Predicting level-2 cognitive outcomes and research clinic diagnosis of MCI and Dementia in Parkinson's Disease from the MoCA: Guidance on the selection of optimal MoCA cutoffs

Tanja Zerenner[1] Sanjay G Manohar[2] Michael Lawton[1] Jamil Razzaque[2]
Falah Al Hajraf[2] Karolien Groenewald[2] Ludo van Hillegondsberg[2]
Tamir Eisenstein[2] Johannes Klein[2] Yoav Ben-Shlomo[1] Michele T Hu[2]

[1]Department of Population Health Sciences, University of Bristol, Bristol, BS8 2PN, UK
[2]Nuffield Department of Clinical Neurosciences, University of Oxford, OX3 9DU, UK

June 30, 2025

Background and Objectives The Montreal Cognitive Assessment (MOCA) is frequently used in cohort studies into Parkinson's disease (PD) as it is a simple and inexpensive tool for assessing the cognitive state of patients. However, cut-off values for distinguishing between normal cognition, mild cognitive impairment (MCI) and dementia differ across the literature. We comprehensively evaluate the accuracy of the MoCA for stratifying patients and investigate whether stratification can be improved by including additional routinely collected information.

Methods We use longitudinal data from PD and healthy control participants of the PPMI cohort which – in addition to the MoCA – conducts detailed level-2 cognitive testing and records diagnoses of MCI and dementia made at the research clinics. The ability of the MoCA to predict level-2 outcomes and diagnoses was assessed using the AUC. Multilevel logistic regression was used to predict higher-order outcomes from MoCA scores in conjunction with other routinely collected information on basic demographics, functional impairment (MDS-UPDRS 1.1) and/or MoCA subscores. Model performance was compared using the AUC. Optimal cut-offs for patient stratification w.r.t. Youden's J, equal proportions, screening and diagnosis were derived.

Results Data from 1094 PD patients and of 267 healthy control participants was analyzed. MoCA total scores predicted level-2 cognitive impairment in PD (2 impaired domains or more) with an AUC of 0.86 (95% CI 0.84–0.88). Youden's J was maximized at cutoff ≤ 24 with sensitivity 74.7 (70.5-79.3) and specificity 83.1 (82.0-84.2); cutoff ≤ 21 equated proportions. Prediction of level-2 cognitive outcomes was not substantially improved by additional covariates, but functional impairment/MDS-UPDRS 1.1 was found to be superior over the MoCA for predicting research clinic diagnosis of PD-MCI or PDD. MDS-UPDRS 1.1 ≥ 1 discriminated diagnosis of any impairment from no impairment with sensitivity

81.9 (79.7, 84.2) and specificity 71.5 (69.9, 73.0); a combination of MDS-UPDRS 1.1 $\geq$ 1 and MoCA $\leq$ 26 equated proportions. A combination of MDS-UPDRS 1.1 $\geq$ 2 and MoCA $\leq$ 20 equated proportions for discrimination of dementia from no dementia.

Discussion Optimal cutoffs for PD-MCI or PDD on the MoCA scale depend on their purpose. In research studies with no immediate effect on patients, we recommend considering a cutoff that equates proportions. Identifying suitable cutoffs from the literature - for research or in clinic use - needs to take into account the respective PD population.

1 Introduction

It is estimated, that 24 - 31% of people living with Parkinson's disease (PD) have dementia [1]. About 25% of PD patients with normal cognition at the time of PD diagnosis convert to mild cognitive impairment (MCI) within 3 years [2]. Dementia prevalence 8 years post PD diagnosis was estimated as 78.2% (95% CI 71.1-84.0) by [3]. A more recent study suggests lower dementia prevalence, estimating a probability of dementia of 9% 10 years after PD diagnosis in the PPMI cohort and 27% in the Penn state cohort, rising to 74% at 20 years in the latter [4].

An accurate assessment of the cognitive state of PD patients is crucial not only in clinic, but also for research purposes, for instance in studies identifying risk factors of cognitive decline and future dementia [5, 6], predictive modeling [7, 8], or biomarker research [9, 10]. While in an ideal scenario a patient's cognitive state is assessed with an extensive cognitive test battery, comprised of multiple tests for each cognitive domain, in practice elaborate cognitive testing is often not feasible and brief cognitive screening tools are utilized instead. One of the today most widely used tools is the Montreal Cognitive Assessment (MoCA), a 10 min test comprising 30 questions assessing executive and visuospatial function, naming, attention, language, abstraction, recall, and orientation [11]. It is widely acknowledged that the MoCA has not been specifically designed for the profile of cognitive impairment specific to PD and hence cutoffs optimal in the general population are not necessarily optimal in a PD population. Numerous studies have thus adressed the derivation of PD specific MoCA cutpoints for PD-MCI and/or PDD [12–24]. Optimal cutoffs reported by above studies lie between $\leq$ 17 to $\leq$ 26 for PD-MCI and $\leq$ 9 to $\leq$ 25 for PDD depending on purpose (e.g., screening vs diagnosis), optimality criteria (e.g., Youden's J, balance of sensitivity or specificity) or study population.

We evaluate the ability of the MoCA to predict outcomes from level-2 cognitive testing as well as research clinic diagnoses of PD-MCI or PDD from the MoCA using longitudinal data from over 1000 subjects diagnosed with PD. Furthermore, we investigate if predictions of higher-order cognitive outcomes and diagnoses can be improved by incorporating additional routinely collected information on basic demographics, MDS-UPDRS or by optimizing the weighting of MoCA subscores. Finally, we consider four different criteria for selecting optimal cutoffs for research and in-clinic use and report respective cutoffs for the best model/covariate combination for each reference outcome.

2 Data

We are utilizing data from the The Parkinson Progression Marker Initiative (PPMI) cohort study. In the following we describe the scales and variables relevant for our analysis. Details on recruitment, inclusion/exclusion criteria and follow-up are described elsewhere `https://www.ppmi-info.org` [25, 26].

Routine variables and assessments collected in the PPMI cohort include basic participant demographics, the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [27] and the MoCA [11]. The MDS-UPDRS predominantly assesses motor function in PD, but also includes a question (MDS-UPDRS 1, Question 1) rating functional impairment due to cognitive deficits as perceived by the patient and/or caregiver on a scale from 0 (no functional impairment) to 4 (severe functional impairment).

In the PPMI cohort level-2 cognitive testing of executive function, memory and visuospatial function has been conducted since study begin in 2010. More recently the battery of level-2 tests has been extended, but due to the comparably low number of complete observations for the new, extended battery, we restrict our analysis to the level-2 tests conducted since study begin: Memory has been assessed with the Hopkins Verbal Learning Test - Revised (HVLT-R)[28] of which we consider the scores for total recall and recognition discrimination. Executive function has been assessed in three distinct tests, animals fluency test [29], Symbol Digit Modalities Test (SDMT) [30, 31] and Letter-Number-Sequencing (LNS) [32]. Visuospatial functioning has been assessed with the 15-item (odd/even) version of the Benton Judgment of Line Orientation (BJL) [33].

3 Methods

3.1 Evaluating level-2 testing

Initially we attempted to evaluate the level-2 cognitive testing data utilizing normative ranges from existing literature, but we found the percentages of patient visits with test scores < 1.5 SD's below norm to vary greatly across the different tests when applying respective published norms. We therefore opted to utilize the level-2 data available for the controls arm of the PPMI cohort to derive our own age, sex and education adjusted normative ranges for LNS, SDMT, animals fluency, BJL, HVLT-recall and HVLT-discrimination using mixed-effects linear and tobit regressions. Details are provided in the Supplementary Material.

To quantify overall performance of level-2 testing for each cognitive domain, the age, sex and education adjusted z-scores of the individual tests are summed within each domain:

$$
\begin{aligned}
s_{executive} &= z_{lns} + z_{animals} \\
s_{memory} &= z_{recall} + z_{discrim} \\
s_{visuospatial} &= z_{BJL}
\end{aligned}
$$

We chose to not include the SDMT in the executive domain score $s_{executive}$ as it appears to be impacted by motor function more strongly than the other scores.

The domain level scores $s_{exeutive}$ and s_{memory} are no longer z-scored, but have a standard deviation

< 1 due to the positive correlation of the individual test scores. Instead of z-scoring a second time we opted to use the 6.5%-quantiles of the healthy controls domain scores as a cutoff for defining level-2 impairment in the respective domain (Supplementary Table 3). In a standard normal distribution the 6.5%-quantile corresponds to -1.5 SD's below expectation/norm. Hence our cut-off resembles one of the criteria routinely applied in the literature to define cognitive impairment from level-2 testing as -1.5 SD's below norm [34, 35].

