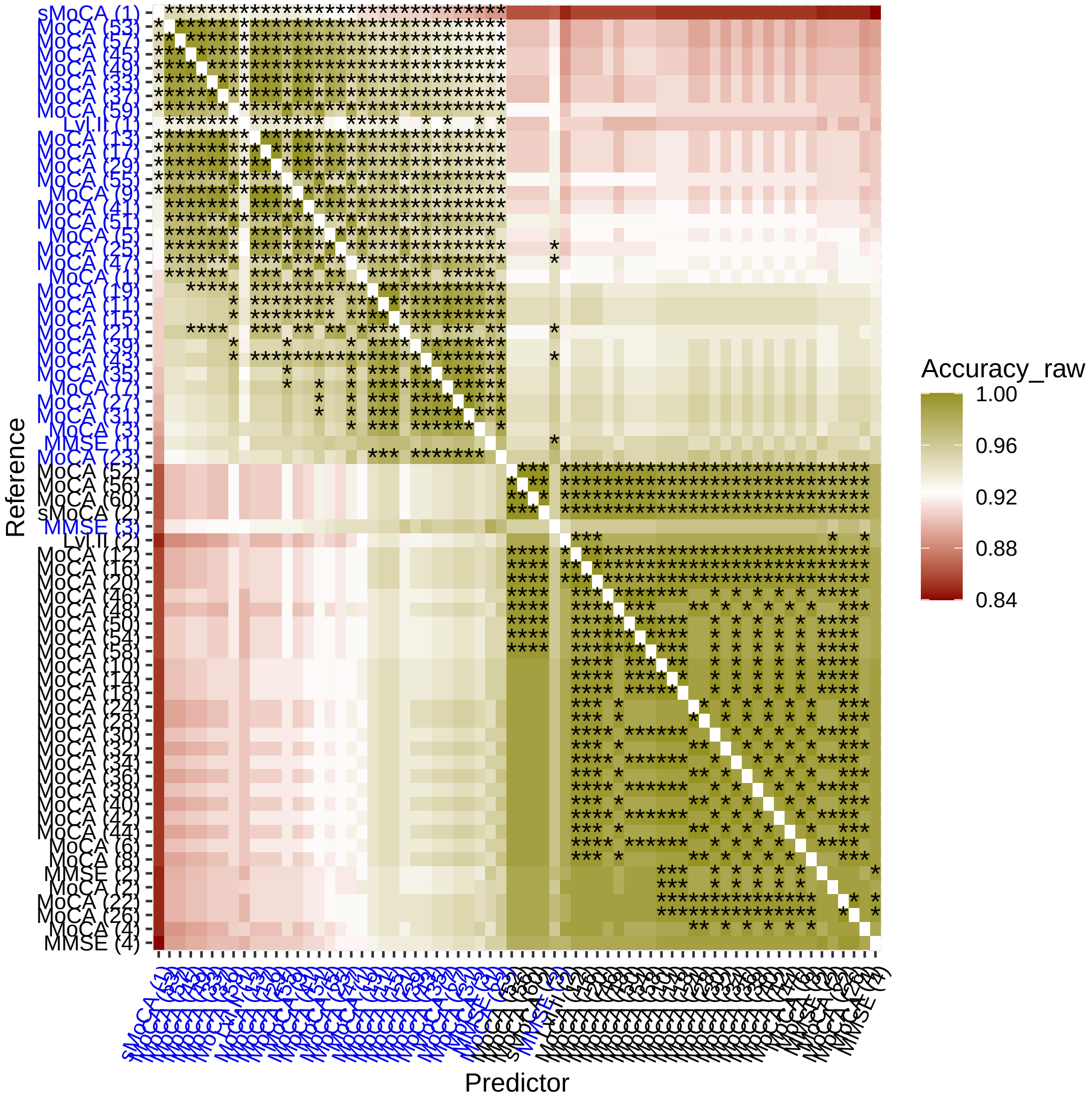


Figure 2

Prediction accuracy matrix. The matrix depict accuracy of



Appendix

Derivation of the Algorithms Set

Both, the original PDD criteria (Dubois et al., 2007) and the call for their change (Kulisevsky et al., 2024) allow for several distinct combinations of items to be used to define cognitive impairment. Consequently, in this study we derived all algorithms for probable PDD on Level I that are in line with published criteria. This procedure parallel the diagnostic algorithm outlined in Table 2 of Dubois et al. (2007). Specifically, in this study, we varied the exact specification of items 3-5 of this table (i.e., the measure of global cognitive impairment, the measure of the impact on IADLs and the measures of impaired cognition).

For each set of criteria (MMSE-based, MoCA-based, sMoCA-based and Level II), we first specified the items and then the thresholds for each item used to define probable PDD. If more than one option was present in either the choice of the item or the choice of the threshold, we created an algorithm for each choice in turn. The final set of algorithms was arrived at by computing the Cartesian product of all possibilities provided by varying items and thresholds. All combinations are presented in Table A1.

For MMSE-based algorithms, the following sets of items served as the basis:

$Global = \{MMSE < 26\}$

$Attention = \{Sevens\ backwards < 4\}$

$Executive = \{Clock\ drawing < 2, Lexical\ fluency\ (S) < 10\}$

$Construction = \{Pentagons < 1\}$

$Memory = \{Three\text{-}words\ recall < 3\}$

$IADL = \{FAQ > 7, FAQ\ (it.9) > 1\}$

The ensuing Cartesian product $Global \times Attention \times Executive \times Construction \times Memory \times IADL$ results in $1 \times 1 \times 2 \times 1 \times 1 \times 2 = 4$ MMSE-based algorithms for probable PDD.