3.2 Model outcomes and covariates

We consider a total of four different cognitive outcomes: (1) impairment in ≥ 1 domain according to level-2 testing, (2) impairment in ≥ 2 domains according to level-2 testing, (3) research clinic diagnosis of cognitive impairment (PD-MCI/PDD vs. PD-NC) and (4) research clinic diagnosis of dementia (PDD vs. PD-NC/PD-MCI). We first assess the discriminative ability of the MOCA total scores. We then proceed to study if discriminative ability can be improved by utilizing MoCA subscores instead of totals, by incorporating routinely collected information on basic demographics (age, gender and years of education) or by incorporating functional impairment as assessed in MDS-UPDRS 1 Question 1. The MoCA subscores we consider are based on the categorization of questions on the MoCA instrument,i.e., visuospatial/executive, naming, attention, language, abstraction, delayed recall, and orientation [11, 36]. The MDS-UPDRS 1.1 responses are included as a factor variable with categories 3 (moderate) and 4 (severe) merged into a single factor due to 4 being rarely recorded (only 10 times in a total 4700 patient visits, 0.2%, Table 1). For the research clinic diagnosis outcomes we further study the discriminative ability of the level-2 testing domain scores, i.e., $s_{exeutive}$, s_{memory} and $s_{visuospatial}$ in comparison to the MoCA.

3.3 Modeling approach

As we use longitudinal data and therefore non independent observations, we employ a mixed-effect logistic regression with a random intercept for each subject. As we aim to derive models for estimating outcome probabilities of 'new' patients for whom no random effect was estimated, we need to obtain and evaluate marginal predictions. Due to the non-linear logit link function of the logistic regression, we cannot simply applying the conditional model with the random effect set to zero [37, 38]. Marginal effects $\beta_{M,m}$, however, can be estimated from the respective conditional effects $\beta_{RE,m}$ using a cumulative Gaussian approximation to the logistic function [38, 39]. Details are provided in the Supplementary Material. One can then proceed to estimate the probability of an outcome for a 'new' patient utilizing the marginal effects $\beta_{RE,m}$, like in a standard, i.e., non-multilevel, logistic regression as

$$P(Y = 1|X = x) = \frac{1}{1 + \exp(-\beta_{RE,0} - \sum_{m=1}^{p} x_m \beta_{RE,m})}. \tag{1}$$

Models were fitted using R's *lme4* package [40] with the *nlminbwrap* optimizer. ROC curves were computed and plotted using *pROC* [41]. Calibration curves were plotted with the *CalibrationCurves* package. Following recommendations for internal validation of logistic regression prediction models [42], we performed a non-parametric bootstrap. Bootstrap estimation of bias/model optimism allows for opti-

mism correction of the AUC estimates. The root-mean-square-error (RMSE) over the bootstrap sample provides a measure of the uncertainty of the estimated AUC. We performed $m = 200$ bootstrap iterations, bootstrapping at subject level. Equations for bias and RMSE are provided in the Supplementary Material.

3.4 Cutoffs

Optimality of a cutoff depends on their purpose. We therefore evaluate 4 distinct cutoffs. The first cutoff is the value optimizing Youden's J, that is the sum of sensitivity and specificity minus one, a widely used statistical optimality criterion [43]. The second cutoff we consider is selected such that the percentages of observations below cutoff equals the percentages of observations with the respective outcome. We hence refer to this cutoff as the "equal proportions" cutoff. While this is not a standard approach to the selection of cutoffs, it is - from a medical literaute perspective - equivalent to using an equipercentile method for equating of scales [44]. At the equal proportions cutoff sensitivity equals positive predictive value (PPV) and specificity negative predictive value (NPV). For screening or diagnosis of patients in clinic, that is use of cutoffs with direct implications for patients, benefits and potential harms to patients, and potentially clinical capacity, in the situation at hand need to be taken into account. We therefore provide a range of cutoff values alongside their respective statistical properties in the Supplementary Material to allow for individual informed choices. In the main manuscript we derive candidate cutoffs for screening and diagnosis following [12, 13, 18, 23] by selecting the lowest cutoff with sensitivity and NPV ≥ 80 as a screening cutoff and the largest cutoff with specificity and PPV ≥ 80 as a diagnostic cutoff.

4 Results

4.1 Demographics and cognitive outcomes

Basic demographics and cognitive test scores for Controls and PDs are summarized in Supplemental Table 1. Detailed information on the level-2 normative ranges derived from the PPMI Controls data is provided in the Supplementary Material. All six scores showed an association with age unlikely due to chance ($p < 0.001$) with scores decreasing with increased age. Five out of six scores were associated with education, with subjects with longer education achieving higher scores on average ($p \leq 0.08$). Only the performance in the HVLT-discrimination test did not show any indications of education dependence. Females achieved higher scores in the memory domain ($p < 0.001$), males in the visuospatial domain ($p < 0.001$). Differences between sexes in the executive function domain are less pronounced with some indication of females scoring higher in SDMT ($p = 0.03$) and animals fluency test ($p = 0.09$). The BLJ scores were associated with form version ($p < 0.001$) with higher scores achieved on the odd version.

In 881 PD patient visits (18.7%) level-2 testing indicated impairment in one domain, in 297 (6.3%) visits in two domains, and in 91 visits (1.9%) in all three domains tested. PD-MCI was diagnosed at at 920 (20%) patient visits, PDD at 104 (2.3%) visits (Table 1). As expected, MoCA scores are decreased and functional impairment increased with less favorable cognitive outcomes. However, while almost half of the PDD cases (47.1%) were recorded to experience moderate or severe functional impairment, only 19.8% percent of cases with l2-impairment in all three tested domains were (Table 1).

4.2 Predicting cognitive outcomes

4.2.1 Level-2 impairment

The education adjusted MoCA total scores discriminate patients with ≥ 1 impaired domain from those with no impairment according to level-2 testing with an AUC of 0.76; those with ≥ 2 impaired domains from those with one or none with an AUC of 0.86 (Table 2, Figure 2A,B). Incorporating demographics did not improve discrimination. MDS-UPDRS 1.1 or using MoCA subscores instead of the total improves discrimination slightly (Table 2, Figure 3A,B). The latter indicates that the standard subscore weighting is slightly suboptimal for predicting level-2 impairment. Except for naming and abstraction, however, all MoCA items were predictive ($p < 0.001$, Supplemental Tables 7,8). Also, all MDS-UPDRS 1.1 factor levels were predictive ($p < 0.001$) indicating that already slight functional impairment is associated with an increased probability of unfavorable performance on level-2 testing.

Internal bootstrap validation shows no indication of overfitting for any of the models with optimism in terms of AUC estimated to be ≤ 0.002 for the majority of models and ≤ 0.005 for the most complex ones with the greatest number of covariates (Table 2). The models are overall reasonably well calibrated, but do exhibit some tendency towards too moderate estimations (calibration slopes > 1, Supplemental Tables 12, 11, Supplemental Figure 2).

Table 3 summarizes selected cutoffs for patient stratification. As no model leads to a substantial improvement over the MoCA - the maximum increase in AUC is ≤ 0.01 - we consider only cutoffso the MoCA scale (Figure 3A,B). Youden's J is maximized at a MoCA cutpoint of ≤ 26 for stratifying ≥ 1 impaired domain; ≤ 24 for stratifying ≥ 2 impaired domains. Equal proportions cutoffs are ≤ 26 and ≤ 21, respectively, with sensitivities of 57.4% and 49.7%, meaning when equating proportions about half of cases with the respective outcome are correctly identified. Screening cutoffs are higher; diagnostic cutoffs lower.

4.2.2 Research clinic diagnoses

The education adjusted MoCA total scores discriminate patients with research clinic diagnosis of any cognitive impairment (PD-MCI/PDD) from those with normal cognition with an AUC of 0.74; those with research clinic diagnosis of dementia from those with no dementia diagnosis with an AUC of 0.90 (Table 2, Figure 2C,D). Incorporating demographic covariates improves discrimination of cognitive impairment ($p = 0.001$) with age being the most relevant demographic covariate. Of two patients with the same MoCA total score, the older patient is more likely diagnosed with cognitive impairment at a research clinic visit ($p < 0.001$, Supplementary Table 9). Though, at the same MoCA total, older age is also associated with an increased probability of being diagnosed with dementia ($p = 0.001$, Supplementary Table 10), in our data set overall discrimination between dementia and no dementia is not improved by incorporating demographic covariates into the model. Also utilizing MoCA subscores instead of the totals does not improve discrimination. Interestingly, however MoCA subscores, namely visospatial, attention and orientation, are the more strongly associated with dementia diagnosis ($p < 0.001$) than the other subscores ($p \geq 0.15$).

The greatest improvement in discrimination is achieved by adding the MDS-UPDRS 1.1 as a model co-

variate. This holds for both, discriminating any cognitive impairment from normal cognition (AUC=0.86) as well as discriminating dementia from no dementia (AUC=0.96). Our data even suggests, that the MDS-UPDRS 1.1 on it's own is a superior predictor of research clinic cognitive diagnoses than the MoCA totals (Table 2, Figure 3C,D). For predicting any impairment all factor levels are relevant ($OR_M \geq 5$, $p < 0.001$), for predicting dementia diagnosis only MDS-UDPRS 1.1 scores of 2 or greater are predictive ($OR_M > 5$, $p < 0.001$) while an MDS-UPDRS 1.1 sore of 1, that is slight functional impairment, exhibited an in comparison weak association with dementia diagnosis ($OR_M \approx 1.4$, $0.18 \leq p \leq 0.39$). Level 2 cognitive test results were not superior over the MoCA in predicting research clinic diagnosis.

Internal bootstrap validation shows no indication of overfitting for any of the models with optimism in terms of AUC estimated to be ≤ 0.002 for the majority of models and ≤ 0.004 for the more complex ones. Only for the more complex models utilizing the MoCA subscores for predicting dementia diagnosis optimism is estimated to be slightly larger (0.007 to 0.011) reflecting the less powerful events per covariate ratios of these models. The models predicting any cognitive impairment show reasonable calibration. The models predicting dementia diagnosis are not reasonably calibrated (Supplemental Tables 11,12 and Figure 2). In this case model estimated probabilities shall not be interpreted as such. The most likely reason is a nonlinear association between the logit of the probability of dementia and cognitive test scores and possibly other covariates. However, due to the low number of cases we did not explore this further, but opted to state respective cutoffs on the linear predictor scale as opposed to model estimated probabilities.

Youden's J is maximized at a MoCA cutpoint of ≤ 25 for discriminating any cognitive impairment from normal cognition, and ≤ 23 for discriminating dementia from no dementia (Table 3). Equal proportions cutpoints are ≤ 24 and ≤ 17, respectively with sensitivities of 47.0% and 32.7%. Screening cutoffs are, as expected, higher, the diagnostic cutoff for any cognitive impairment lower. For dementia no MoCA cutpoint meets the specified diagnostic requirements of $\geq 80\%$ Sensitivity and PPV. Incorporating functional impairment as recorded in the MDS-UPDRS 1.1 notably improves discrimination which is reflected by higher sensitivity and/or specificity compared to MoCA totals only (Table 3).

Also a simpler approach of combining a MoCA cutoff with a MDS-UPDRS 1.1 cutoff improves discrimination compared to a pure MoCA cutoff. For discriminating any cognitive impairment from none, a combination of MDS-UPDRS 1.1 ≥ 1 and MoCA ≤ 26 optimizes w.r.t. our equal-proportions criterion with a sensitivity of 57.2%, an increase of 10% compared to MoCA totals only. For discriminating dementia from no dementia, a combination of MDS-UPDRS 1.1 ≥ 2 and MoCA ≤ 20 optimizes w.r.t. the equal-proportions criterion with a sensitivity of 46.1%, an increase of about 13% compared MoCA totals only. If in favor of a slightly higher sensitivity one might alternatively choose the combination of MDS-UPDRS 1.1 ≥ 2 and MoCA ≤ 21, as the hypothetical optimum that truly equates proportions is estimated to lie in between.

5 Discussion

The MoCA predicts higher-order cognitive outcomes and diagnoses with AUCs between 0.73-0.90, a range commonly regarded "acceptable" at its lower end and "excellent" at is upper end [45]. Incorporating demographic covariates or using MoCA subscores for predicting level-2 cognitive outcomes leads to a

statistically significant, but small increase in AUC of ≤ 0.01 even for the most complex models. We therefore recommend applying a cutoffs on original MoCA scale over a more complex approach. For predicting cognitive research clinic diagnosis we recommend a combination of MoCA totals and the MDS-UPDRS 1.1, which increases the AUC by ≈ 0.06 compared to MoCA totals only. In fact, the MDS-UPDRS 1.1 by itself is a better predictor of cognitive diagnosis than the MoCA. We believe that this finding reflects the extent of dialogue between the clinician and patient or relative/caregiver when assessing functional impairment and its relevance to the clinician's overall diagnosis. We are aware of the circularity of this part of our analysis. Clinicians making a diagnosis following the PPMI protocol are asked to evaluate functional impairment and may review cognitive test scores including the MoCA. Our findings however do suggest that the assessment of functional impairment is the dominant factor in the research clinic diagnosis and therefore, in absence of a diagnosis, a highly valuable predictor of what a diagnosis would have been. Interestingly, functional impairment was not only predictive of dementia diagnosis, but also of MCI diagnosis. The graded assessment of functional impairment in the MDS-UPDRS 1.1 allows to capture the slight to mild functional impairments that clinicians associate with an overall diagnosis of MCI.

Our finding that the level-2 domain scores are not superior predictors of research clinic diagnosis compared to the MoCA totals may seem surprising at first. However, due to the circular nature of this part of our analysis this is no indication that level-2 testing of the in three in PD most relevant cognitive domains may be an inferior representation of a patient's cognitive state, but rather suggests that the performance on the MoCA plays an equally great role in the research clinic diagnosis. Reviewing performance on 5 (or more) tests is more time-consuming and might require re-consideration of the respective normative ranges while the interpretation of MoCA scores is comparably straightforward.

Two further statistical remarks are to be made. The finding of better discrimination / larger AUC for the more severe cognitive outcomes is, from a statistical perspective, not surprising. Reasonable cutoffs for more severe outcomes lie in less dense regions of the statistical distribution of the MoCA totals, which reduces mis-classification risk. Secondly, we are aware that in the medical literature, adjustments of p-values for multiple comparisons are often appropriate. For our study, we deliberately chose not to do so, as the different model versions utilize common covariates and are therefore not independent.

Which cutoffs to use depends on the respective purpose. To provide suggestions for screening and diagnostic cutoffs we have followed criteria widely applied across the literature [12, 13, 19, 23]. However, we'd like to emphasize that in any applications with immediate effect on patients, such as decisions on specialist referrals, counseling or treatments, situation specific decisions on suitable cutoffs need to be made taking into consideration benefits and potential harms to patients. The Supplementary Material therefore provides the discriminative properties of a wide range of cutoffs to inform individual decisions.

In pure research applications on the other hand, for instance when harmonizing data from distinct cohorts for incidence or prevalence studies, in studies identifying risk factors or developing prediction models for future cognitive decline or dementia, or for stratification of bio samples, we recommend considering the equal proportions cutoffs. It is selected such that the percentage of observations below cutoff equals the percentage of observations with the respective outcome, and holds thus an intuitive interpretation.

Given the variety of literature on PD specific MoCA cutoffs, we here only explicitly compare our

findings with the two studies with the greatest sample sizes known to us. [20] have analyzed data from 106 PD subjects with normal cognition, 280 with PD-MCI, and 84 with PDD and identified cutoffs by optimizing Youden's J. For discriminating PDD from PD without dementia they found the same cutoff of ≤ 23 as the present study. The optimal cutoff to discriminate PD-MCI of ≤ 26 from [20] is one point higher than in the present study. The difference in Youden's J between our optimal ≤ 25 cutoff and a cutoff of ≤ 26 is however small. So, all in all our findings are consistent with [20].

The study with the largest sample size we are aware of is by [23] who analyzed data from 700 PD subjects with normal cognition, 706 PD-MCI, 374 PDD subjects. They derive their screening and diagnostic cutoffs using the same criteria as the present study, but conclude throughout lower cutoff values - a screening cutoff of 17 for dementia for instance. The mean MoCA score in their PD-NC group was 24.5, 20.2 in the PD-MCI group and 13.4 in the PDD group, that is several points lower than in our study. Their reference, a clinical diagnosis, is based on the MDS diagnostic criteria for MCI and dementia [46, 47] including level-2 diagnostic testing of all 5 cognitive domains evaluated using respective Italian norms. While not fully equivalent in terms of study design, the magnitude of the differences between our findings and [23] is surprising. In the original publication of the MoCA a score of ≤ 25 was described to signify mild cognitive impairment. A mean MoCA score below that threshold in a group classified as normal cognition therefore appears counterintuitive at first glance. However, [48] found a strong education dependence of the MoCA scores and in particular, a more pronounced decline with age in those who those who spent less years in education. This suggests that potentially the difference in education levels between the study population of [23] and the PPMI cohort might explain some of the differences. However, [23] reported that just like in the regression models of the present study, demographic variables, including years of education and age, were not among the key features of their decision trees.