For MoCA-based algorithms, the following sets of items served as the basis:

$Global = \{MoCA < 26\}$

$Attention = \{Sevens\ backwards < 3\}$

$Executive = \{Clock\ drawing < 2, Clock\ drawing < 3, Lexical\ fluency\ (K) < 11\}$

$Construction = \{Cube\ drawing < 1\}$

$Memory = \{Five\text{-}words\ recall < 1, Five\text{-}words\ recall < 2, Five\text{-}words\ recall < 3, Five\text{-}words\ recall < 4, Five\text{-}words\ recall < 5\}$

$Language = \{Abstraction < 2, Animal\ naming < 3\}$

$IADL = \{FAQ > 7, FAQ\ (it.9) > 1\}$

Note that the additional language domain adds complexity to establishing a diagnostic algorithm because simply by adding it to the set of items, the number of potential algorithms doubles. Further complexity is added by the fact that there are so far no guidelines for selecting a diagnostic threshold for Clock drawing and Five-words recall tests, both of which differ from their counterparts used by Dubois et al. (2007). Finally, although the Sevens backwards item has different thresholds in MoCA-based compared to MMSE-based algorithms, this difference is solely due to a difference in scoring whereby 3 points in MoCA correspond to 4 or 5 points in MMSE. The Seven backwards item threshold for MoCA-based algorithms used in this study is thus equivalent to its MMSE-based counterpart.

Computing the Cartesian product $Global \times Attention \times Executive \times Construction \times Memory \times Language \times IADL$ yields $1 \times 1 \times 3 \times 1 \times 5 \times 2 \times 2 = 60$ distinct MoCA-based algorithms for probable PDD.

For sMoCA-based algorithms, the following sets of items served as the basis:

$Global = \{sMoCA < 13\}$

$IADL = \{FAQ > 7, FAQ\ (it.9) > 1\}$

yielding $Global \times IADL$, i.e., $1 \times 2 = 2$ distinct sMoCA-based algorithms for probable PDD.

Finally, the Level II algorithms were based on the following sets of items:

$Attention = \{z(TMT\ A) < -1.5\ \cup z(WAIS\ DSB) < -1.5\}$

$Executive = \{z(CF\ A) < -1.5\ \cup z(PST\ C) < -1.5\}$

$Construction = \{z(JoLO) < -1.5\ \cup z(CLOXI) < -1.5\}$

$Memory = \{z(RAVLT\ DR) < -1.5\ \cup z(BVMTR\ DR) <$

$-1.5 \cup z(\mathit{WMS\text{-}III\ Family\ Pictures}) < -1.5\}$

$\mathit{Language} = \{z(\mathit{WAIS\ Similarities}) < -1.5 \cup z(\mathit{BNT}\ 60) < -1.5\}$

$\mathit{IADL} = \{\mathit{FAQ} > 7, \mathit{FAQ}\ (\mathit{it}.9) > 1\}$

where $z()$ denotes calculation of age, sex and education adjusted z-score. This yields $1 \times 1 \times 1 \times 1 \times 1 \times 2 = 2$ distinct Level II algorithms for probable PDD in the current study. All but the BNT 60 item were evaluated using regression norms published by Bezdicek et al. (2017). Since the original article used BNT 30 instead of BNT 60, we approximated the deficit in BNT 60 by comparing patients' raw score to age- and education-specific normative values reported by Zemanová et al. (2016). Specifically, patients whose BNT 60 score fell below 5$^{\text{th}}$ percentile of their demographic group in Table 6 of Zemanová et al. (2016) were considered to show signs of impaired performance.

Operationalization of Impaired Cognition

In the original criteria, item 4 of Level I criteria, i.e., impaired cognition, was defined as follows: *"The proposed diagnostic criteria require a profile of cognitive deficits, typical of those described for PD-D, in two or more of four domains."* (Dubois et al., 2007, p. 2316) Consequently, we defined impaired cognition as a deficit in two or more domains of four in MMSE-based criteria and as a deficit in two or more of five domains in MoCA-based criteria. sMoCA-based criteria omitted the "impaired cognition" item altogether because they were intended as a shorter screening alternative to classical Level I assessment. Finally, for the Level II criteria, we employed standard definition of impaired cognition as the *"[i]mpairment on at least two neuropsychological tests, represented by either two impaired tests in one cognitive domain or one impaired test in two different cognitive domains."* (Litvan et al., 2012, Table 1)

Theoretical and Empirical Estimands

In this study, we follow the framework proposed by Lundberg et al. (2021) for specifying targets of inference (i.e., the estimands) in qunatitative sciences to increase transparency and connect statistical evidence to relevant theory. Table A2 contains verbal description of the components relating to each of our proclaimed research objectives and map them to the population quantity of interest (the theoretical estimand), data-dependent quantity that could be estimated (the empirical estimand) and quantities that are reported in the study (statistical estimates).

The RO1 - to estimate the prevalence of PDD and evaluate the diagnostic variability and concordance across different algorithms of probable PDD - was divided into four distinct research objectives:

- to estimate the rate of PDD within PD (RO1.1),
- to estimated variability of this rate (RO1.2),
- to evaluate predictive value of demographic variables for probable PDD classification (RO1.3) and
- to evaluate concordance between different probable PDD operationalizations and criteria (RO1.4).

Estimates relating to RO1.1 and RO1.3 cannot be safely generalized beyond a population of PD patients that are candidates for DBS due to the systematic differences between DBS candidates pool and general PD population (such as the lower age of DBS candidates compared to the general PD population). On the other hand, the estimates relating to RO1.4 (and to a lesser degree to RO1.2[12]) may not be substantially influenced by the sample at hand as the primary source of their variance might come from variability in measures employed (e.g., MMSE vs MoCA to assess global cognitive performance) rather than variability in patients' performance. Assuming that there is no substantial Differential Item Functioning for DBS candidates compared to a broader population of patients with PD, the estimates relating to RO1.4 can be cautiously generalized beyond the current sample.

[12] Because the quantity of interest is a rate and could thus be though of as a sum of binomially distributed PDD occurences divided by the total number of patients, its variance will likely systematically vary with its mean. Specifically, as the rate goes from extremes to 0.5, the variance increases. Consequently, if our estimate of the rate was lower than the true population rate, e.g., because our sample includes younger patients compared to the general PD population, our estimate of variance would also lower than the true variance of PDD rate in the general PD population. Nonetheless, the between-algorithm variability may not be affected by this phenomenon as unlike variability of PDD rate, we do not have reason to assume it comes about by summing independent binomial events

Finally, for the RO2, the theoretical estimand is defined as the set of diagnostic components whose variation systematically alters the probability of a probable PDD diagnosis. This aspect of the study is exploratory in nature. Empirically, we assess the contribution of each diagnostic feature by examining how variations in operational definitions (e.g., domain-specific thresholds, criteria for functional impairment) influence the statistical estimates derived for the first objective. This allows us to identify the diagnostic elements most responsible for between-algorithm discrepancies.