A diagnosis made by a clinician or the results higher-order cognitive testing are solid references to validate the MoCA against, but not a perfect representation of a patient's cognitive state neither. While diagnostic criteria are outlined in great detail in the literature [46, 47, 49], they do include passages which leave room for interpretation, such as "Impairment on neuropsychological tests may be demonstrated in several ways: performance between 1 to 2 standard deviations (SDs) below age, education, gender, and culturally appropriate norms, [...]" [47]. Already a choice of a cuttoff of either 1 or 1.5 SD's below norm will lead to systematically different categorizations of patients and thus lead to differences in MoCA cutoffs optimized to resemble such categorization. Further, as we have found working with different published norms to evaluate the level-2 cognitive testing in the PPMI cohort, classification further depends on the choice of normative data sets for each test. We thus believe that deriving our own normative ranges for all tests utilizing data from the controls arm of the PPMI cohort constitutes one of the strengths of our study. We evaluate all tests in a consistent manner using controls data from subjects matched to the PD participants in terms of age, gender and education levels. The main limitation of our study and in fact any of the studies deriving MoCA cutoffs is their uncertain level of generalizability. Therefore when screening literature for guidance on suitable MoCA cutoffs either for use in clinic or for research purposes, we recommend to not only evaluate sample sizes, but also compare the population one intends to apply the cutoffs to to the population from which optimal cutoffs were derived. With the amount of literature on PD specific MoCA cutoffs, we believe a systematic comparison and meta-analysis is warranted. Further

future work should address the generalizability of MDS-UPDRS 1.1 cutoffs in comparison to the MoCA across different study populations.

References

[1]D. Aarsland, J. Zaccai, and C. Brayne, "A systematic review of prevalence studies of dementia in parkinson's disease", Movement disorders: official journal of the Movement Disorder Society **20**, 1255–1263 (2005).

[2]D. Saredakis, L. E. Collins-Praino, D. S. Gutteridge, B. C. Stephan, and H. A. Keage, "Conversion to mci and dementia in parkinson's disease: a systematic review and meta-analysis", Parkinsonism & related disorders **65**, 20–31 (2019).

[3]D. Aarsland, K. Andersen, J. P. Larsen, and A. Lolk, "Prevalence and characteristics of dementia in parkinson disease: an 8-year prospective study", Archives of neurology **60**, 387–392 (2003).

[4]J. Gallagher, C. Gochanour, C. Caspell-Garcia, R. D. Dobkin, D. Aarsland, R. N. Alcalay, M. J. Barrett, L. Chahine, A. S. Chen-Plotkin, C. S. Coffey, et al., "Long-term dementia risk in parkinson disease", Neurology **103**, e209699 (2024).

[5]A. Schrag, U. F. Siddiqui, Z. Anastasiou, D. Weintraub, and J. M. Schott, "Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed parkinson's disease: a cohort study", The Lancet Neurology **16**, 66–75 (2017).

[6]L. L. Gibson, D. Weintraub, R. Lemmen, G. Perera, K. R. Chaudhuri, P. Svenningsson, and D. Aarsland, "Risk of dementia in parkinson's disease: a systematic review and meta-analysis", Movement Disorders **39**, 1697–1709 (2024).

[7]D. A. Morales, Y. Vives-Gilabert, B. Gómez-Ansón, E. Bengoetxea, P. Larrañaga, C. Bielza, J. Pagonabarraga, J. Kulisevsky, I. Corcuera-Solano, and M. Delfino, "Predicting dementia development in parkinson's disease using bayesian network classifiers", Psychiatry Research: NeuroImaging **213**, 92–98 (2013).

[8]T. Phongpreecha, B. Cholerton, I. F. Mata, C. P. Zabetian, K. L. Poston, N. Aghaeepour, L. Tian, J. F. Quinn, K. A. Chung, A. L. Hiller, et al., "Multivariate prediction of dementia in parkinson's disease", npj Parkinson's Disease **6**, 20 (2020).

[9]Y. Compta, J. B. Pereira, J. Ríos, N. Ibarretxe-Bilbao, C. Junqué, N. Bargalló, A. Cámara, M. Buongiorno, M. Fernández, C. Pont-Sunyer, et al., "Combined dementia-risk biomarkers in parkinson's disease: a prospective longitudinal study", Parkinsonism & related disorders **19**, 717–724 (2013).

[10]D. G. Coughlin and D. J. Irwin, "Fluid and biopsy based biomarkers in parkinson's disease", Neurotherapeutics **20**, 932–954 (2023).

[11]Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment", Journal of the American Geriatrics Society **53**, 695–699 (2005).

[12]S. Hoops, S. Nazem, A. Siderowf, J. Duda, S. Xie, M. Stern, and D. Weintraub, "Validity of the moca and mmse in the detection of mci and dementia in parkinson disease", Neurology **73**, 1738–1745 (2009).

[13]J. Dalrymple-Alford, M. MacAskill, C. Nakas, L. Livingston, C. Graham, G. Crucian, T. Melzer, J. Kirwan, R. Keenan, S. Wells, et al., "The moca: well-suited screen for cognitive impairment in parkinson disease", Neurology **75**, 1717–1725 (2010).

[14]C. Marras, M. J. Armstrong, C. A. Meaney, S. Fox, B. Rothberg, W. Reginold, D. F. Tang-Wai, D. Gill, P. J. Eslinger, C. Zadikoff, et al., "Measuring mild cognitive impairment in patients with parkinson's disease", Movement Disorders **28**, 626–633 (2013).

[15]R. Biundo, L. Weis, S. Facchini, P. Formento-Dojot, A. Vallelunga, M. Pilleri, and A. Antonini, "Cognitive profiling of parkinson disease patients with mild cognitive impairment and dementia", Parkinsonism & related disorders **20**, 394–399 (2014).

[16]N. Kandiah, A. Zhang, A. R. Cenina, W. L. Au, N. Nadkarni, and L. C. Tan, "Montreal cognitive assessment for the screening and prediction of cognitive decline in early parkinson's disease", Parkinsonism & related disorders **20**, 1145–1148 (2014).

[17]T. Lucza, K. Karádi, J. Kállai, R. Weintraut, J. Janszky, A. Makkos, S. Komoly, and N. Kovács, "Screening mild and major neurocognitive disorders in parkinson's disease", Behavioural neurology **2015**, 983606 (2015).

[18]A. Federico, A. Maier, G. Vianello, D. Mapelli, M. Trentin, G. Zanette, A. Picelli, M. Gandolfi, and S. Tamburin, "Screening for mild cognitive impairment in parkinson's disease: comparison of the italian versions of three neuropsychological tests", Parkinson's disease **2015**, 681976 (2015).

[19]A. Federico, M. Trentin, G. Zanette, D. Mapelli, A. Picelli, N. Smania, M. Tinazzi, and S. Tamburin, "Diagnosing mild cognitive impairment in parkinson's disease: which tests perform best in the italian population?", Neurological Sciences **38**, 1461–1468 (2017).

[20]H. M. Kim, C. Nazor, C. P. Zabetian, J. F. Quinn, K. A. Chung, A. L. Hiller, S.-C. Hu, J. B. Leverenz, T. J. Montine, K. L. Edwards, et al., "Prediction of cognitive progression in parkinson's disease using three cognitive screening measures", Clinical parkinsonism & related disorders **1**, 91–97 (2019).

[21]R.-L. Yu, W.-J. Lee, J.-Y. Li, Y.-Y. Chang, C.-C. Chen, J.-J. Lin, Y.-F. Sung, T.-K. Lin, and J.-L. Fuh, "Evaluating mild cognitive dysfunction in patients with parkinson's disease in clinical practice in taiwan", Scientific reports **10**, 1014 (2020).