Supplementary Presentation of Results

Figure A1

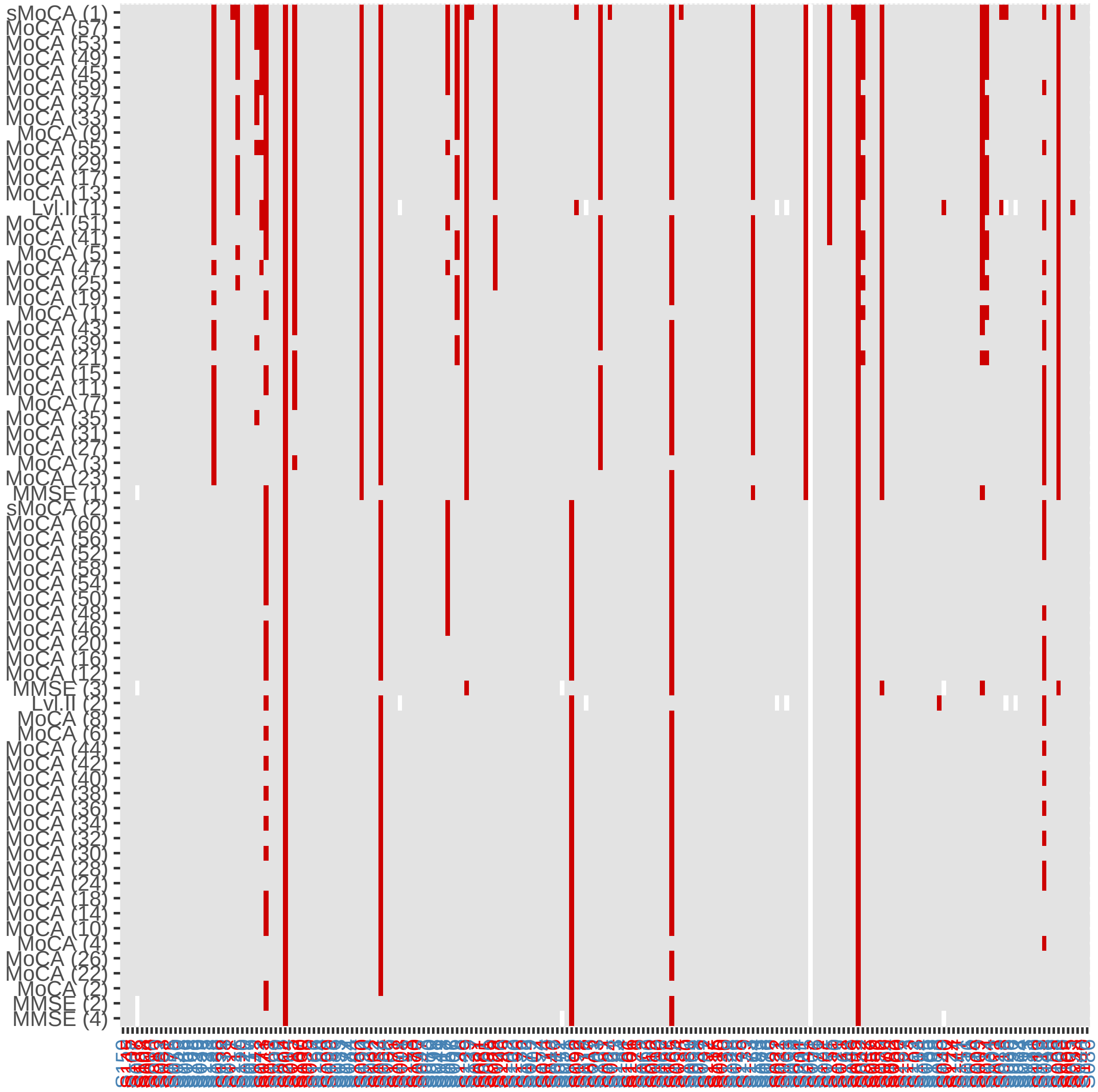

sMoCA (1)
MoCA (57)
MoCA (53)
MoCA (49)
MoCA (45)
MoCA (59)
MoCA (37)
MoCA (33)
MoCA (9)
MoCA (55)
MoCA (29)
MoCA (17)
MoCA (13)
Lvl.II (1)
MoCA (51)
MoCA (41)
MoCA (5)
MoCA (47)
MoCA (25)
MoCA (19)
MoCA (1)
MoCA (43)
MoCA (39)
MoCA (21)
MoCA (15)
MoCA (11)
MoCA (7)
MoCA (35)
MoCA (31)
MoCA (27)
MoCA (3)
MoCA (23)
MMSE (1)
sMoCA (2)
MoCA (60)
MoCA (56)
MoCA (52)
MoCA (58)
MoCA (54)
MoCA (50)
MoCA (48)
MoCA (46)
MoCA (20)
MoCA (16)
MoCA (12)
MMSE (3)
Lvl.II (2)
MoCA (8)
MoCA (6)
MoCA (44)
MoCA (42)
MoCA (40)
MoCA (38)
MoCA (36)
MoCA (34)
MoCA (32)
MoCA (30)
MoCA (28)
MoCA (24)
MoCA (18)
MoCA (14)
MoCA (10)
MoCA (4)
MoCA (26)
MoCA (22)
MoCA (2)
MMSE (2)
MMSE (4)

Table A1

Summary of all algortihms for probable PDD used in the study.

Algorithm	Global deficit	Attention	Executive function
Lvl.II (1)	-	TMT A < -1.5 OR WAIS DS < -1.5	CF A < -1.5 OR PST C < -1.5
Lvl.II (2)	-	TMT A < -1.5 OR WAIS DS < -1.5	CF A < -1.5 OR PST C < -1.5
MMSE (1)	Total score < 26	Sevens < 4	Clock Drawing < 2
MMSE (2)	Total score < 26	Sevens < 4	Clock Drawing < 2
MMSE (3)	Total score < 26	Sevens < 4	VF S < 10
MMSE (4)	Total score < 26	Sevens < 4	VF S < 10
MoCA (1)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (10)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (11)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (12)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (13)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (14)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (15)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (16)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (17)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (18)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (19)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (2)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (20)	Total score < 26	Sevens < 3	Clock drawing < 2
MoCA (21)	Total score < 26	Sevens < 3	VF K < 11
MoCA (22)	Total score < 26	Sevens < 3	VF K < 11
MoCA (23)	Total score < 26	Sevens < 3	VF K < 11
MoCA (24)	Total score < 26	Sevens < 3	VF K < 11
MoCA (25)	Total score < 26	Sevens < 3	VF K < 11
MoCA (26)	Total score < 26	Sevens < 3	VF K < 11
MoCA (27)	Total score < 26	Sevens < 3	VF K < 11
MoCA (28)	Total score < 26	Sevens < 3	VF K < 11

Table A2

Mapping between research objectives and quantities of interest in the current study.

To estimate the rate of PDD within PD. To assess variability of PDD diagnosis depending on the algorithm applied. To evaluate predictive information provided by demographic variables for probable PDD diagnosis. To evaluate the diagnostic concordance between different PDD algorithms within and between PDD criteria.
To identify algorithms' components that contribute to variability in probable PDD diagnosis within and acros

Table A3

Estimates of the rate of probable PDD in the sample.

Algorithm	N	Rate
sMoCA (1)	203	34 (16.75%)
MoCA (53)	203	24 (11.82%)
MoCA (57)	203	24 (11.82%)
MoCA (45)	203	23 (11.33%)
MoCA (49)	203	23 (11.33%)
MoCA (33)	203	22 (10.84%)
MoCA (37)	203	22 (10.84%)
MoCA (59)	203	22 (10.84%)
Lvl.II (1)	196	21 (10.71%)
MoCA (13)	203	21 (10.34%)
MoCA (17)	203	21 (10.34%)
MoCA (29)	203	21 (10.34%)
MoCA (55)	203	21 (10.34%)
MoCA (9)	203	21 (10.34%)
MoCA (41)	203	20 (9.85%)
MoCA (51)	203	20 (9.85%)
MoCA (5)	203	19 (9.36%)
MoCA (25)	203	18 (8.87%)
MoCA (47)	203	18 (8.87%)
MoCA (1)	203	16 (7.88%)
MoCA (19)	203	16 (7.88%)
MoCA (11)	203	15 (7.39%)
MoCA (15)	203	15 (7.39%)
MoCA (21)	203	15 (7.39%)
MoCA (39)	203	15 (7.39%)
MoCA (43)	203	15 (7.39%)
MoCA (35)	203	14 (6.90%)
MoCA (7)	203	14 (6.90%)

Table A4

Level I algorithms for probable PDD as predictors of Level II classification as the reference arranged by their balanced accuracy score.