[22]A. D'Iorio, E. N. Aiello, M. Amboni, C. Vitale, F. Verde, V. Silani, N. Ticozzi, A. Ciammola, B. Poletti, and G. Santangelo, "Validity and diagnostics of the italian version of the montreal cognitive assessment (moca) in non-demented parkinson's disease patients", Aging Clinical and Experimental Research **35**, 2157–2163 (2023).

[23]E. Fiorenzato, S. Cauzzo, L. Weis, M. Garon, F. Pistonesi, V. Cianci, M. L. Nasi, F. Vianello, A. L. Zecchinelli, G. Pezzoli, et al., "Optimal mmse and moca cutoffs for cognitive diagnoses in parkinson's disease: a data-driven decision tree model", Journal of the Neurological Sciences **466**, 123283 (2024).

[24]R. O'Caoimh, M. J. Foley, S. Timmons, and D. W. Molloy, "Screening for cognitive impairment in movement disorders: comparison of the montreal cognitive assessment and quick mild cognitive impairment screen in parkinson's disease and lewy body dementia", Journal of Alzheimer's Disease Reports **8**, 971–980 (2024).

[25]K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, et al., "The parkinson progression marker initiative (ppmi)", Progress in neurobiology **95**, 629–635 (2011).

[26]K. Marek, S. Chowdhury, A. Siderowf, S. Lasch, C. S. Coffey, C. Caspell-Garcia, T. Simuni, D. Jennings, C. M. Tanner, J. Q. Trojanowski, et al., "The parkinson's progression markers initiative (ppmi)–establishing a pd biomarker cohort", Annals of clinical and translational neurology **5**, 1460–1477 (2018).

[27]C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, et al., "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results", Movement disorders: official journal of the Movement Disorder Society **23**, 2129–2170 (2008).

[28]L. G. Ralph H.B. Benedict David Schretlen and J. Brandt, "Hopkins verbal learning test – revised: normative data and analysis of inter-form and test-retest reliability", The Clinical Neuropsychologist **12**, 43–55 (1998).

[29]A. Ardila, F. Ostrosky-Solís, and B. Bernal, "Cognitive testing toward the future: the example of semantic verbal fluency (animals)", International Journal of Psychology **41**, 324–332 (2006).

[30]A. Smith, "Symbol digit modalities test", The clinical neuropsychologist (1973).

[31]R. H. Benedict, J. DeLuca, G. Phillips, N. LaRocca, L. D. Hudson, R. Rudick, and M. S. O. A. Consortium, "Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis", Multiple Sclerosis Journal **23**, 721–733 (2017).

[32]D. Wechsler III, "Wms-iii administration and scoring manual. san antonio, tx: the psychological corporation", (1997).

[33]R. D. Venderploeg, L. V. LaLone, P. Greblo, and J. A. Schinka, "Odd-even short forms of the judgment of line orientation test", Applied Neuropsychology **4**, 244–246 (1997).

[34]D. Weintraub, T. Simuni, C. Caspell-Garcia, C. Coffey, S. Lasch, A. Siderowf, D. Aarsland, P. Barone, D. Burn, L. M. Chahine, et al., "Cognitive performance and neuropsychiatric symptoms in early, untreated parkinson's disease", Movement Disorders **30**, 919–927 (2015).

[35]K. Lugosi, M. A. Engh, Z. Huszár, P. Hegyi, P. Mátrai, G. Csukly, Z. Molnár, K. Horváth, D. Mátis, and Z. Mezei, "Domain-specific cognitive impairment in multiple sclerosis: a systematic review and meta-analysis", Annals of Clinical and Translational Neurology **11**, 564–576 (2024).

[36]S. Nazem, A. D. Siderowf, J. E. Duda, T. Ten Have, A. Colcher, S. S. Horn, P. J. Moberg, J. R. Wilkinson, H. I. Hurtig, M. B. Stern, et al., "Montreal cognitive assessment performance in patients with parkinson's disease with "normal" global cognition according to mini-mental state examination score", Journal of the American Geriatrics Society **57**, 304–308 (2009).

[37]A. Skrondal and S. Rabe-Hesketh, "Prediction in multilevel generalized linear models", Journal of the Royal Statistical Society Series A: Statistics in Society **172**, 659–687 (2009).

[38]M. Pavlou, G. Ambler, S. Seaman, and R. Z. Omar, "A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes", BMC Medical Research Methodology **15**, 1–6 (2015).

[39]S. L. Zeger, K.-Y. Liang, and P. S. Albert, "Models for longitudinal data: a generalized estimating equation approach", Biometrics, 1049–1060 (1988).

[40]D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4", Journal of statistical software **67**, 1–48 (2015).

[41]X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "Proc: an open-source package for r and s+ to analyze and compare roc curves", BMC bioinformatics **12**, 1–8 (2011).

[42]E. W. Steyerberg, F. E. Harrell Jr, G. J. Borsboom, M. Eijkemans, Y. Vergouwe, and J. D. F. Habbema, "Internal validation of predictive models: efficiency of some procedures for logistic regression analysis", Journal of clinical epidemiology **54**, 774–781 (2001).

[43]M. Hassanzad and K. Hajian-Tilaki, "Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in roc analysis: an update review", BMC Medical Research Methodology **24**, 84 (2024).

[44]M. J. Kolen and R. L. Brennan, *Test equating: methods and practices* (Springer Science & Business Media, 2013).

[45]D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression* (John Wiley & Sons, 2013).

[46]B. Dubois, D. Burn, C. Goetz, D. Aarsland, R. G. Brown, G. A. Broe, D. Dickson, C. Duyckaerts, J. Cummings, S. Gauthier, et al., "Diagnostic procedures for parkinson's disease dementia: recommendations from the movement disorder society task force", Movement disorders **22**, 2314–2324 (2007).

[47]I. Litvan, J. G. Goldman, A. I. Tröster, B. A. Schmand, D. Weintraub, R. C. Petersen, B. Mollenhauer, C. H. Adler, K. Marder, C. H. Williams-Gray, et al., "Diagnostic criteria for mild cognitive impairment in parkinson's disease: movement disorder society task force guidelines", Movement disorders **27**, 349–356 (2012).

[48]H. C. Rossetti, L. H. Lacritz, C. M. Cullum, and M. F. Weiner, "Normative data for the montreal cognitive assessment (moca) in a population-based sample", Neurology **77**, 1272–1275 (2011).

[49]M. Emre, D. Aarsland, R. Brown, D. J. Burn, C. Duyckaerts, Y. Mizuno, G. A. Broe, J. Cummings, D. W. Dickson, S. Gauthier, et al., "Clinical diagnostic criteria for dementia associated with parkinson's disease", Movement disorders: official journal of the Movement Disorder Society **22**, 1689–1707 (2007).

Table 1: Demographics and cognition in the PD arm of the PPMI cohort stratified by the number of cognitive domains impaired according to level-2 testing and by cognitive research clinic diagnosis. Listed are means and standard deviations or, where applicable, percentages and counts.

	#l2-impaired domains				**research clinic diagnosis**		
	0	1	2	3	PD-NC	PD-MCI	PDD
N*	3431 (73%)	881 (18.7%)	297 (6.3%)	91 (1.9%)	3545 (77.8%)	910 (20.0%)	104 (2.3%)
age	65.1 (9.4)	66.5 (10.2)	68.5 (10.4)	67.3 (10.9)	64.5 (9.5)	70.0 (8.5)	72.4 (9.2)
female	37.5% (1286)	40.2% (354)	35.7% (106)	38.6% (35)	40.1% (1420)	31.2% (284)	32.7% (34)
education years	15.9 (3.2)	15.3 (3.5)	14.8 (4.2)	14.3 (4.7)	15.8 (3.4)	15.5 (3.3)	14.7 (4.3)
PD duration [years]	5.4 (3.3)	5.4 (3.3)	6.2 (3.5)	6.5 (3.3)	5.3 (3.2)	6.6 (3.5)	7.6 (3.1)
MoCA total (adj)	27.4 (2.4)	25.3 (3.2)	22.4 (4.0)	18.0 (4.8)	27.1 (2.6)	24.6 (3.7)	19.4 (5.1)
MDS-UPDRS 1.1							
0 (none)	65.5% (2244)	50.3% (442)	37.4% (111)	15.4% (14)	71.5% (2532)	19.6% (178)	4.8% (5)
1 (slight)	27.7% (948)	34.1% (299)	34.7% (103)	39.6% (36)	25.9% (916)	47.0% (427)	10.6% (11)
2 (mild)	5.6% (192)	11.5% (101)	16.9% (50)	25.3% (23)	2.4% (85)	26.1% (237)	36.5% (38)
3 (moderate)	1.2% (40)	3.8% (33)	10.4% (31)	14.3% (13)	0.3% (10)	7.2% (65)	40.4% (42)
4 (severe)	0.0% (0)	0.3% (3)	0.7% (2)	5.5% (5)	0.0% (0)	0.2% (2)	7.7% (8)
PD-NC	85% (2829)	67.8% (576)	41.9% (121)	20.9% (19)	-	-	-
PD-MCI	14.5% (484)	29.7% (252)	44.6% (129)	49.5% (45)	-	-	-
PDD	0.5% (17)	2.5% (21)	13.5% (39)	29.7% (27)	-	-	-
#l2-impaired domains							
0	-	-	-	-	80.0% (2827)	53.3% (484)	16.4% (17)
1	-	-	-	-	16.3% (576)	27.6% (251)	20.2% (21)
2	-	-	-	-	3.5% (121)	14.2% (129)	37.5% (39)
3	-	-	-	-	0.5% (19)	4.9% (45)	26.0% (27)

* N $\widehat{=}$ number of patient visits; 1094 distinct patients; 1086 had a research clinic diagnosis at at least one visit

Table 2: Area under the ROC curve (AUC) of prediction models for the different cognitive outcomes utilizing different sets of covariates. Bias/Optimism correction and RMSE were estimated from $n = 200$ bootstrap iterations.

outcome	covariates	AUC (corrected)	Bias	RMSE	AUC (original)
$\geq$ 1 impaired domain (l2-testing)	MoCA total (adj)	0.757	-0.000	0.012	0.756
	MoCA total (raw), demographics	0.755	0.004	0.012	0.759
	MoCA subscores	0.763	0.001	0.011	0.764
	MoCA subscores, demographics	0.760	0.005	0.012	0.765
	MDS-UPDRS 1.1	0.622	-0.001	0.013	0.622
	MoCA total (adj), MDS-UPDRS 1.1	0.763	0.002	0.012	0.765
	MoCA total (raw), demogr., MDS-UPDRS 1.1	0.766	0.002	0.012	0.769
$\geq$ 2 impaired domains (l2-testing)	MoCA total (adj)	0.861	-0.000	0.014	0.861
	MoCA total (raw), demographics	0.862	0.001	0.014	0.864
	MoCA subscores	0.869	0.002	0.011	0.871
	MoCA subscores, demographics	0.870	0.003	0.012	0.873
	MDS-UPDRS 1.1	0.687	-0.002	0.018	0.685
	MoCA total (adj), MDS-UPDRS 1.1	0.866	0.001	0.013	0.867
	MoCA total (raw), demogr., MDS-UPDRS 1.1	0.865	0.004	0.013	0.869
impairment (research clinic diagnosis)	MoCA total (adj)	0.738	-0.002	0.014	0.736
	MoCA total (raw), demographics	0.756	0.000	0.012	0.756
	MoCA subscores	0.740	0.002	0.013	0.742
	MoCA subscores, demographics	0.762	0.004	0.012	0.766
	l2 domain scores	0.728	0.001	0.013	0.729
	l2 domain scores, demographics	0.775	0.001	0.012	0.776
	MDS-UPDRS 1.1	0.812	0.000	0.008	0.812
	MoCA total (adj), MDS-UPDRS 1.1	0.863	0.000	0.008	0.863
	MoCA total (raw), demogr., MDS-UPDRS 1.1	0.870	0.001	0.008	0.871
dementia (research clinic diagnosis)	MoCA total (adj)	0.897	0.001	0.015	0.898
	MoCA total (raw), demographics	0.883	0.004	0.019	0.887
	MoCA subscores	0.897	0.007	0.016	0.903
	MoCA subscores, demographics	0.891	0.011	0.022	0.901
	l2 domain scores	0.887	0.002	0.020	0.889
	l2 domain scores, demographics	0.915	0.006	0.014	0.921
	MDS-UPDRS 1.1	0.914	0.003	0.017	0.917
	MoCA total (adj), MDS-UPDRS 1.1	0.963	0.001	0.008	0.964
	MoCA total (raw), demogr., MDS-UPDRS 1.1	0.961	0.004	0.010	0.964

Table 3: Selected cutoffs for 4 different cognitive outcomes and their respective specificity, sensitivity, positive and negative predictive values and accuracy (percentage correctly identified) with corresponding 95%-CIs. Shown are the cutoffs maximizing Youden's J=Sensitivity + Specificity, an equal proportions cutoff for which the percentage of observations below the cutoff is approximately equal to the percentage of observations with the respective outcome, the lowest cutoff with sensitivity ≥ 80 and NPV ≥ 80 as a screening cutoff and the highest cutoff with specificity ≥ 80 and PPV ≥ 80 as a diagnostic cutoff. Firstly cutoffs on the MoCA scale, precisely MoCA totals adjusted for years of education. Second, for research clinic outcomes, combined MDS-UPDRS 1.1 and MoCA cutoffs. For diagnosis of impairment the MDS-UPDRS 1.1. cutoff is ≥ 1, for dementia ≥ 2. Third, cutoffs for selected prediction models on the predicted probability scale (p) and, for dementia where the model exhibited unsatisfactory calibration on the linear predictor scale (lp).