Level II (1)[a]						Level II (2)[b]				
Predictor	κ	Accuracy	p	Sensitivity	Specificity	Predictor	κ	Accuracy	p	Ser
Top five										
sMoCA (1)	0.70	0.93	.061	0.95	0.93	MoCA (12)	0.85	0.99	.028	
MoCA (45)	0.69	0.94	.019	0.76	0.96	MoCA (16)	0.85	0.99	.028	
MoCA (49)	0.69	0.94	.019	0.76	0.96	MoCA (20)	0.85	0.99	.028	
MoCA (53)	0.67	0.93	.035	0.76	0.95	MoCA (52)	0.79	0.98	.078	
MoCA (57)	0.67	0.93	.035	0.76	0.95	MoCA (56)	0.79	0.98	.078	
Bottom five										
MoCA (31)	0.55	0.93	.061	0.48	0.98	MoCA (58)	0.70	0.98	.168	
MoCA (35)	0.53	0.92	.098	0.48	0.98	MoCA (22)	0.66	0.98	.168	
MoCA (39)	0.51	0.92	.148	0.48	0.97	MoCA (26)	0.66	0.98	.168	
MMSE (1)	0.53	0.93	.061	0.43	0.99	MMSE (2)	0.66	0.98	.168	
MMSE (3)	0.41	0.92	.143	0.30	0.99	MMSE (4)	0.53	0.97	.296	

[a] IADL deficite was defined as FAQ (total score) > 7

[b] IADL deficite was defined as FAQ (item 9) > 1

κ: Cohen's κ; p: p-value associated with a one-sided Exact Binomial Test comparing the Accuracy to the No Information Rate; The table shows five most accurate (Top five) and five least accurate (Bottom five) Level I algrithms for Parkinson's Disease Dementia (PDD) in predicting Level II classficiation of PDD. The algorithms were grouped by their definition of the deficit in Instrumental Activities of Daily Living (IADLs). The items comprising each listed algorithm can be found in Table A1.

Figure A2

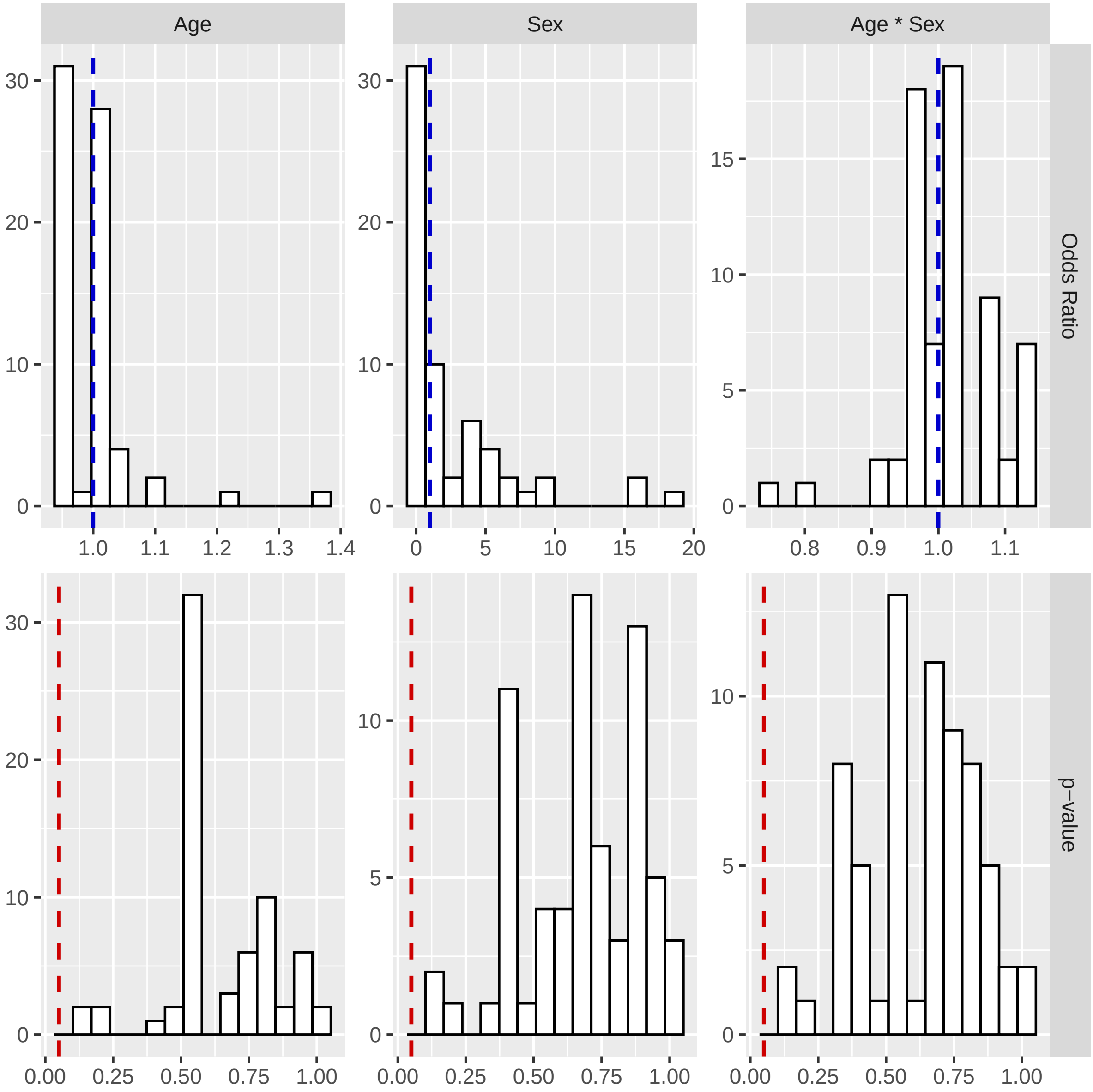

Age
Sex
Age * Sex
Odds Ratio
p−value
0
10
20
30
1.0
1.1
1.2
1.3
1.4
0
5
10
15
20
0.8
0.9
1.0
1.1
0.00
0.25
0.50
0.75
1.00

Figure A3

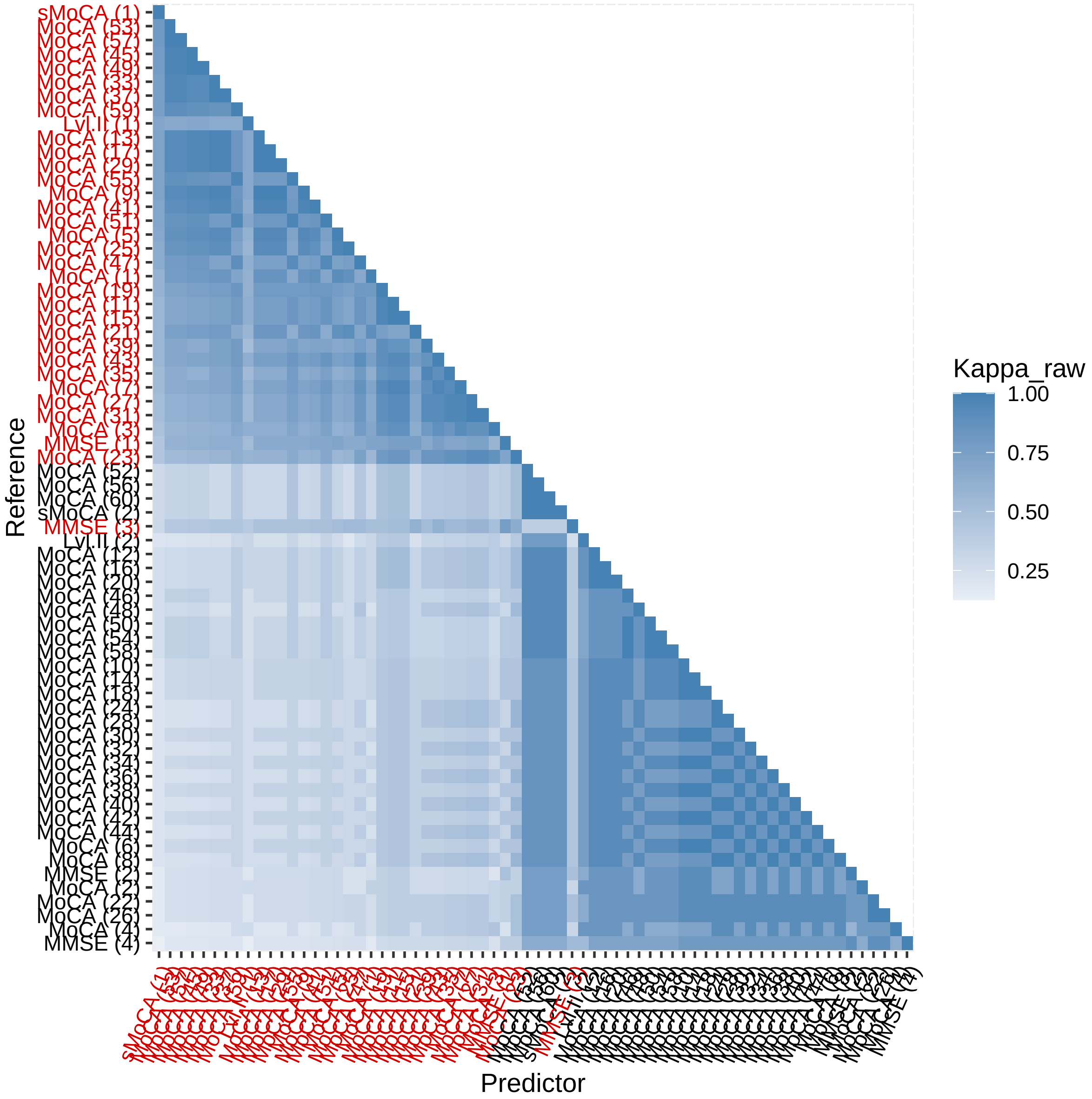

sMoCA (1)
MoCA (53)
MoCA (57)
MoCA (45)
MoCA (49)
MoCA (33)
MoCA (37)
MoCA (59)
Lvl.II (1)
MoCA (13)
MoCA (17)
MoCA (29)
MoCA (55)
MoCA (9)
MoCA (41)
MoCA (51)
MoCA (5)
MoCA (25)
MoCA (47)
MoCA (1)
MoCA (19)
MoCA (11)
MoCA (15)
MoCA (21)
MoCA (39)
MoCA (43)
MoCA (35)
MoCA (7)
MoCA (27)
MoCA (31)
MoCA (3)
MMSE (1)
MoCA (23)
MoCA (52)
MoCA (56)
MoCA (60)
sMoCA (2)
MMSE (3)
Lvl.II (2)
MoCA (12)
MoCA (16)
MoCA (20)
MoCA (46)
MoCA (48)
MoCA (50)
MoCA (54)
MoCA (58)
MoCA (10)
MoCA (14)
MoCA (18)
MoCA (24)
MoCA (28)
MoCA (30)
MoCA (32)
MoCA (34)
MoCA (36)
MoCA (38)
MoCA (40)
MoCA (42)
MoCA (44)
MoCA (6)
MoCA (8)
MMSE (2)
MoCA (2)
MoCA (22)
MoCA (26)
MoCA (4)
MMSE (4)
Reference
Predictor
Kappa_raw
1.00
0.75
0.50
0.25