		Criterion	Sensitivity	Specificity	PPV	NPV	Accuracy
$\geq$ 1 impaired domain (l2-testing)							
MoCA	Youden's J	MoCA $\leq$ 26	68.0 (65.4, 70.6)	69.5 (68.0, 71.0)	45.2 (43.0, 47.4)	85.5 (84.1, 86.7)	69.1 (67.8, 70.3)
	equal proportions	MoCA $\leq$ 25	57.4 (54.6, 60.2)	80.1 (78.6, 81.4)	51.6 (49.0, 54.2)	83.5 (82.3, 84.8)	73.9 (72.6, 75.2)
	screening	MoCA $\leq$ 28	88.7 (86.9, 90.5)	39.5 (37.8, 41.1)	35.1 (33.5, 36.8)	90.5 (88.9, 91.9)	52.8 (51.3, 54.2)
	diagnostic	MoCA $\leq$ 21	23.7 (21.4, 26.1)	97.8 (97.3, 98.3)	80.0 (76.1, 84.0)	77.6 (76.4, 78.8)	77.8 (76.7, 79.0)
$\geq$ 2 impaired domains (l2-testing)							
MoCA	Youden's J	MoCA $\leq$ 24	74.7 (70.5, 79.3)	83.1 (82.0, 84.2)	28.5 (25.8, 31.2)	97.3 (96.8, 97.9)	82.5 (81.3, 83.5)
	equal proportions	MoCA $\leq$ 21	49.7 (44.7, 55.0)	95.8 (95.2, 96.3)	51.3 (46.2, 56.5)	95.5 (94.8, 96.1)	92.0 (91.2, 92.7)
	screening	MoCA $\leq$ 25	81.4 (77.6, 85.3)	74.6 (73.2, 75.9)	22.4 (20.3, 24.5)	97.8 (97.3, 98.3)	75.2 (73.8, 76.4)
	diagnostic	MoCA $\leq$ 15	10.3 (7.5, 13.4)	99.8 (99.7, 99.9)	81.6 (70.2, 92.0)	92.5 (91.8, 93.3)	92.4 (91.7, 93.2)
impairment (research clinic diagnosis)							
MoCA	Youden's J	MoCA $\leq$ 25	58.0 (55.1, 61.0)	78.1 (76.7, 79.4)	43.1 (40.5, 45.6)	86.7 (85.5, 87.8)	73.6 (72.3, 74.9)
	equal proportions	MoCA $\leq$ 24	47.0 (43.9, 50.0)	85.7 (84.6, 86.8)	48.5 (45.4, 51.6)	85.0 (83.8, 86.1)	77.1 (75.9, 78.3)
	screening	MoCA $\leq$ 28	87.2 (85.1, 89.2)	37.5 (35.9, 39.1)	28.5 (26.9, 30.0)	91.1 (89.6, 92.5)	48.5 (47.1, 49.9)
	diagnostic	MoCA $\leq$ 16	5.4 (4.1, 6.8)	99.6 (99.4, 99.8)	80.9 (70.7, 89.3)	78.7 (77.5, 79.8)	78.7 (77.5, 79.9)
MoCA & UPDRS 1.1 $\geq$ 1	Youden's J	(MoCA $\leq$ 30)	81.9 (79.7, 84.2)	71.5 (69.9, 73.0)	45.1 (42.9, 47.3)	93.3 (92.4, 94.2)	73.8 (72.5, 75.1)
	equal proportions	MoCA $\leq$ 26	57.2 (54.3, 60.4)	88.9 (87.8, 89.9)	59.6 (56.5, 62.8)	87.9 (86.8, 89.0)	81.8 (80.7, 83.0)
	screening	(MoCA $\leq$ 30)	81.9 (79.7, 84.2)	71.5 (69.9, 73.0)	45.1 (42.9, 47.3)	93.3 (92.4, 94.2)	73.8 (72.5, 75.1)
	diagnostic	MoCA $\leq$ 21	19.8 (17.4, 22.3)	98.6 (98.2, 99.0)	80.4 (75.3, 85.2)	81.1 (80.0, 82.3)	81.1 (80.0, 82.2)
MoCA & UPDRS 1.1 model	Youden's J	$p \geq 0.19$	78.4 (75.9, 80.9)	77.4 (76.1, 78.8)	49.8 (47.5, 52.2)	92.6 (91.7, 93.5)	77.7 (76.5, 78.8)
	equal proportions	$p \geq 0.32$	63.3 (60.5, 66.1)	89.6 (88.7, 90.6)	63.6 (60.7, 66.5)	89.5 (88.5, 90.5)	83.8 (82.8, 84.8)
	screening	$p \geq 0.17$	82.6 (80.3, 84.8)	71.6 (70.2, 73.1)	45.4 (43.3, 47.7)	93.5 (92.6, 94.4)	74.0 (72.9, 75.3)
	diagnostic	$p \geq 0.57$	42.2 (39.2, 45.2)	97.0 (96.5, 97.5)	80.1 (77.0, 83.4)	85.4 (84.3, 86.5)	84.8 (83.8, 85.8)
dementia (research clinic diagnosis)							
MoCA	Youden's J	MoCA $\leq$ 23	77.9 (70.1, 85.6)	85.8 (84.8, 86.8)	11.4 (9.2, 13.7)	99.4 (99.2, 99.6)	85.6 (84.6, 86.6)
	equal proportions	MoCA $\leq$ 17	32.7 (24.0, 41.9)	98.6 (98.3, 99.0)	35.4 (26.9, 45.5)	98.4 (98.1, 98.8)	97.1 (96.6, 97.6)
	screening	MoCA $\leq$ 25	91.3 (85.7, 96.3)	71.5 (70.2, 72.8)	7.0 (5.7, 8.3)	99.7 (99.5, 99.9)	72.0 (70.6, 73.2)
	diagnostic	-					
MoCA & UPDRS 1.1 $\geq$ 2	Youden's J	MoCA $\leq$ 29	84.6 (78.2, 91.2)	91.6 (90.8, 92.3)	19.0 (15.7, 22.7)	99.6 (99.4, 99.8)	91.4 (90.6, 92.2)
	equal proportions	MoCA $\leq$ 20	46.1 (36.9, 55.9)	99.0 (98.7, 99.3)	52.2 (42.4, 62.5)	98.8 (98.4, 99.1)	97.8 (97.4, 98.2)
	screening	MoCA $\leq$ 28	83.7 (76.9, 90.4)	92.3 (91.5, 93.0)	20.2 (16.7, 24.0)	99.6 (99.4, 99.8)	92.1 (91.3, 92.8)
	diagnostic	-					
MoCA & UPDRS 1.1 model	Youden's J	$lp \geq -3.41$	91.3 (85.4, 96.2)	89.5 (88.7, 90.4)	16.9 (13.9, 20.0)	99.8 (99.6, 99.9)	89.6 (88.7, 90.4)
	equal proportions	$lp \geq -1.27$	52.9 (43.3, 61.7)	98.8 (98.5, 99.1)	51.4 (41.8, 61.1)	98.9 (98.6, 99.2)	97.8 (97.4, 98.2)
	screening	$lp \geq -2.31$	81.7 (74.3, 88.7)	96.0 (95.5, 96.6)	32.6 (27.0, 38.3)	99.6 (99.4, 99.7)	95.7 (95.2, 96.3)
	diagnostic	$lp \geq 1.82$	4.8 (1.0, 9.3)	100.0 (99.9, 100.0)	83.3 (50.0, 100.0)	97.8 (97.4, 98.2)	97.8 (97.4, 98.2)

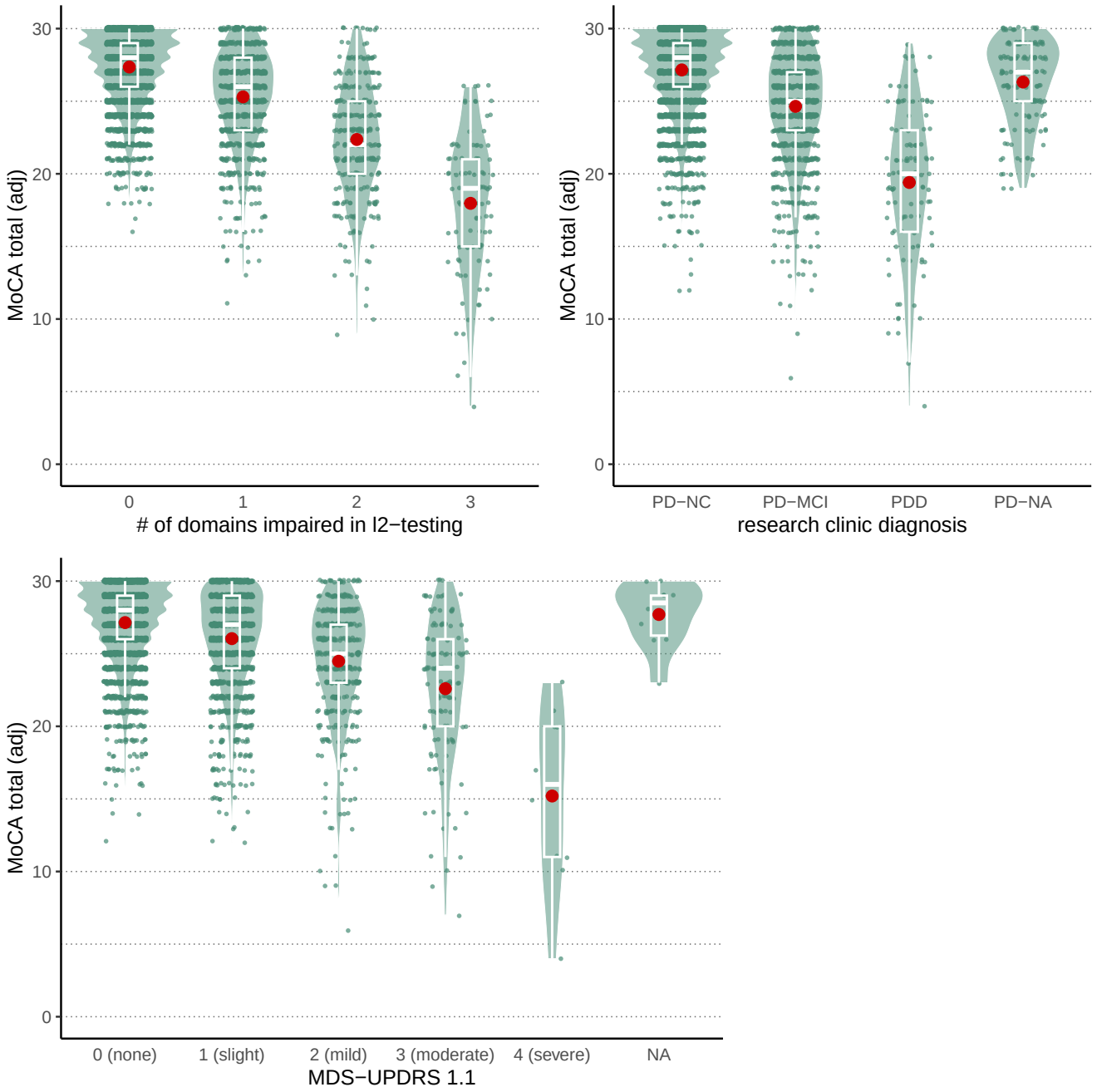

Figure 1: MoCA total scores (adjusted for years of education) in the PD arm of the PPMI cohort stratified by number of cognitive domains impaired in level-2 testing (A), cognitive research clinic diagnosis (B) and functional impairment as as recorded in MDS-UPDRS 1 Question 1 (C). Means, standard deviations and numbers of observations are listed in Table 1.