Figure A4

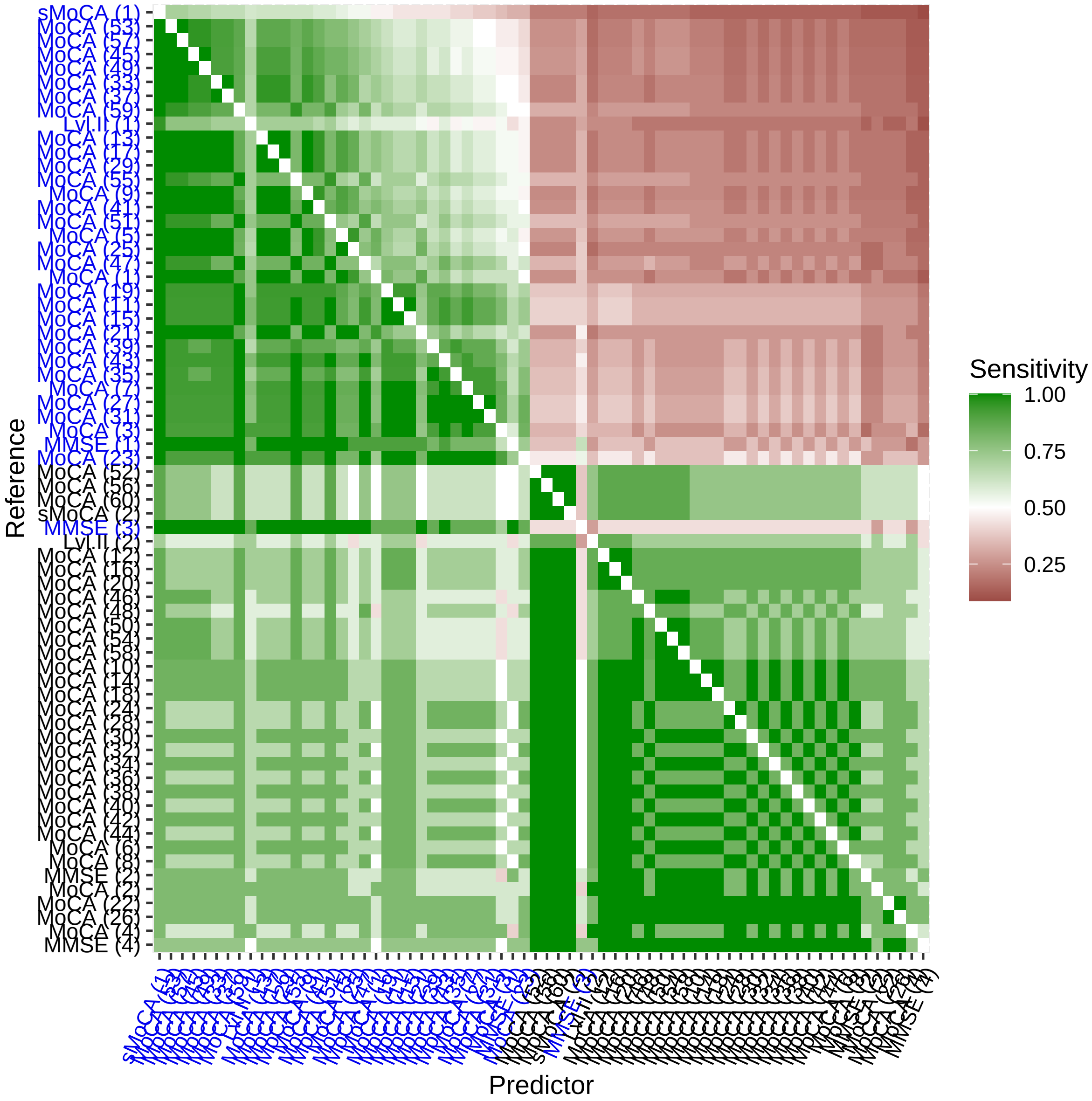
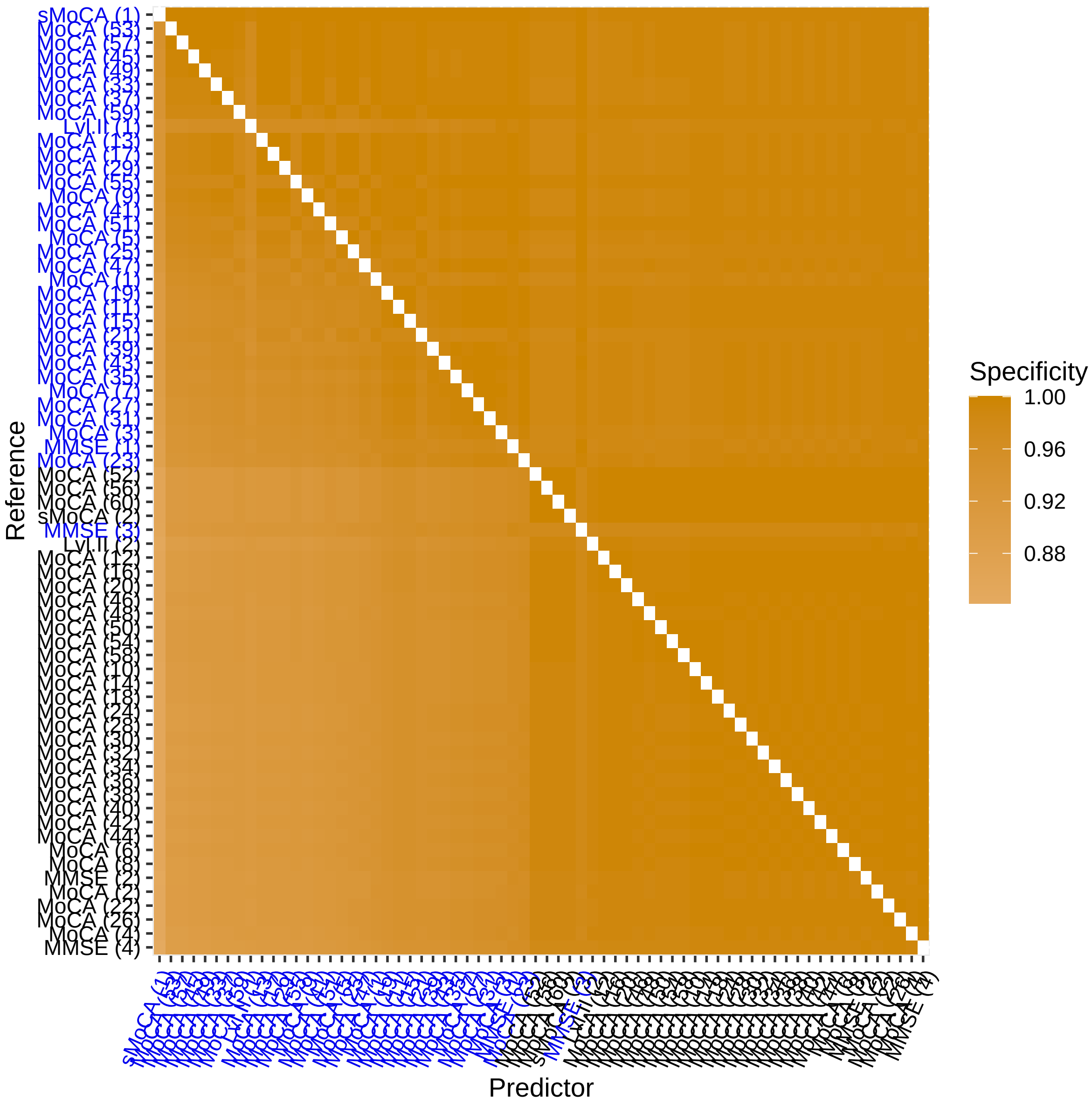

sMoCA (1)
MoCA (53)
MoCA (57)
MoCA (45)
MoCA (49)
MoCA (33)
MoCA (37)
MoCA (59)
Lvl.II (1)
MoCA (13)
MoCA (17)
MoCA (29)
MoCA (55)
MoCA (9)
MoCA (41)
MoCA (51)
MoCA (5)
MoCA (25)
MoCA (47)
MoCA (1)
MoCA (19)
MoCA (11)
MoCA (15)
MoCA (21)
MoCA (39)
MoCA (43)
MoCA (35)
MoCA (7)
MoCA (27)
MoCA (31)
MoCA (3)
MMSE (1)
MoCA (23)
MoCA (52)
MoCA (56)
MoCA (60)
sMoCA (2)
MMSE (3)
Lvl.II (2)
MoCA (12)
MoCA (16)
MoCA (20)
MoCA (46)
MoCA (48)
MoCA (50)
MoCA (54)
MoCA (58)
MoCA (10)
MoCA (14)
MoCA (18)
MoCA (24)
MoCA (28)
MoCA (30)
MoCA (32)
MoCA (34)
MoCA (36)
MoCA (38)
MoCA (40)
MoCA (42)
MoCA (44)
MoCA (6)
MoCA (8)
MMSE (2)
MoCA (2)
MoCA (22)
MoCA (26)
MoCA (4)
MMSE (4)
Reference
Predictor
Sensitivity
1.00
0.75
0.50
0.25

Figure A5


Reference
sMoCA (1)
MoCA (53)
MoCA (57)
MoCA (45)
MoCA (49)
MoCA (33)
MoCA (37)
MoCA (59)
Lvl.II (1)
MoCA (13)
MoCA (17)
MoCA (29)
MoCA (55)
MoCA (9)
MoCA (41)
MoCA (51)
MoCA (5)
MoCA (25)
MoCA (47)
MoCA (1)
MoCA (19)
MoCA (11)
MoCA (15)
MoCA (21)
MoCA (39)
MoCA (43)
MoCA (35)
MoCA (7)
MoCA (27)
MoCA (31)
MoCA (3)
MMSE (1)
MoCA (23)
MoCA (52)
MoCA (56)
MoCA (60)
sMoCA (2)
MMSE (3)
Lvl.II (2)
MoCA (12)
MoCA (16)
MoCA (20)
MoCA (46)
MoCA (48)
MoCA (50)
MoCA (54)
MoCA (58)
MoCA (10)
MoCA (14)
MoCA (18)
MoCA (24)
MoCA (28)
MoCA (30)
MoCA (32)
MoCA (34)
MoCA (36)
MoCA (38)
MoCA (40)
MoCA (42)
MoCA (44)
MoCA (6)
MoCA (8)
MMSE (2)
MoCA (2)
MoCA (22)
MoCA (26)
MoCA (4)
MMSE (4)
Predictor
Specificity
1.00
0.96
0.92
0.88