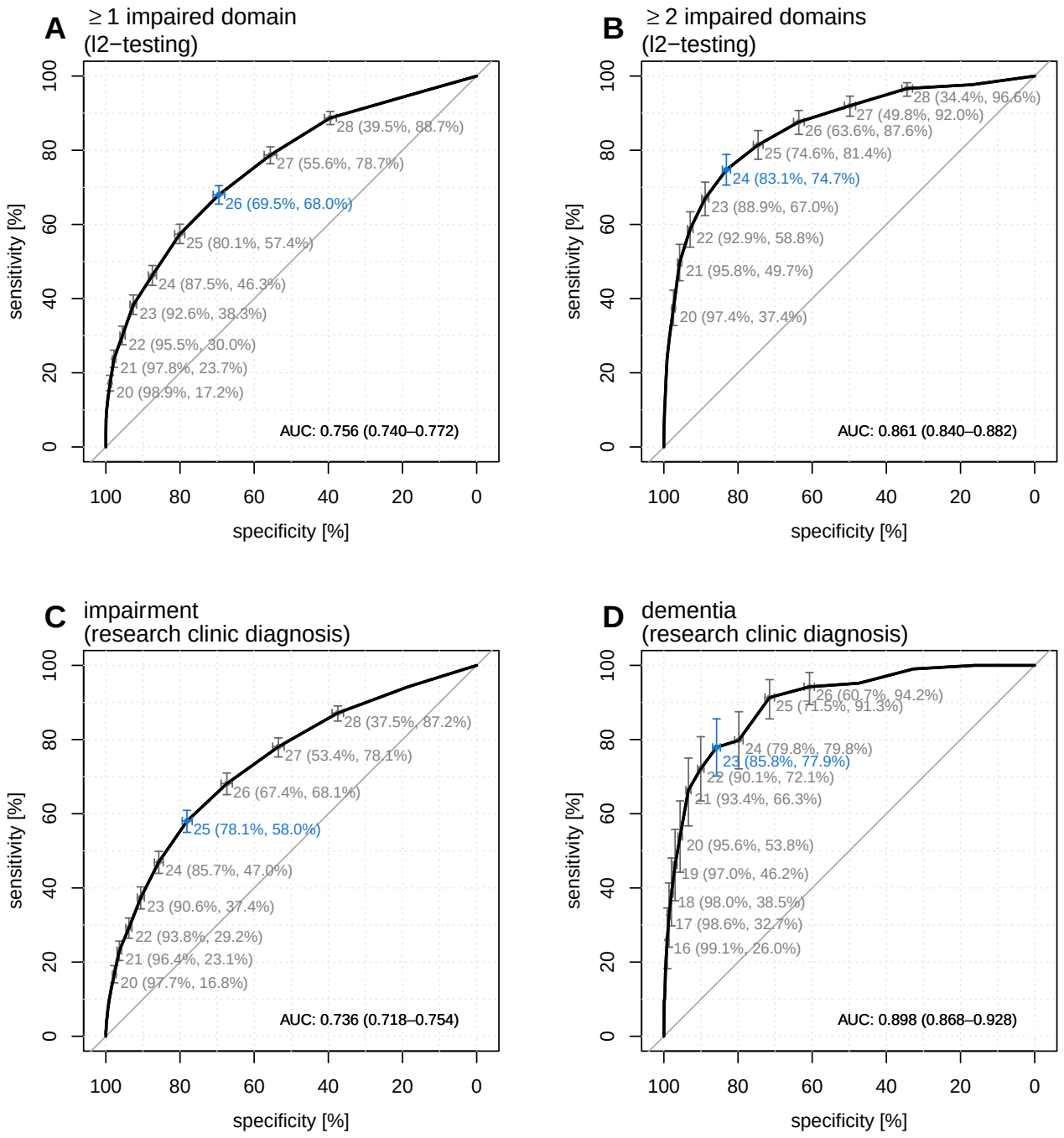

Figure 2: Discriminative ability of the MoCA total scores (adjusted for years of education): ROC curves and AUC's for 4 different cognitive outcomes: ≥ 1 impaired domain in l2-testing, ≥ 2 impaired domains in l2-testing, research clinic diagnosis of any impairment (PD-MCI or PDD) vs. none (PD-NC), research clinic diagnosis of dementia (PDD) vs. no dementia (PD-MCI or PD-NC). The cutpoints maximizing Youden's J = sensitivity + specificity -1 are highlighted in blue.

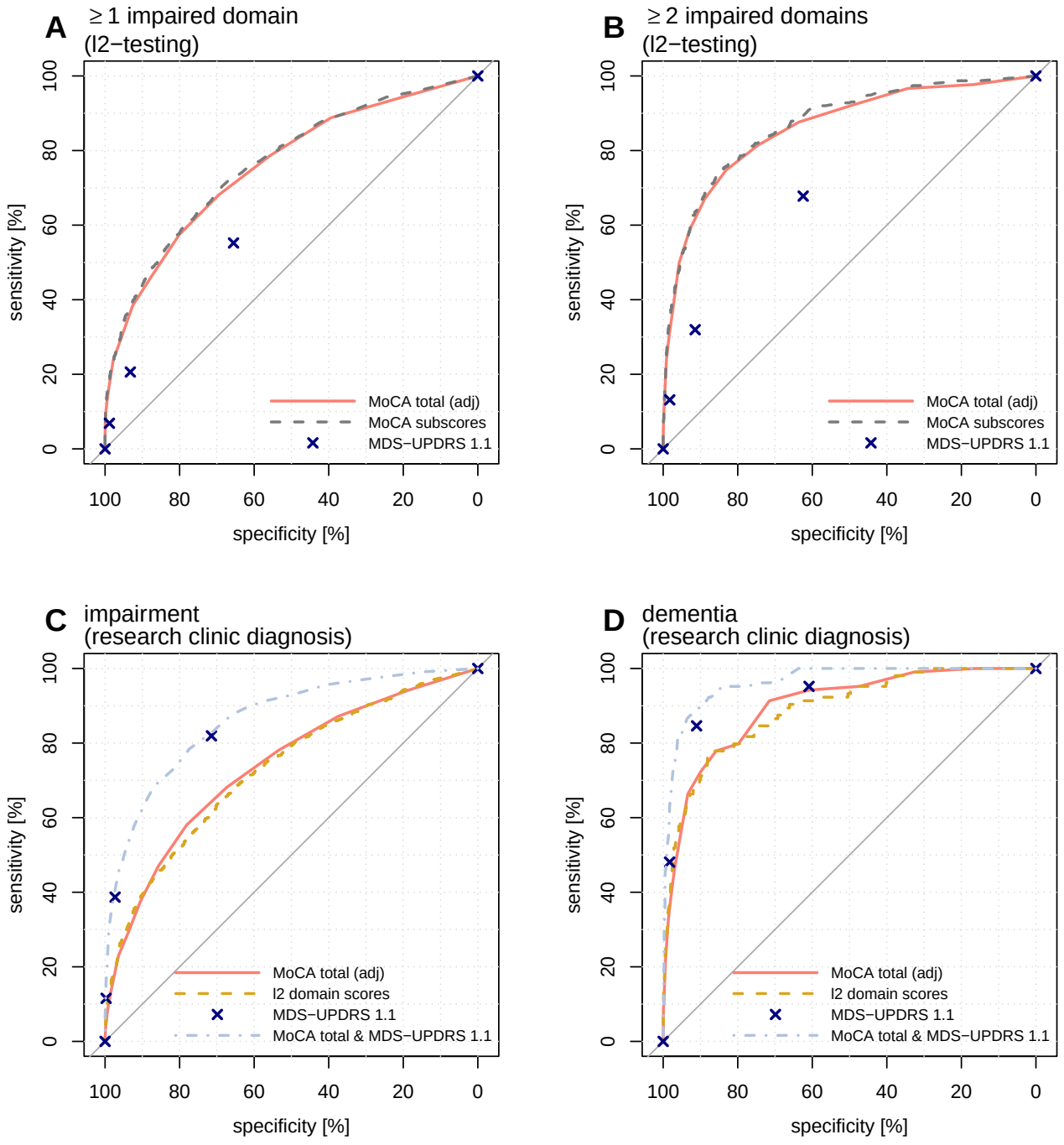

Figure 3: ROC curves of MoCA total scores, MDS-UDPDRS 1 Question 1 and selected prediction models (see legends in panels A-D) for 4 different cognitive outcomes. The corresponding AUC's are listed in Table 2.